

Structure and Dynamics of End-to-End Loop Formation of the Penta-Peptide Cys-Ala-Gly-Gln-Trp in Implicit Solvents

In-Chul Yeh* and Anders Wallqvist

Biotechnology High Performance Computing Software Applications Institute, Telemedicine and Advanced Technology Research Center, U.S. Army Medical Research and Materiel Command, Fort Detrick, Maryland

Received: May 1, 2009; Revised Manuscript Received: July 15, 2009

To investigate the effects of implicit solvents on peptide structure and dynamics, we performed extensive molecular dynamics simulations on the penta-peptide Cys-Ala-Gly-Gln-Trp. Two different implicit solvent models based on the CHARMM22 all-atom force field were used. Structural properties of the peptide such as distributions of end-to-end distances and dihedral angles obtained from molecular dynamics simulations with implicit solvent models were in a good agreement with those obtained from a previous explicit solvent simulation using the same force field. Representative structures observed in explicit solvent were sampled by implicit solvent models but with different relative probabilities. However, we observed significant differences in dynamical properties in explicit and implicit solvent models when we used traditional methods for the temperature control, such as Nosé–Hoover or Berendsen thermostats. The explicitly solvated peptide displayed the slowest dynamics in both end-to-end contact formation and intrinsic diffusive motion of end-to-end distances. A closer agreement between implicit and explicit solvated peptide dynamics was observed when Langevin dynamics with a friction coefficient of 10 ps^{-1} was used to maintain the temperature of the systems.

Introduction

Molecular dynamics (MD) simulation techniques are widely used to study structure and dynamics of biomolecular systems in atomic detail. An accurate and efficient treatment of solvation continues to be a challenge in simulations of biomolecular systems. The development of computationally efficient methods such as the particle mesh Ewald method^{1,2} for calculating long-range electrostatic forces contributed to a wider adoption of explicit representation of water molecules in simulations of biological macromolecules. However, the computational cost of employing explicit solvent simulations is still significant. In addition, the large number of degrees of freedom due to explicit solvent water molecules poses significant challenges in applying advanced sampling techniques such as replica-exchange MD.^{3–6} Various implicit solvent models^{7–13} have been developed as an alternative to expensive explicit-solvent simulations. These implicit solvent models have been mainly used to explore structural properties of proteins. Zhou and Berne¹⁴ compared the free energy landscapes of the C-terminal β -hairpin of protein G obtained by explicit and implicit solvent simulations based on the OPLS-AA¹⁵ force field with an efficient conformational sampling by a highly parallel replica-exchange method. They found that the free energy landscape from the surface-generalized Born (SGB)¹² implicit solvent model is quite different from that of the explicit solvent model. In the free energy landscape calculated by the SGB implicit solvent model, the lowest free energy state was not the native β -strand structure observed in the explicit solvent simulation and the experiments. They attributed this behavior to an overly strong salt bridge effect^{16,17} between charged residues in the SGB implicit solvent model. Felts et al.¹⁸ obtained a better agreement with the explicit solvent

results by using the effective potential consisting of the OPLS-AA force field and the analytical generalized Born plus nonpolar (AGBNP)¹³ implicit solvent model. Increased dielectric screening incorporated in the AGBNP model reduced the stability of incorrectly formed salt bridges, which tend to disrupt the formation of the β -hairpin. This underscores that a proper balance between the nonpolar effective potential terms, which drive compaction, and the electrostatic interactions, which promote secondary structure formation in the implicit solvent model, has to be made to reproduce structural properties observed with the explicit solvent simulations or experiments.¹⁸ Fan et al.¹⁹ compared performances of three different generalized Born (GB) models in reproducing the backbone root-mean-square deviation from the native structure, the number of hydrogen bonds retained in the simulation, and the experimental and calculated radius of gyration for 10 different proteins. Recently, Hamelberg et al.²⁰ probed the role of hydration on protein dynamics by MD simulations using a generalized Born (GB) implicit solvation model based on the AMBER force field.¹¹ Similarly, Feig²¹ examined kinetic properties of the alanine dipeptide and selected proteins in an implicit solvent modeled by the generalized Born molecular volume (GBMV) method.^{9,10} Here we present theoretical studies of end-to-end loop formation dynamics of a small peptide employing different implicit solvent models in order to investigate dynamical artifacts introduced by implicit solvents.

The end-to-end contact formation of amino acid residues is one of the elementary processes in protein folding and has been extensively studied both experimentally^{22–26} and theoretically.^{25,27–36} The end-to-end contact formation dynamics of the penta-peptide Cys-Ala-Gly-Gln-Trp (CAGQW) has been experimentally studied with triplet quenching.²² The time scale of the end-to-end contact formation of this peptide was determined to be on the order of 100 ns, a time-scale which is accessible in atomically detailed MD simulations. Extensive MD simulations of the peptide in explicit solvent with an aggregate

* To whom correspondence should be addressed. Address: Biotechnology HPC Software Applications Institute, ATTN: MCMR-TT, Building 363 Miller Drive, Fort Detrick, Maryland 21702-5012. Tel.: (301) 619-0702. Fax: (301) 619-1983. E-mail: icy@bioanalysis.org.

simulation time on the order of microsecond have been performed with both AMBER94³⁷ and CHARMM22³⁸ force fields.³⁰ The time scale of the end-to-end contact formation estimated by the simulation is on the order of 10 ns. The discrepancy of time scale between the experiment and the simulation was attributed to the finite quenching rate of the Trp by the Cys residue. Indeed, a recent fluorescence-based measurement²⁵ obtained a much faster time scale of 14 ns for the end-to-end contact formation of a similar penta-peptide where the Cys of the CAGQW peptide was replaced by an efficient fluorescent probe, 2,3-diazabicyclo[2.2.2]oct-2-ene-labeled asparagine (Dbo). Explicit solvent simulation studies by Roccatano et al.^{25,31} on the CAGQW and Dbo-AGQW penta-peptides with GROMOS96³⁹ and OPLS-AA/L⁴⁰ force fields also found that the simulated dynamics is faster than the experimentally measured dynamics. They also noted that the force field had a large influence on the dynamical properties on the peptide based on the effective size of the barrier to formation of the compact structures. Faster end-to-end formation kinetics was observed for GROMOS96 and OPLS-AA/L force fields as compared with AMBER94 or CHARMM22 force fields.

In this study, we investigated the effect of treating the solvent implicitly on structural and dynamical properties of the CAGQW peptide by performing extensive MD simulations with generalized Born implicit solvent models based on the CHARMM22 force field and comparing the result to the explicit-solvent simulation using the same force field.³⁰ We observed that structural properties of the peptide sampled by explicit solvent simulations can be reasonably represented by simulations with generalized Born implicit solvent models. However, we observed a significant difference in dynamical properties of the peptide obtained by simulations with explicit and implicit solvent models when we employed conventional methods for the temperature control. This difference can partly be removed by performing simulations coupled with Langevin dynamics at a friction coefficient of 10 ps⁻¹, as has been noted previously.^{20,21}

Methods

We performed MD simulations of the CACQW penta-peptide with the CHARMM MD simulation program,⁴¹ version c31b2. The MMTSB Tool Set⁴² was used to facilitate the simulations. As in the experiments,²² the C-terminus of the peptide was amidated while the N-terminus remained unblocked. We used two generalized Born implicit solvent models available in CHARMM, GBSW (generalized Born with a smooth switching)⁸ and GBMV2 (generalized Born molecular volume version 2).¹⁰ In the GBMV2 model, the β_S -parameter determines the smoothness of the molecular volume at the protein-solvent dielectric boundary, while the S_0 -variable is adjusted to match Poisson solvation energies. Here we used the original set of parameters with values of β_S of -20 and S_0 of 0.7.^{10,43} In the GBSW model, three parameters are specified, a smoothing length w , a parameter that modulates the Coulomb field term a_0 , and an empirical correction term a_1 . We employed the set of parameter values denoted as GBSW2,^{43,44} which were optimized to fit the Lee-Richards molecular surface⁴⁵ Poisson results. Thus, we set the smoothing length w to 0.2 Å and a_0 and a_1 to 1.2045 and 0.1866, respectively. We also performed simulations with the parameter set in the original GBSW model denoted as GBSW1 with w , a_0 , and a_1 of 0.3 Å, -0.1801, and 1.8174, respectively. The atomic input radii optimized by Nina et al.⁴⁶ and Chen et al.⁴⁷ have been used for simulations with the GBSW1 model. The input radii for simulations with GBMV2 and GBSW2 were taken to be the standard CHARMM

atomic radii as these models were optimized with these radii.^{10,43,44} These implicit solvent models have been successfully applied in recent simulation studies to calculate thermodynamic properties of the SH3 domain of α -spectrin⁴⁸ and the absolute ligand binding free energy to a ribosome-targeting protein.⁴⁹ The nonpolar solvation energy was accounted for by the default energy terms expressed as the linear product of the solvent-exposed surface area of the solute and a phenomenological surface tension coefficient (γ) set to 0.030 and 0.015 kcal/(mol·Å²) for the GBSW2 and GBMV2 models, respectively. The same γ value of 0.015 kcal/(mol·Å²) was also used in a previous simulation with GBMV2 model.⁴³ The differences in γ used in GBSW2 and GBMV2 can partly be attributed to the different surface representations used to calculate the solvent exposed surface area.⁴⁹ Two different γ values of 0.030 and 0.005 kcal/(mol·Å²) used in previous simulations^{8,47} with the GBSW1 model were employed to gauge the effect of the choice of γ on the structural properties of the peptide. We also performed MD simulations of the CACQW penta-peptide in dielectric continuum with varying dielectric constants ϵ of 1, 4, and 80. We used data from a previous simulation of the CACQW peptide in explicit solvent for comparison and further analysis.³⁰

We used several different methods for temperature control, a Nosé-Hoover thermostat^{50,51} with a thermal inertia parameter of 50 kcal/(mol ps²), a Berendsen thermostat⁵² with a coupling constant of 5 ps, and Langevin dynamics⁵³ with friction coefficients of 2, 10, and 60 ps⁻¹ applied to non-hydrogen atoms. Although the true friction is nonisotropic and atom dependent, a single isotropic friction coefficient was applied to all heavy atoms for simplicity. A MD integration time step of 2 fs was used except for Langevin dynamics simulations employing a friction coefficient of 60 ps⁻¹, where a shorter time step of 1 fs was required. Distances used for the onset of damping non-bonded interaction, the cutoff for nonbonded interactions, and the cutoff for nonbonded list generation were 20, 22, and 25 Å, respectively. For a small 5-residue peptide used in our study, these cutoff distances amount to calculations without any cutoff distances. Covalent bonds between the heavy atoms and hydrogens were constrained by the SHAKE algorithm.⁵⁴

The fully extended conformation of the CACQW peptide was used as a starting configuration of a 1 ns MD simulation at a temperature of 700 K controlled by a Nosé-Hoover thermostat. Fifty configurations sampled from this simulation trajectory from each implicit solvent model were then used as starting configurations for 50 independent production runs at 300 K for the corresponding implicit solvent model studied in this work. Each of the 50 simulations lasted 21 ns and trajectories were sampled after an initial 100 ps equilibration unless specified otherwise. A simple equal time-block analysis showed no significant differences in thermodynamic and structural properties calculated with the data from different time blocks, indicating that the proper equilibration was achieved. Each 21 ns simulation required about 53.7 and 21.5 h to finish on a single core of 3.0 GHz Intel Woodcrest processor with GBMV2 and GBSW2 models, respectively. Coordinates were saved at every 1 ps for further analysis.

The end-to-end distance (r_c) of the peptide was defined as the distance between the sulfur atom of the initial cysteine side chain and the closest non-hydrogen atom of the tryptophan indole ring. We defined end-to-end contact formation as taking place when the end-to-end distance is less than 4 Å. The same value of the contact distance (d_c) was used in previous experimental²² and computational³⁰ studies. We also used contact

distances estimated from simulations for the analysis of end-to-end contact formation dynamics. Then $S_{\text{open}}(t)$, the survival probability of the open noncontact state of the peptide between time 0 and t , is 1 for $t < t_c$, where t_c is the first-contact time, and 0 for $t \geq t_c$. We assumed that every saved noncontact configuration ($r_c > d_c$) for each MD simulation run is a valid starting point. The reported survival probability $S_{\text{open}}(t)$ was estimated by the averaged survival probabilities over all starting points of the respective simulations. The average contact formation time τ_{open} or the lifetime of the peptide in the noncontact state was estimated by calculating the mean first passage time from the following expression^{27,30}

$$\tau_{\text{open}} = \int_0^{\infty} S_{\text{open}}(t) dt \quad (1)$$

Conversely, we can define the survival probability of the contact state $S_{\text{closed}}(t)$ between time 0 and t , as 1 for $t < t_o$, where t_o is the first contact-opening time, and 0 for $t \geq t_o$. Integration of this survival probability yields the corresponding τ_{closed} , the lifetime of the peptide in the closed contact state. The contact formation rate k_c was defined as τ_{open}^{-1} , and the corresponding contact opening rate k_o as $\tau_{\text{closed}}^{-1}$. The rate constants estimated from this definition are exact only when $S(t)$ is exponential and approximate otherwise.

Clustering analysis of structures of the CAGQW peptide that form end-to-end contacts was performed with the average-linkage clustering algorithm⁵⁵ implemented in the *ptraj* module of Amber, version 10.⁵⁶

Error bars reported in this work were based on the 95% confidence interval assuming random sampling (1.96 times standard error) and obtained by analysis of 50 independent simulations runs for each model.

Results and Discussions

Structural Properties. End-to-End Distance Distribution.

In general, solvation in aqueous media provides the solute with the capability to form a multitude of hydrogen-bonded interactions with individual water molecules. For a small peptide, this effect is typically manifested in the attenuation of intramolecular hydrogen bonding interactions. Figure 1 shows distributions of end-to-end distances of the CAGQW peptide in dielectric-continuum simulations with dielectric constants ϵ of 1, 4, and 80 contrasted with the results obtained from using explicit solvent water molecules. In the vacuum simulation ($\epsilon = 1$) where the Coulomb interactions are not shielded, conformations characterized by strong internal hydrogen bonding between carbonyl oxygens and N-terminal hydrogens dominated. This is shown by the strong preference of end-to-end distances located around 8 Å in the vacuum simulation. As the dielectric screening is increased from vacuum to a simulated general hydrophobic environment of ($\epsilon = 4$) to an aqueous environment ($\epsilon = 80$), the strength of the internal hydrogen bonding formed by the peptide-groups diminishes. However, the simple dielectric screening model is not capable of capturing the full effect of the explicit solvation. In particular, the high dielectric simulation, corresponding to an aqueous environment ($\epsilon = 80$), overemphasizes the contact conformation ($r_c \leq 4$ Å) and underrepresents the extended conformations ($r_c \geq 10$ Å). The extended conformations in the high dielectric approximation essentially lack the stabilizing peptide/water hydrogen bonding that occurs in the explicit solvent model.

Influence of Thermostats on Implicit Models. From a statistical mechanical point of view, MD trajectories are used

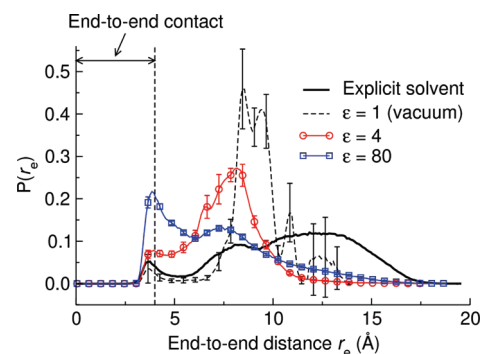


Figure 1. $P(r_c)$, probability distributions of the end-to-end distance (r_c) of the CAGQW peptide obtained by simulations with explicit solvent model and with different dielectric continuum models. The distribution obtained with the explicit solvent model is shown as a solid line. The results for the explicit solvent model shown in this and the following figures were obtained from a previous simulation of the CAGQW peptide in explicit solvent.³⁰ Distributions with continuum models with dielectric constant ϵ of 1 (vacuum), 4, and 80 are represented by a dotted line and lines with circles and squares, respectively. A dashed vertical line ($r_c = 4$ Å) indicates a boundary of end-to-end contact formation. Error bars in this and the following figures show the 95% confidence interval estimated by multiplying 1.96 to standard errors obtained by analysis of 50 independent simulations runs for each model.

to generate conformations from which one can form thermodynamic averages. Provided that the sampling is ergodic and that the time-step used in the simulations is small enough, the method used to generate the trajectories should not influence the results. In practice, for smaller systems where coupling to solvent degrees of freedom can dominate the mechanisms of temperature equilibration, the choice of thermostat to control the temperature can and will influence calculated properties.^{20,21,57,58} Figure 2a illustrates this effect for distributions of end-to-end distances calculated from MD simulations using the GBSW2 implicit solvent model with five different thermostats.

Overall, the distributions of end-to-end distances derived from simulations with the GBSW2 model capture the major features of the explicit solvent simulation in sharp contrast to the dielectric continuum model results shown in Figure 1. In particular, the implicit model qualitatively captures the stabilization of the extended conformations ($r_c \geq 10$ Å). However, the details of the distributions depended on the choice of thermostats, both at distances near the contact distance and at extended distances. In simulations with simplified models without explicit solvent, Nosé–Hoover or Berendsen temperature control methods with a single thermostat were shown to introduce erroneous descriptions of thermodynamic and kinetic properties of systems.^{21,58,59} Figure 2a shows that simulations with Nosé–Hoover or Berendsen temperature control methods result in slightly increased populations of more compact states with smaller end-to-end distances compared to simulations with the explicit solvent model and Langevin dynamics. It was recently shown that simulations with GBMV2 implicit solvent model combined with Langevin dynamics with a friction coefficient of 10 ps^{-1} reproduced kinetic properties of peptides and proteins obtained from explicit solvent simulations and experiments.²¹ From our simulations, we can see that Langevin dynamics thermostats with different friction coefficients resulted in a generally similar distribution of end-to-end distances and a closer agreement with the explicit solvent model. However, a smaller time step of 1 fs and a longer equilibration time of 20 ns were necessary for the simulation employing a Langevin friction coefficient of 60 ps^{-1} in order to obtain a distribution of end-

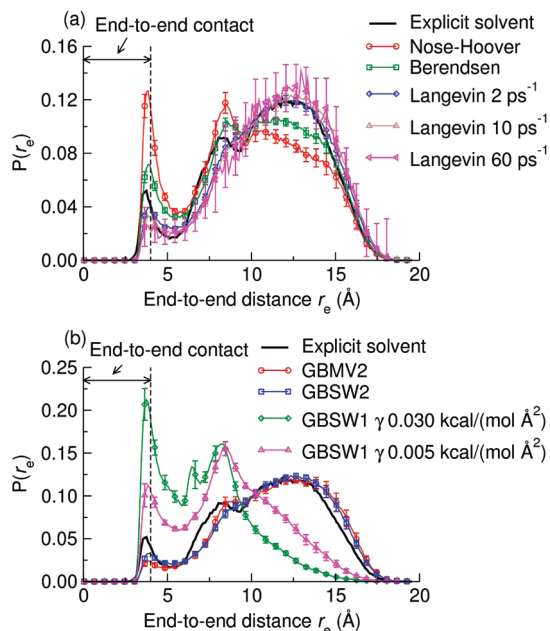


Figure 2. (a) Probability distributions of end-to-end distance of the CAGQW peptide obtained by simulations with the GBSW2 implicit solvent model using different temperature-control methods. The distribution obtained with the explicit solvent model is also shown as a solid line. Distributions obtained with Nosé–Hoover and Berendsen thermostats are shown with lines with circles and squares, respectively. Distributions obtained with simulations coupled with Langevin dynamics with friction coefficients with 2, 10, and 60 ps^{-1} are represented by lines with diamonds and triangles pointing up and left, respectively. For the distribution generated with a friction coefficient of 60 ps^{-1} , we only used data from the last 1 ns of each 21 ns trajectory due to the long equilibration required for this simulation. (b) Comparison of probability distributions of end-to-end distance of the CAGQW peptide obtained by simulations with explicit solvent (solid line), GBMV2 (a line with circles), GBSW2 (a line with triangles), GBSW1 with γ of 0.030 $\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$ (a line with diamonds), and GBSW1 with γ of 0.005 $\text{kcal}/(\text{mol}\cdot\text{\AA}^2)$ (a line with triangles). The results from the GBMV2, GBSW2, and GBSW1 models were obtained with the Langevin friction coefficient of 10 ps^{-1} .

to-end distances similar to those obtained with smaller friction coefficients. Consistent with earlier work,²¹ choices of the friction coefficient 10 ps^{-1} or below tend to have minimal impact on overall structural and dynamical properties. Thus, we adopted Langevin dynamics with a friction coefficient of 10 ps^{-1} as the method of choice to generate ensemble trajectories for generalized Born implicit solvent models.

GBSW2 versus GBMV2 and GBSW1 Models. Figure 2b compares distributions of end-to-end distances obtained from GBMV2, GBSW1, and GBSW2 implicit solvent models to the explicit solvent model. Overall, both GBSW2 and GBMV2 implicit solvent models showed reasonable agreements of end-to-end distance distributions compared with explicit solvent simulations. Both implicit models have significant populations at extended conformations with $r_e \geq 10$ \AA . However, both implicit models consistently generated smaller populations of peptides whose end-to-end distances are within the contact distance of 4 \AA as compared to the explicit solvent simulation. Further, neither model reproduced the small but pronounced peak near the end-to-end distance of 8.3 \AA . This peak is characterized by peptide conformations where the end residues Cys and Trp make simultaneous contacts with the Gln residue in the middle. These interactions are diminished in the implicit models; in particular the Cys/Gln interaction in the GBSW2 model is not properly accounted for. In contrast, distributions

of end-to-end distances calculated from simulations with the GBSW1 model overestimated populations within the contact distance and near the 8.3 \AA peak and underestimated the population of extended conformations. Even though a smaller phenomenological surface tension coefficient γ with the GBSW1 model increased the relative population of extended conformations, a significant difference remained between distributions of end-to-end distances with GBSW1 and explicit solvent models.

Dihedral Angle Distributions. To further characterize structural properties of the peptide, we calculated the potential of mean force (PMF) associated with the dihedral angles ϕ and ψ of the three middle amino acid residues, Ala, Gly, and Gln. Figure 3 shows PMF surfaces obtained from explicit solvent simulations together with those obtained from GBMV2 and GBSW2 models. For Ala and Gln, both implicit models showed dihedral angle PMFs similar to what was obtained from the explicit solvent simulation. The PMF surface of the Gly residue located in the middle of the CAGQW peptide is associated with significant populations for dihedral angles near $\phi = \pm 180$ degrees and $\psi = \pm 180$ degrees, consistent with the extended conformations shown in Figure 2b. These are qualitatively similar for both implicit models and the explicit solvent results. However, there are small differences in the PMF surfaces that can primarily be related to the shape of the extended conformations. Overall, we obtained a good agreement between PMF surfaces of dihedral angles calculated with implicit and explicit solvent models. The CAGQW peptide studied here does not have any charged residues, which could present problematic salt bridge effects as observed in free energy landscapes of the C-terminal β -hairpin of protein G calculated with the SGB implicit solvent model.¹⁴ In addition, both GBMV2 and GBSW2 models used in this study were developed to reproduce a dielectric boundary defined by Lee–Richards molecular surface,⁴⁵ which minimizes the overestimation of the solvation energy due to the interstitial high dielectric regions that occurs in the simplistic treatment of the dielectric boundary.^{10,43,60} Indeed, a good agreement with the PMF distribution of dihedral angles calculated with the explicit solvent model was also observed in other studies with generalized Born models where the interstitial high dielectric regions are properly corrected.^{20,21,60}

End-to-End Contact Ensemble. While the above dihedral angle analysis showed the average distribution derived from all conformational samples, we would like to focus on the differences exhibited in the set of peptide conformations that form the compact contact state $r_e \leq 4$ \AA . We selected 1000 structures whose end-to-end distances are within the contact distance of 4 \AA from the peptide conformations generated by each solvent model. We performed clustering analysis on these structures as described in Methods and selected the top three out of five representative clusters. The top three clusters typically contained more than 95% of tested contact-conformations. Figure 4 shows representative end-to-end contact structures from these clusters obtained by the explicit solvent simulation. These contact structures were all sampled by the simulations performed with implicit solvent models but with different probabilities. The structures from the most dominant cluster (65% of the selected structures belonged to this cluster) in explicit solvent simulation (Cluster 1) resembled those in the most populated cluster generated by the GBMV2 model. However, structures from the second most dominant (20%) cluster seen in the explicit solvent simulations (Cluster 2) resembled those in the third largest cluster in simulations with the GBMV2 model. Structures from the most populated (54%) clusters with the GBSW2 model

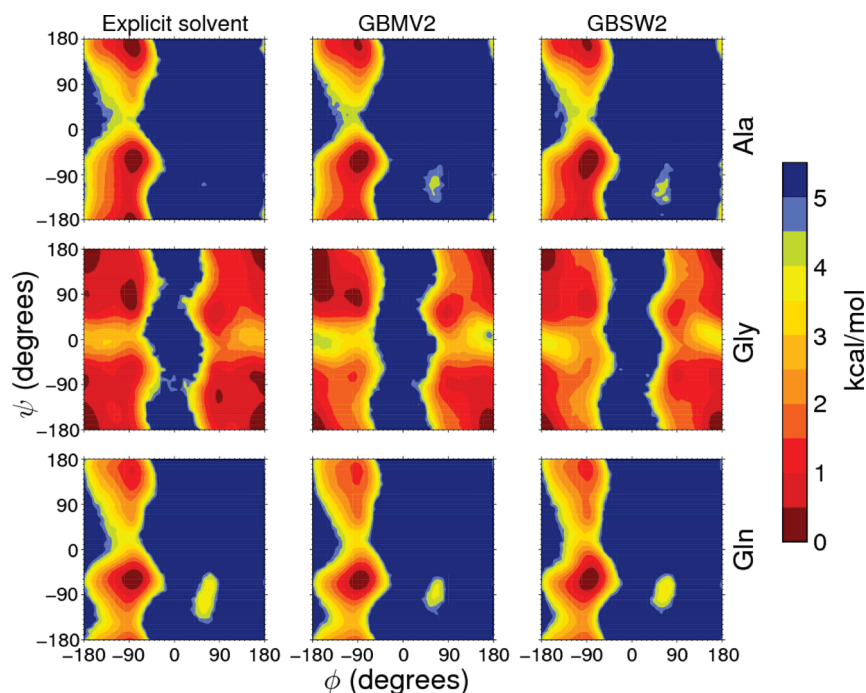


Figure 3. The potential of mean force (PMF) distributions of dihedral angle of the three central residues of the CAGQW peptide in explicit and implicit solvent models. The PMF was calculated from $-RT \log(P(\phi, \psi))$, where $P(\phi, \psi)$ is the probability distribution of the dihedral angles ϕ and ψ normalized by its maximum, R is the gas constant, and T is the temperature (300 K).

resembled those in the third largest (14%) cluster in explicit solvent simulations (Cluster 3). The relative importance or weight of these three contact cluster conformations could not be captured by the implicit models. Figure 4 also shows dihedral angle ϕ and ψ distributions in terms of PMF surfaces for the three representative conformations for each cluster. The dihedral angle distribution of Gly displays a distinct pattern for each cluster, essentially relating how the peptide bends slightly differently to form the three distinct contact states. The Ala residue in Cluster 2 and the Gln residue in Cluster 1 favor β -strand like dihedral conformations.

Dynamical Properties. End-to-End Contact Formation Dynamics. We calculated the survival probability of the open (noncontact) state $S_{\text{open}}(t)$ for implicit-solvent models coupled with Langevin friction coefficient of 10 ps^{-1} and explicit solvent simulation. $S_{\text{open}}(t)$ in Figure 5a shows that the open noncontact state of the CAGQW peptide survived longer in explicit solvent than in implicit solvents models. Average lifetimes of the peptide in the open state τ_{open} were estimated by applying eq 1 to the time dependent survival probabilities in Figure 5a. This yielded τ_{open} of 1.6, 0.9, and 0.6 ns for explicit, GBMV2, and GBSW2 solvent models, respectively, as summarized in Table 1. Relative magnitudes of τ_{open} values are consistent with the respective barrier heights of the transition from noncontact to contact state in the potential of mean force (PMF) distribution of end-to-end distances shown in Figure 5b. The larger the barrier the longer the average lifetime of the peptide is in the open state. In addition, Table 1 lists τ_{open} obtained for simulations with different thermostats and friction coefficients. τ_{open} values estimated from implicit-solvent simulations coupled with Langevin dynamics with the friction coefficient of 10 ps^{-1} are significantly larger than those obtained with Nosé–Hoover or Berendsen thermostats and are in better agreement with the explicit solvent simulation. It is also clear that the choice of the friction coefficient determines the dynamics of the contact formation and provides a heuristic that can be used to match the explicit-solvent dynamics results. We also calculated the

survival probability of the closed (contact) state as a function of time (data not shown) and the average lifetimes of the peptide in the closed state τ_{closed} listed in Table 1.

Unlike structural properties of the peptide, dynamics of end-to-end contacts are sensitive to the choice of the contact distance used in the analysis. Therefore, we calculated survival probability distributions and lifetimes of open and closed states of the peptide with contact distances defined as the end-to-end distances corresponding to peak positions separating open and closed states in the PMF distribution shown in Figure 5b. The contact distances according to this definition were 5.6, 6.1, and 5.7 Å for explicit solvent, GBSW2, and GBMV2 models, respectively. The results are shown in Figure 5c and Table 2. Overall, we observed faster end-to-end contact formations and slower contact openings with larger contact distances. However, the relative speeds of contact formation and opening among different solvent models remained the same. With this definition of contact distances, we estimated the equilibrium fraction of contact states as $k_c/(k_c + k_o)$ or $\tau_{\text{closed}}/(\tau_{\text{open}} + \tau_{\text{closed}})$ and compared them with those observed in the simulations in Table 2. Considering the approximations involved in characterizing complex dynamics of formation and opening of various end-to-end contacts with single rate constants, the agreement between estimated and observed values for the equilibrium fraction of contact states is quite good.

Diffusion Dynamics Associated with the End-to-End Distance. To further understand effects of different solvent models on dynamical properties of peptides, we analyzed the dynamics associated with end-to-end fluctuations on short time scales. Feig²¹ considered C_{α} root-mean-square-deviation (rmsd) changes of proteins as a function of time to examine differences in kinetic properties based on solvent models, thermostats, and Langevin friction coefficients. In a similar spirit, we calculated mean square displacement of end-to-end distances as function of time to estimate relative magnitudes of diffusive motions for the solvent models studied here. As the diffusive motion of r_c is limited by the extent of the peptide, it is the short time behavior

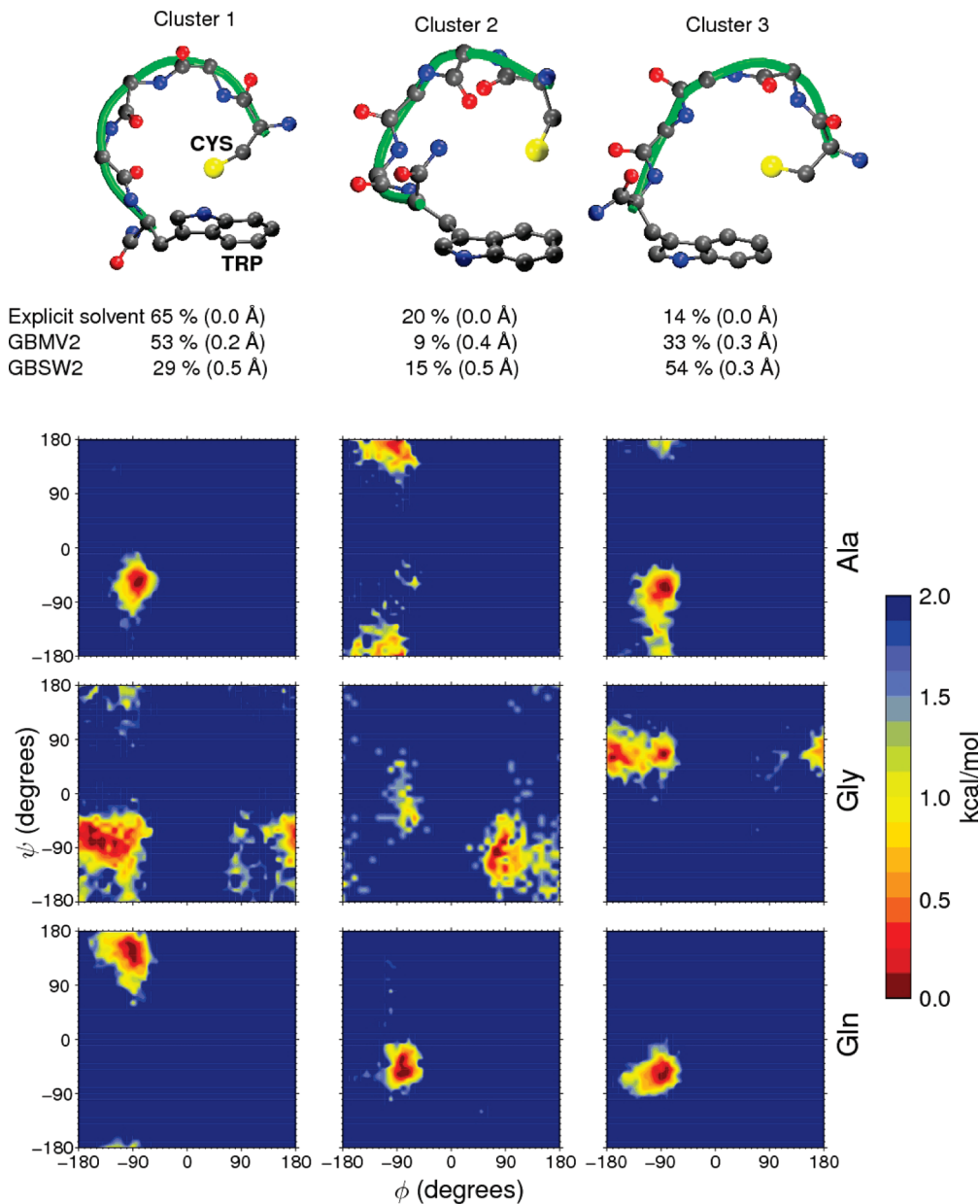


Figure 4. Representative structures of the three most populated clusters from the CAGQW structures that form end-to-end contact in explicit solvent simulations. The relative populations of the representative structures in each implicit solvent are also shown. The numbers in parentheses indicate backbone RMSDs between the average structure of a cluster as generated from the explicit simulation and the structures generated with implicit solvents. The potential of mean force (PMF) distributions of dihedral angles of the three central residues of the CAGQW peptide conformations that form end-to-end contacts are also shown. For each cluster, the data from explicit and implicit models are combined to prepare the PMF distributions of the dihedral angles.

that allows us to gauge the relative speeds of diffusive dynamics of the peptide.

Figure 6a shows mean square displacements of the end-to-end distance as a function of time from simulations in explicit solvent and with the GBSW2 implicit solvent model with different thermostats. The results obtained with the GBMV2 model using the same Langevin friction coefficient of 10 ps^{-1} showed virtually identical results, and for brevity we limit the following discussion to the GBSW2 model. As expected, the mean square displacement of the end-to-end distance is not linear with respect to the time interval even at short time scales as the internal diffusive motion of the peptide is not free but constrained by intramolecular covalent bonds. Therefore, a simple linear regression cannot be applied to obtain a diffusion coefficient as in the case of a free diffusion. However, the mean square displacement of the end-to-end distance at a short time

scale is expected to provide information on relative magnitudes of diffusive motions of end-to-end distances in different solvent models. For example, we can see clearly that the diffusive movement in explicit solvent is the slowest with a mean square displacement of 0.6 \AA^2 after 1 ps, the shortest recorded time difference. This is consistent with the slow end-to-end contact formation dynamics shown in Figure 5 and Tables 1 and 2 for the explicit solvent model.

We examined the dynamical behavior of the peptide as a function of the friction coefficient. Figure 6b shows the mean square displacement of the end-to-end distance after 1 ps as a function of the friction coefficient (we assigned a zero friction coefficients for simulations that did not employ Langevin dynamics). This clearly illustrated that the friction coefficient plays an important role in defining the internal diffusive dynamics of the peptide. Nosé–Hoover or Berendsen thermo-

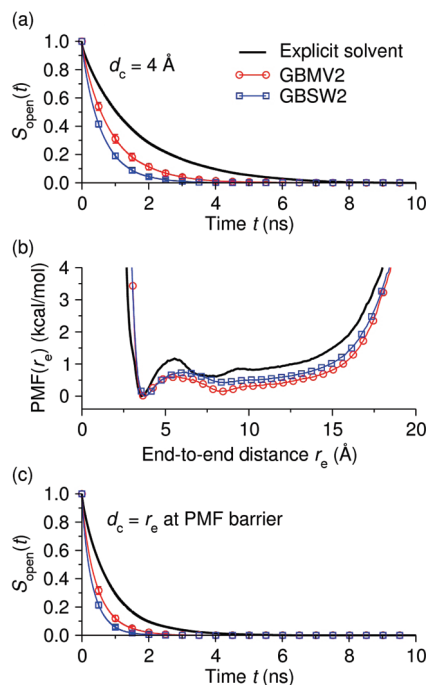


Figure 5. (a) Survival probability $S_{\text{open}}(t)$ of the open noncontact state calculated with explicit (a solid line), GBMV2 (a line with circles), and GBSW2 (a line with squares) solvent models. The implicit solvent calculations were done using a Langevin friction coefficient of 10 ps^{-1} . The contact distance of 4 \AA was used in the analysis. (b) The PMF distributions of end-to-end distances calculated from $-RT \log(P(r_e)/r_e^2)$, where $P(r_e)$ is the probability distribution of the end-to-end distance r_e as shown in Figure 2b. The PMF distributions were shifted along the y-axis so that minima near the contact distance coincide at 0.0 kcal/mol . (c) Same as in panel a but analyzed with different contact distances listed in Table 2, which correspond to r_e values at PMF barriers dividing open and closed states.

TABLE 1: Lifetimes of Open and Closed States for the Fixed Contact Distance of 4 \AA ^a

solvent model	temperature control	τ_{open} (ps)	τ_{closed} (ps)
explicit	Berendsen	1590 (188)	6.1 (1.4)
GBSW2	Nosé–Hoover	188 (22)	2.7 (0.3)
GBSW2	Berendsen	136 (9)	2.1 (0.2)
GBSW2	Langevin 2 ps^{-1}	213 (12)	1.8 (0.1)
GBSW2	Langevin 10 ps^{-1}	595 (43)	2.3 (0.1)
GBSW2	Langevin 60 ps^{-1}	3759 (1041)	4.5 (0.5)
GBMV2	Langevin 10 ps^{-1}	876 (80)	2.8 (0.2)

^a Errors in parentheses are based on the 95% confidence interval estimated by 1.96 times standard error. Standard errors were estimated by analysis of 50 independent simulations for implicit solvent models. For explicit solvent simulations with varying durations, standard errors were estimated by dividing simulations into 5 groups containing 10 simulations each. τ_{open} and τ_{closed} are average lifetimes of open and closed states, respectively. The results for the explicit solvent model in this and the following table were obtained from the previous simulation of the CACQW peptide in explicit solvent.³⁰

stats yielded relatively faster local diffusive motions with significantly larger mean square displacement at 1 ps than obtained with the explicit solvent. However, using these thermostats with implicit solvent models could lead to artifacts in calculated structural properties as shown in Figure 2a. In addition, Feig²¹ observed that the use of a Nosé–Hoover thermostat instead of Langevin dynamics greatly enhances local conformational sampling but slows down the crossing of barriers, which may be important in global sampling, by at least an order of magnitude, due to the lack of solute–solvent

TABLE 2: Lifetimes of Open and Closed States for Alternate Contact Distance Values^a

solvent model	d_c (Å)	τ_{open} (ps)	τ_{closed} (ps)	$\tau_{\text{closed}}/(\tau_{\text{open}} + \tau_{\text{closed}})$	f_c
explicit	5.6	835 (17)	60 (21)	0.067 (0.023)	0.069 (0.009)
GBSW2	6.1	322 (24)	11.6 (0.8)	0.035 (0.003)	0.067 (0.004)
GBMV2	5.7	451 (38)	12.8 (2.0)	0.028 (0.005)	0.044 (0.005)

^a The contact distance d_c in this table was defined as end-to-end distance corresponding to peak positions separating open and closed states in the PMF distribution shown in Figure 5b. Errors in parentheses are based on the 95% confidence interval estimated by 1.96 times standard error. Standard errors were estimated by analysis of 50 independent simulations for implicit solvent models. For explicit solvent simulations with varying durations, standard errors were estimated by dividing simulations into 5 groups containing 10 simulations each. f_c : fraction of contact states calculated from $P(r_e)$.

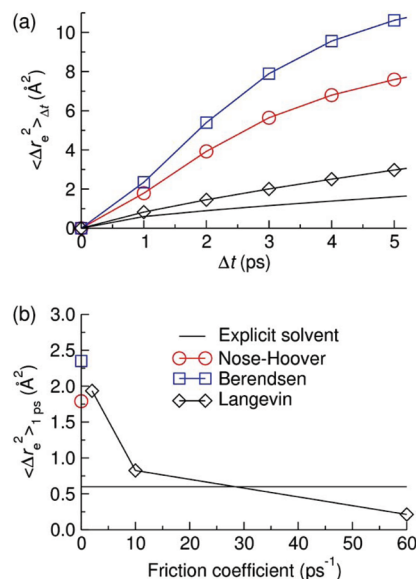


Figure 6. (a) Mean square displacement of end-to-end distance r_e between t and $t + \Delta t$ as calculated from the trajectories generated with the GBSW2 model. The explicit solvent values are indicated by the solid line. Lines with circles, squares, and diamonds represent results obtained from the implicit model using a Nosé–Hoover thermostat, a Berendsen thermostat, and Langevin dynamics with a friction coefficient of 10 ps^{-1} , respectively. (b) Mean square displacement of end-to-end distance r_e between t and $t + 1 \text{ ps}$ as a function of the friction coefficient. The same types of lines or symbols used in panel a have been adopted to represent results with different thermostats.

stochastic collisions. Figure 6b shows that a smaller Langevin friction coefficient of 2 ps^{-1} could achieve a similar degree of local conformational sampling obtained with Nosé–Hoover or Berendsen thermostats. As was observed in a recent simulation study with GBMV2,²¹ implicit-solvent simulations coupled with Langevin dynamics with an applied friction coefficient of 10 ps^{-1} give a reasonable agreement with the explicit solvent result. In a comparative study of generalized Born models, good agreements between calculated and experimentally derived structural properties were also obtained by simulations with the 10 ps^{-1} friction coefficient. On the basis of these considerations, Langevin dynamics with an applied friction coefficient of 10 ps^{-1} or smaller would be a reasonable choice for a thermostat in implicit-solvent simulations.

Figure 7a shows the mean displacement of the end-to-end distance after 1 ps as a function of its starting value. There is a good agreement between distributions of the mean displace-

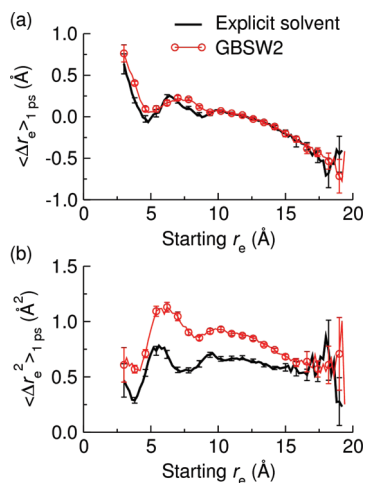


Figure 7. (a) Mean displacements associated with end-to-end distance r_e of the peptide between t and $t + 1$ ps as a function of starting r_e distance. A line with circles represents results with the GBSW2 implicit solvent model. The solid line represents results from simulations with explicit solvent. (b) Mean square displacements associated with end-to-end distance r_e of the peptide between t and $t + 1$ ps as a function of starting r_e distance.

ment calculated from explicit and GBSW2 implicit solvent models. At small and large starting end-to-end distances, the mean displacement is positive and negative, respectively, corresponding to a reflective motion in the bound potential. Deviation from zero mean displacement indicates that dynamics of end-to-end distances is influenced by conformation specific preferences rather than random diffusive motions. Except for large ($r_e > 15$ Å) or small ($r_e < 4$ Å) end-to-end distances where the population density is low, the mean displacement is close to zero, which indicates that internal dynamics on the shorter time scale of 1 ps is largely diffusive.

Figure 7b shows the mean square displacement of the end-to-end distance after 1 ps as a function of its starting value. As can be seen, the random diffusive motion associated with the end-to-end distance increases close to the barrier and slows near contact conformations of the peptide compared to the average value (Figure 6b). At the barrier ($r_e \sim 6$ Å), dynamics are accelerated as the peptide seeks a lower energy conformation by either closing or opening up the peptide. In the contact conformation ($r_e \sim 4$ Å), the peptide is energetically trapped and the motion restricted, which is reflected in the lowered mean square displacement. In general, the implicit model exhibits greater local variations than the explicit solvent model. This is compatible with the faster dynamics observed in the contact and noncontact lifetimes for the implicit models. Ultimately, the dynamics is governed by the potential energy landscape, and the faster dynamics exhibited by the implicit model are indicative of a less rugged energy landscape.²⁰

Summary and Conclusions

In this study, we investigated structural and dynamical properties of a small peptide CAGQW with different solvent models. Simulations with the GBSW2 and GBMV2 implicit solvent models showed similar distributions of end-to-end distances and dihedral angles as with the explicit solvent model. A close examination of end-to-end contact conformations of the peptide revealed that representative structures observed in explicit solvent were sampled by implicit solvent models but with different relative probabilities. These calculations were based on the same CHARMM22 force field. This is in contrast

with the small overlap (2% within 2 Å rmsd) between end-to-end contact conformations observed in explicit solvent simulations using two different force fields (AMBER94 and CHARMM22).³⁰

We observed significant differences in dynamical properties of peptides between simulations performed by explicit and implicit solvent models. The peptide in explicit solvent displayed the slowest dynamics in both contact formation and the intrinsic diffusive motion. The slow internal diffusive motion of the peptide in explicit solvent could reasonably be accounted for in implicit models by using Langevin dynamics and applying a friction coefficient of 10 ps^{-1} . However, details related to the energy landscape between the explicit and implicit model drive the observed difference in dynamical properties.

The study of the small peptide presented here allowed us in great detail and with high confidence to calculate the differences between explicit and implicit solvent models. Our study highlighted the similarities and differences that can be expected when using a generalized Born implicit solvent to model structural and dynamical properties of peptides and proteins. A more detailed study with larger peptides and diverse set of amino acid sequences would be needed for a thorough comparison of explicit and implicit solvent models. We noted that structural and dynamical properties obtained from implicit solvent simulations are quite sensitive to details of the calculation, such as the choice of thermostat. However, we also affirmed that simulations with GBMV2 or GBSW2 implicit solvent models coupled with Langevin dynamics with a friction coefficient near 10 ps^{-1} can give good agreements with explicit-solvent simulations.

Acknowledgment. We thank Dr. Michael S. Lee and Dr. Mark A. Olson for helpful suggestions and discussions. Funding support for this work came from the Department of Defense (DoD) High Performance Computing (HPC) Modernization Program Office under the HPC Software Applications Institute initiative, the U.S. Army Medical Research and Materiel Command. Computational time was provided by the U.S. Army Research Laboratory DoD Supercomputing Resource Center. The opinions or assertions contained herein are the private views of the authors and are not to be construed as official or as reflecting the views of the U.S. Army or of the U.S. Department of Defense.

References and Notes

- (1) Darden, T. A.; York, D. M.; Pedersen, L. G. *J. Chem. Phys.* **1993**, *98*, 10089.
- (2) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- (3) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141.
- (4) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. *J. Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749.
- (5) Kannan, S.; Zacharias, M. *Proteins* **2007**, *66*, 697.
- (6) Yeh, I.-C.; Olson, M. A.; Lee, M. S.; Wallqvist, A. *Biophys. J.* **2008**, *95*, 5021.
- (7) Felts, A. K.; Gallicchio, E.; Wallqvist, A.; Levy, R. M. *Proteins* **2002**, *48*, 404.
- (8) Im, W.; Lee, M. S.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1691.
- (9) Lee, M. S.; Salsbury, F. R.; Brooks, C. L. *J. Chem. Phys.* **2002**, *116*, 10606.
- (10) Lee, M. S.; Feig, M.; Salsbury, F. R.; Brooks, C. L. *J. Comput. Chem.* **2003**, *24*, 1348.
- (11) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.
- (12) Ghosh, A.; Rapp, C. S.; Friesner, R. A. *J. Phys. Chem. B* **1998**, *102*, 10983.
- (13) Gallicchio, E.; Levy, R. M. *J. Comput. Chem.* **2004**, *25*, 479.
- (14) Zhou, R. H.; Berne, B. *J. Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777.

- (15) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225.
- (16) Simmerling, C.; Strockbine, B.; Roitberg, A. E. *J. Am. Chem. Soc.* **2002**, *124*, 11258.
- (17) Okur, A.; Wickstrom, L.; Simmerling, C. *J. Chem. Theory Comput.* **2008**, *4*, 488.
- (18) Felts, A. K.; Harano, Y.; Gallicchio, E.; Levy, R. M. *Proteins* **2004**, *56*, 310.
- (19) Fan, H.; Mark, A. E.; Zhu, J.; Honig, B. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6760.
- (20) Hamelberg, D.; Shen, T. Y.; McCammon, J. A. *J. Chem. Phys.* **2006**, *125*, 7.
- (21) Feig, M. *J. Chem. Theory Comput.* **2007**, *3*, 1734.
- (22) Lapidus, L. J.; Eaton, W. A.; Hofrichter, J. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 7220.
- (23) Bieri, O.; Wirz, J.; Hellrung, B.; Schutkowski, M.; Drewello, M.; Kiefhaber, T. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9597.
- (24) Hudgins, R. R.; Huang, F.; Gramlich, G.; Nau, W. M. *J. Am. Chem. Soc.* **2002**, *124*, 556.
- (25) Roccatano, D.; Sahoo, H.; Zacharias, M.; Nau, W. M. *J. Phys. Chem. B* **2007**, *111*, 2639.
- (26) Huang, F.; Hudgins, R. R.; Nau, W. M. *J. Am. Chem. Soc.* **2004**, *126*, 16665.
- (27) Szabo, A.; Schulten, K.; Schulten, Z. *J. Chem. Phys.* **1980**, *72*, 4350.
- (28) Pastor, R. W.; Zwanzig, R.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 3878.
- (29) Zhou, H. X.; Szabo, A. *Biophys. J.* **1996**, *71*, 2440.
- (30) Yeh, I.-C.; Hummer, G. *J. Am. Chem. Soc.* **2002**, *124*, 6563.
- (31) Roccatano, D.; Nau, W. M.; Zacharias, M. *J. Phys. Chem. B* **2004**, *108*, 18734.
- (32) Ping, G.; Dastidar, S. G.; Duan, Y. *J. Chem. Phys.* **2007**, *126*, 045108.
- (33) Hamelberg, D.; de Oliveira, C. A. F.; McCammon, J. A. *J. Chem. Phys.* **2007**, *127*, 155102.
- (34) Feige, M. J.; Paci, E. *J. Mol. Biol.* **2008**, *382*, 556.
- (35) Doucet, D.; Roitberg, A.; Hagen, S. J. *Biophys. J.* **2007**, *92*, 2281.
- (36) Portman, J. J. *J. Chem. Phys.* **2003**, *118*, 2381.
- (37) 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.
- (38) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586.
- (39) Schuler, L. D.; Daura, X.; Van Gunsteren, W. F. *J. Comput. Chem.* **2001**, *22*, 1205.
- (40) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474.
- (41) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, *4*, 187.
- (42) Feig, M.; Karanickolas, J.; Brooks, C. L. *J. Mol. Graph. Model.* **2004**, *22*, 377.
- (43) Chocholoušová, J.; Feig, M. *J. Comput. Chem.* **2006**, *27*, 719.
- (44) Feig, M.; Onufriev, A.; Lee, M. S.; Im, W.; Case, D. A.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 265.
- (45) Lee, B.; Richards, F. M. *J. Mol. Biol.* **1971**, *55*, 379.
- (46) Nina, M.; Beglov, D.; Roux, B. *J. Phys. Chem. B* **1997**, *101*, 5239.
- (47) Chen, J. H.; Im, W. P.; Brooks, C. L. *J. Am. Chem. Soc.* **2006**, *128*, 3728.
- (48) Yeh, I.-C.; Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2008**, *112*, 15064.
- (49) Lee, M. S.; Olson, M. A. *J. Phys. Chem. B* **2008**, *112*, 13411.
- (50) Nosé, S. *J. Chem. Phys.* **1984**, *81*, 511.
- (51) Hoover, W. G. *Phys. Rev. A* **1985**, *31*, 1695.
- (52) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Di Nola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684.
- (53) Allen, M. P.; Tildesley, D. J. *Computer Simulations of Liquids*; Oxford University Press: New York, 1987.
- (54) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
- (55) Shao, J. Y.; Tanner, S. W.; Thompson, N.; Cheatham, T. E. *J. Chem. Theory Comput.* **2007**, *3*, 2312.
- (56) Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, I. T. E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. *Comput. Phys. Commun.* **1995**, *91*, 1.
- (57) Tuckerman, M. E.; Marx, D. *Phys. Rev. Lett.* **2001**, *86*, 4946.
- (58) Mor, A.; Ziv, G.; Levy, Y. *J. Comput. Chem.* **2008**, *29*, 1992.
- (59) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635.
- (60) Mongan, J.; Simmerling, C.; McCammon, J. A.; Case, D. A.; Onufriev, A. *J. Chem. Theory Comput.* **2007**, *3*, 156.

JP904064Z