Using free energy perturbation to predict effects of changing force field parameters on computed conformational equilibriums of peptides

Zanxia Cao and Haiyan Liu[a)]
School of Life Sciences and Hefei National Laboratory for Physical Sciences at the Microscale, University of Science and Technology of China (USTC), Hefei, Anhui 230027, People's Republic of China

(Received 24 March 2008; accepted 21 May 2008; published online 1 July 2008)

We propose a method that may allow data about the conformational equilibriums of peptides to enter the parameter calibration phase in force field developments. The method combines free energy perturbation with techniques for extensive sampling in the conformational space. It predicts shifts in computed conformational equilibriums in response to separate or combined perturbations of force field parameters. As an example we considered a force field associated with an implicit solvent model. We considered two different approaches to define conformational states of four peptides. One is based on reaction coordinates and two-dimensional free energy surfaces. The other is based on the clustering analysis of sampled conformations. Effects of perturbing various model parameters on the equilibriums between nativelike states with other conformational states were considered. For one type of perturbation predicted to have consistent effects on different peptides, the predictions have been verified by actual simulations using a perturbed model. © *2008 American Institute of Physics*. [DOI: 10.1063/1.2944248]

I. INTRODUCTION

For a biomolecular force field to predict correctly conformational equilibriums of various peptides and proteins,[1,2] an appropriate formulation of the potential energy function to capture all key aspects of the underlying physics is necessary but insufficient. Extensive parametrization is required to achieve quantitative descriptions of the involved complex interactions and their subtle balances. Primary targets of such parametrizations usually include a range of properties of small molecules in gas and in condensed phases. Calibrations exclusively based on such data, however, lead to several limits. First, the reference data available from experiments or from first-principles calculations are usually insufficient for deriving a unique set of parameters. Second, not all parameters important for conformational equilibriums of macromolecules can be well constrained by reproducing properties of small molecules. To constrain the parameters and their combinations by conformational equilibriums of peptides and proteins, repetitious and expensive trial-and-error simulations on different peptides and proteins are usually required

Here, we propose an approach to predict how conformational equilibriums change upon perturbations of force field parameters. The method combines extensive sampling in the conformational space and the free energy perturbation (FEP) theory[3] and may substantially reduce the amount of trial-and-error simulations. Potentially, the approach may also be employed to gain insights into how various physical interactions contribute to the conformational equilibriums of interest. To demonstrate the method, we consider a previously parametrized generalized Born/solvent accessible surface area (GBSA) model[4] for the GROMOS43A1 force field.[5] The model is applied to four peptides. Based on conformations sampled using temperature replica exchange molecular dynamics (T-REMD) simulations,[6–8] we predict how perturbations of various force field parameters would shift the computed equilibriums between the native conformational states and other conformational states of different systems. The predictions are verified by subsequent T-REMD simulations with perturbed parameters.

II. MATERIALS AND METHODS

A. Peptides, potential energy function, and sampling using T-REMD

We considered four peptides (noted as S_1,[9] S_2,[10], S_3,[11], and S_4,[12] see Table I) known to fold into different secondary structures. The peptides have been described by the GROMOS43A1 force field, with solvent treated by a GBSA model parametrized by Zhu *et al.*[13] Stochastic dynamics simulations[14] have been performed with an atomic frictional coefficient of 91 ps^{-1}, an integration time step of 2 fs, and covalent bond lengths constrained by SHAKE (Ref. 15) with a relative tolerance of 10^{-4}. Sixteen replicas have been simulated at temperatures of 200, 217, 236, 256, 279, 303, 330, 358, 390, 424, 461, 501, 544, 592, 643, and 700 K. Each replica had been equilibrated at its respective temperature for 100 ps. Then 20 ns T-REMD simulations were performed; replica exchanges were attempted every 0.1 ps based on the Metropolis criterion. Coordinates and energies have been recorded every 0.2 ps. The 16 trajectories spanning the last 12 ns have been analyzed.

[a)]Author to whom correspondence should be addressed. Tel.: (+86) 551-3607451. FAX: (+86) 551-3601443. Electronic mail: hyliu@ustc.edu.cn.

TABLE I. Peptides and definitions of native and near-native conformations based on the reaction coordinates.

Peptides	Sequences and native secondary structures	Native conformations[a]	Conformations of native state[b]	Conformations of near-native state
S peptide analog (S_1)	AETAAAKFLREHMDS	N_{HB}=7	$N_{HB} \in [6,7]$	$N_{HB} \in [4,7]$
	α-helix	R_g=7.5	$R_g \in (6.8,8.2)$	$R_g \in (6.4,8.6)$
E6-interacting peptide	IPESSELTLQELLGEERR	N_{HB}=6	$N_{HB} \in [5,6]$	$N_{HB} \in [3,6]$
(S_2)	α-helix	R_g=8.7	$R_g \in (8.0,9.4)$	$R_g \in (7.6,9.8)$
β-peptide (S_3)	GEWTYDDATKTFTVTE	N_{HB}=6	$N_{HB} \in [5,6]$	$N_{HB} \in [3,6]$
	β-hairpin	R_g=8.3	$R_g \in (7.6–9.0)$	$R_g \in (7.2–9.4)$
1le0 (S_4)	SWTWEGNKWTWK	N_{HB}=4	$N_{HB} \in [3,4]$	$N_{HB} \in [2,4]$
	β-hairpin	R_g=6.7	$R_g \in (6.0–7.4)$	$R_g \in (5.6–7.8)$

[a]N_{HB} is the number of native backbone-backbone hydrogen bonds. A hydrogen bond is counted if the distance between two atoms (H and O in this case) is less than 2.5 Å and the angle N—H···O is larger than 135.0°. R_g is the radius of gyration in Å.
[b]Conformations are assigned to the native state if they have at most one native hydrogen bond missing and have R_g's within a 0.14 nm window centered around the R_g of the corresponding native structures, and near-native state if they are not assigned to the native states and have at most half of the native hydrogen bonds missing and have R_g's within a 0.22 nm window centered around the R_g of the corresponding native structures.

B. Predicted shifts in conformational equilibriums upon perturbations of force field parameters using FEP

We refer the force field used for sampling as the reference model (represented by the Hamiltonian $H_{\text{reference}}$). A perturbed model ($H_{\text{perturbed}}$) refers to a force field with perturbed parameters. A conformational state *C* is defined as a particular region in the conformational space. According to the FEP theory, for a given state C, the free energy change associated with switching the Hamiltonian from the reference to the perturbed one is given by

$$\Delta F_C = F_{C,\text{perturbed}} - F_{C,\text{reference}} = -k_B T \ln\langle \exp^{-(H_{\text{perturbed}}-H_{\text{reference}})/k_B T}\rangle_{C,\text{reference}}, \quad (1)$$

in which $k_B T$ is the Boltzmann constant multiplied by the temperature and $\langle\cdots\rangle_{C,\text{reference}}$ refers to the average over the equilibrium distribution of conformations within state *C* under the reference Hamiltonian.

Although ΔF_C itself is artificial, a thermodynamic cycle can be constructed to predict how the computed equilibrium between different conformational states would be shifted by switching from $H_{\text{reference}}$ to $H_{\text{perturbed}}$,

$$\Delta\Delta F_{C_1,C_2} = \Delta F_{C_2} - \Delta F_{C_1} = -k_B T \ln \frac{\langle \exp^{-(H_{\text{perturbed}}-H_{\text{reference}})/k_B T}\rangle_{C_2,\text{reference}}}{\langle \exp^{-(H_{\text{perturbed}}-H_{\text{reference}})/k_B T}\rangle_{C_1,\text{reference}}}. \quad (2)$$

If $H_{\text{perturbed}}$ and $H_{\text{reference}}$ differ only in the value of a parameter p by an amount Δp, the following free energy derivative predicts how sensitive the computed equilibrium between C_1 and C_2 is to changes in p,

$$\frac{d\Delta F_{C_1,C_2}}{dp} \approx \frac{\Delta\Delta F_{C_1,C_2}}{\Delta p}. \quad (3)$$

We emphasize that Eqs. (2) and (3) contain averages over the equilibrium distributions of the reference model only; thus equilibrium sampling using the reference model suffices to predict the effects of different perturbations.

C. Defining different conformational states based on reaction coordinates and free energy surfaces

We construct free energy surfaces (FESs) based on reaction coordinates (RCs) commonly used in peptide folding studies. The two RCs are the radius of gyration (R_g) and the number of native backbone-backbone hydrogen bonds in secondary structures excluding turns (N_{HB}). The corresponding FESs indicated that except for S_1, the native states do not correspond to the lowest minima in the two dimensional spaces. The native states being too high on the reference FESs, sampling distributions in the native and near-native states by brutal-force extensions of the simulations are impractical. We chose to modify the reference model so that sufficient sampling of near-native and native conformations could be obtained for averaging. The modifications are additional half-side harmonic potentials with a force constant of 1000 kJ mol^{-1} nm^{-2}. These potentials restrain the distances between oxygen and hydrogen atoms forming hydrogen bonds in the native states to be within length of 2.5 Å. Two dimensional FESs using a modified $H_{\text{reference}}$ containing the restraining potentials were also constructed. We note that in Eq. (2), the distributions within different states are normalized separately and independently; the restraining potentials would have minor effects on the computed relative free energy changes if they do not disturb the internal distributions of conformations within the native and near-native states. Based on the unrestrained and/or restrained FESs, native, near-native, and non-native states have been defined using the RC ranges given in Table I.

D. Defining different conformation states based on clustering analysis of sampled conformations

Although the above approach of using two dimensional RCs to visually represent the conformational space of peptides has been a simple and widely used way, unavoidably, the free energy contour maps depend on the RCs. Another way to partition sampled conformations into different states is to cluster them based on their mutual root-mean-square deviations of Cα positions ($\text{RMSD}_{C\alpha}$). For each peptide, we consider conformations sampled by the replica at 303 K. A total of 15 000 conformations from the last 12 ns trajectory

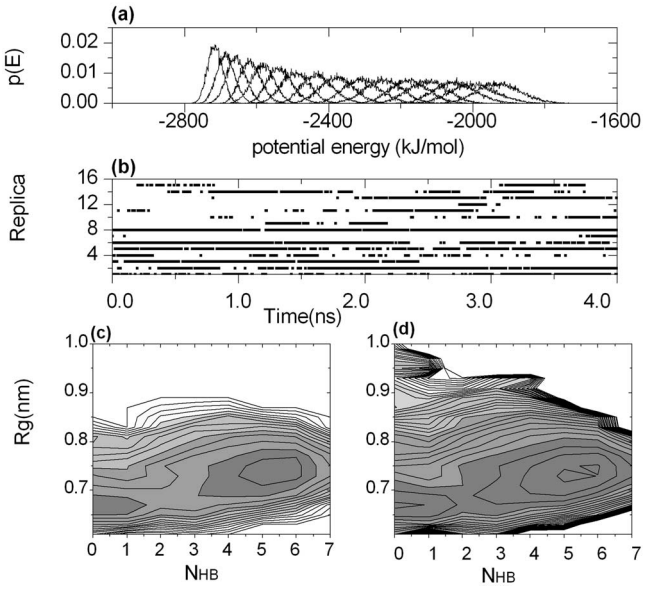

FIG. 1. Results for the T-REMD simulation and T-WHAM of S_1 are shown here. (a) The distributions of total potential energies at different temperatures. (b) The code (a number between 1 and 16) of the trajectory exchanged to the 303 K replica at different time points of the simulation. (c) A two dimensional FES at 303 K constructed using data of a single replica. (d) A two dimensional FES at 303 K constructed using T-WHAM.

(1 conformation every 0.8 ps) were clustered based on their pairwise $\mathrm{RMSD}_{C\alpha}$. Only residues in native secondary structures have been compared. The criteria for clustering have been that for any conformation in a cluster, there is at least one other conformation in the same cluster with a $\mathrm{RMSD}_{C\alpha}$ less than 0.1 nm from the conformation, and there should be no conformation in any other cluster with a $\mathrm{RMSD}_{C\alpha}$ less than 0.1 nm. In addition, all conformations in the same cluster should be connected by the $\mathrm{RMSD}_{C\alpha}$ criterion.

III. RESULTS AND DISCUSSIONS

A. T-REMD simulations and T-WHAM

Effective T-REMD requires sufficient overlaps between the energy distributions at different temperatures. Figure 1(a) shows for S_1 the overlaps between the energy distributions of neighboring replicas. The ratios of successful exchange attempts were between 25% and 40%. Figure 1(b) shows that configurations from all of the 16 continuous trajectories contributed to the replica at 303 K. Results for other peptides are similar. Figures 1(c) and 1(d) shows the FESs for S_1 at 303 K, constructed by analyzing a single replica and by the temperature-weighted histogram analysis (T-WHAM) approach,[16] respectively. For regions associated with lower free energies, the two FESs are essentially the same, while for regions of higher free energies, the T-WHAM results provided more details. The percentages of native and near-native states (defined by criteria in Table I) obtained by analyzing a single replica are $23 \pm 1\%$ and $68 \pm 6\%$, respectively, and by applying T-WHAM are $21 \pm 1\%$ and $65 \pm 6\%$, respectively (to estimate the statistical errors, the total 12 ns trajectory set has been evenly divided into three 4 ns blocks and the standard deviation between averages over individual blocks have been computed, which are the same below). Results for other peptides produced similar comparisons and we used T-WHAM in later analyses.

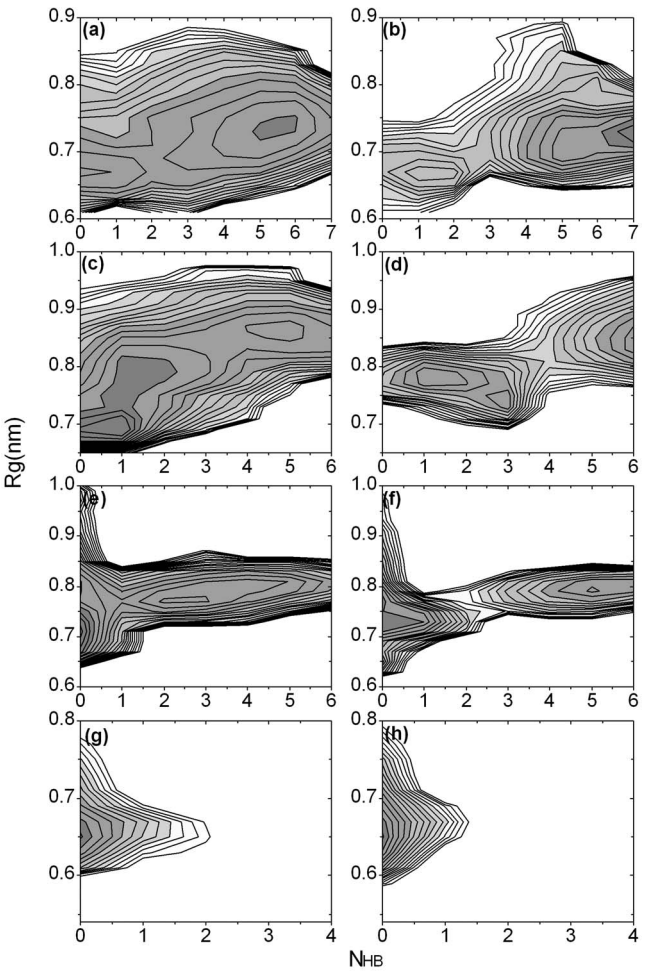

FIG. 2. FESs constructed from unrestrained T-REMD simulations, contoured at $0.5k_BT$ intervals. Darker regions correspond to lower free energies. (a), (c), (e), and (g) correspond to results using the reference model for peptides S_1, S_2, S_3, and S_4, respectively. (b), (d), (f), and (h) correspond to the same results using a model with increased charges on polar backbone atoms.

B. FESs obtained using the reference model

For S_1 [Fig. 2(a)], the native state corresponds to the global minimum. For the other three systems, this is, however, not the case. For S_2 [Fig. 2(c)], the FES shows two distinctive minima: one corresponding to the native state (with an R_g of ~8.5 Å and 5 native hydrogen bonds) and the other to collapsed random structures. For S_3 [Fig. 2(e)] and S_4 [Fig. 2(g)], the FESs have no minimum close to the respective native states. Table II also shows that for these peptides, unrestrained simulations could barely or not sample in the native or near-native regions.

The restrained FESs are shown in Fig. 3. For S_2 [Fig. 3(a)], there are two minima: one is the native state with a free energy of 0.8 kJ mol^{-1} relative to the non-native minima. For each of the two β-hairpin forming peptides, there is only a single native minimum [Figs. 3(c) and 3(e)].

TABLE II. Percentages of native state conformations (%), and both native and near-native state conformations (data in parentheses), among all sampled conformations.

Simulations	S_1[a,b]	S_2[a,b]	S_3[a,b]	S_4[a,b]
Unrestrained, reference model	$21\pm1(65\pm6)$	$12\pm5(28\pm5)$	$4\pm0(18\pm4)$	0(0)
	$23\pm1(67\pm4)$	$11\pm5(25\pm3)$	$4\pm0(16\pm4)$	0(0)
Unrestrained, perturbed model	$63\pm2(94\pm6)$	$38\pm4(45\pm5)$	$15\pm1(20\pm4)$	0(0)
	$63\pm1(94\pm5)$	$38\pm5(57\pm4)$	$15\pm1(20\pm4)$	0(0)
Restrained, reference model	···	$26\pm3(54\pm1)$	$51\pm2(97\pm0)$	$80\pm0(97\pm0)$
	···	$44\pm3(70\pm5)$	$53\pm1(97\pm0)$	$80\pm0(97\pm0)$
Restrained, perturbed model	···	$83\pm2(90\pm1)$	$71\pm2(99\pm0)$	$93\pm0(99\pm0)$
	···	$88\pm2(89\pm5)$	$83\pm1(99\pm0)$	$93\pm0(99\pm0)$

[a]Data in the first row for each type of simulations correspond to definitions of different states based on N_{HB} and R_g. Data in the second row correspond to definitions based on N_{HB} and $RMSD_{C\alpha}$.
[b]The error ranges are standard deviations between three 4 ns trajectory blocks.

Another widely used RC in literature has been the $RMSD_{C\alpha}$ from a reference structure (usually the native one). We have analyzed the sampled conformations using this RC as well. In general, $RMSD_{C\alpha}$ correlates well with N_{HB}, and the FESs in the $RMSD_{C\alpha}$ and N_{HB} space are qualitatively the same as those in the R_g and N_{HB} space.

For S_3 and S_4, although the non-native states may correspond to minima on the unrestrained FESs, the constituting conformations distribute diversely in the conformational space. If they were clustered based on $RMSD_{C\alpha}$, no outstandingly large cluster could be found.

We also analyzed the distributions of the backbone (φ,ψ) angles for each peptide (Fig. 4, data for different residues contained in native secondary structure have been pooled together). Interestingly, despite the fact that the reference model failed to reproduce the β-hairpin native states as minima on the respective FESs for S_3 and S_4, it does lead to significantly more sampling in the β regions for these two peptides relative to the helical peptides.

It is worth mentioning that the peptides we considered as examples have been simulated by others for other purposes such as to reveal mechanisms of folding. Besides many performed using explicit solvent models, at least two studies have employed GBSA models associated with other force fields. For the S_3 peptide, Zhou and Berne[17] have reported a FES under GBSA, which is quite similar to the results obtained here. Yang *et al.*[18] have analyzed the folding of S_4 peptide in implicit solvent. They found that the averaged number of native hydrogen bond is only about 2 at 300 K.

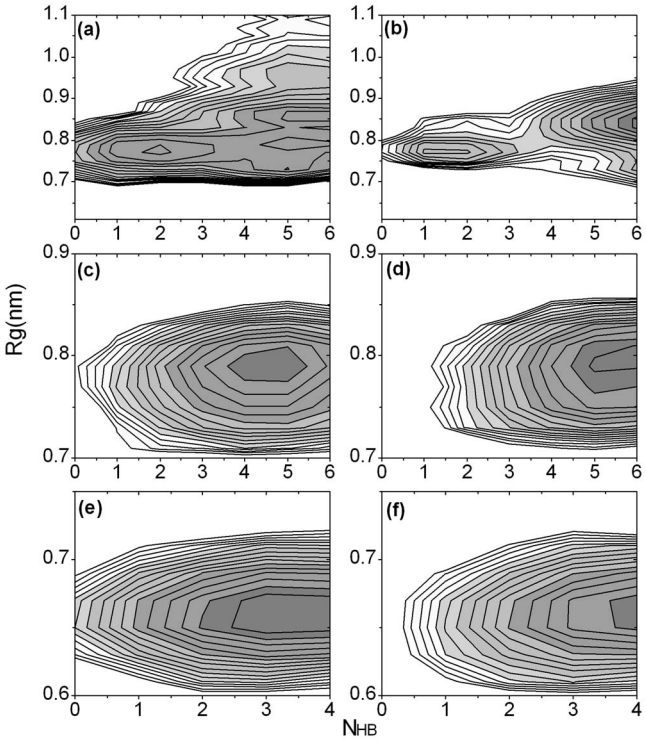

FIG. 3. As in Fig. 2, except that the FESs have been constructed from the restrained T-REMD sampling, with (a), (c), and (e) corresponding to simulations using the reference model for S_2, S_3, and S_4, respectively, and (b), (d), and (f) corresponding to the same results using a model with increased charges on polar backbone atoms.

C. Conformation clusters

By the clustering criteria, conformations sampled in the unrestrained simulations fall into clusters of varying sizes (see Fig. 5). For S_1, 42 clusters have been obtained. Among them, 3 contain at least 500 conformations. One such cluster contains at least one conformation with a $RMSD_{C\alpha}$ less than 0.1 nm from the native structure, and this cluster contains 92% conformation in total. For S_2, 17 clusters have been obtained. Among them, 2 contain at least 500 conformations. One such cluster contains at least one conformation with a $RMSD_{C\alpha}$ less than 0.1 nm from the native structure, and this cluster contains 50% conformation in total. For S_3, 263 clusters have been obtained. Among them, 4 contain at least 500 conformations. One such cluster contains at least one conformation with a $RMSD_{C\alpha}$ less than 0.1 nm from the native structure, and this cluster contains 16% conformation in total. For S_4, 176 clusters have been obtained. Among them, 5 contain at least 500 conformations. No cluster contains at least one conformation with a $RMSD_{C\alpha}$ less than 0.1 nm from the native structure. Conformations sampled in the restrained simulations of S_2, S_3, and S_4 form single clusters (see Fig. 5), and the respective native structures can be assigned to these clusters by the $RMSD_{C\alpha}$ criterion.

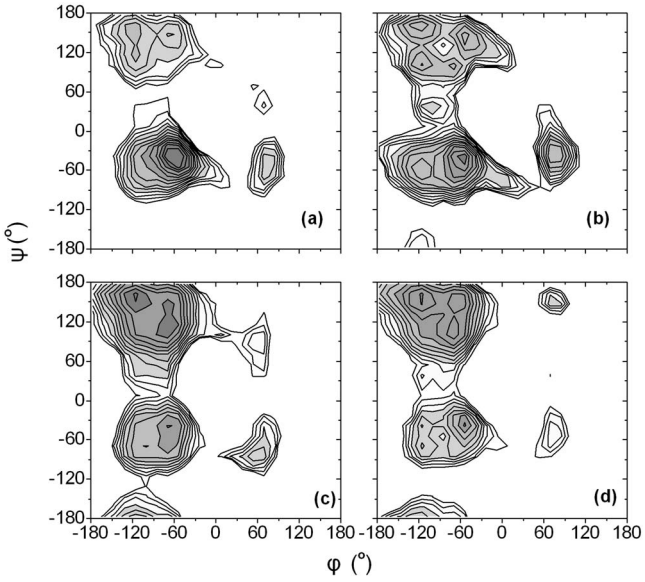

FIG. 4. Potentials of mean forces derived from the backbone (φ, ψ) distributions sampled in the unrestrained T-REMD simulations using the reference model. The contours are at $0.5k_BT$ intervals, and darker regions correspond to lower free energies. (a)–(d) correspond to results for S_1–S_4, respectively. For each peptide, data for residues contained in the native secondary structure (residues 4–11, 5–12, 2–6, and 11–15, and 2–4 and 9–11 for peptides S_1, S_2, S_3, and S_4, respectively) have been pooled together.

D. Predicted shifts in conformational equilibriums in response to perturbation of force field parameters according to RCs and FESs

The derivatives of free energy differences in Eq. (3) with respect to the five parameters P_1–P_5 in the GBSA model, the van der Waals radii of three backbone atom types (R_C, R_N, and R_{CH}), and the sizes of partial charges on the backbone N—H and C=O groups have been computed, and $\Delta p = 0.02p$. Figure 6 indicates that perturbing the GB parameters and van der Waals parameters usually has small effects on the conformational equilibriums of interest. Exceptions are the effects of changing P_1 on the conformational equilibriums of S_2 and S_3. These effects are, however, mutually inconsistent; perturbing P_1 in one direction could stabilize the native and near-native states of one peptide relative to the non-native one but could then destabilize those of the other peptide.

Figure 6 suggests that perturbing the partial charges produced effects consistent across different peptides. The free energy derivatives indicate that increasing the absolute partial charges would stabilize the native states relative to the near-native states or near-native states relative to the non-native states.

We then considered simultaneous perturbations of these charges. For increasing the absolute charges by 20%, applying Eq. (2) predicted that the computed relative free energies of native states relative to near-native states would be lowered by -5.2 ± 0.6, -0.2 ± 0.1, -2.9 ± 0.2, and -3.2 ± 0.2 kJ mol^{-1}, and of near-native states relative to non-native states would be lowered by -7.4 ± 1.6, -2.5 ± 1.0, -4.5 ± 0.6, and -1.6 ± 0.2 kJ mol^{-1} for S_1, S_2, S_3, and S_4, respectively.

If we replace the R_g criteria in Table I by $\mathrm{RMSD}_{C\alpha}$ < 0.2 nm for the native state and $\mathrm{RMSD}_{C\alpha} < 0.3$ nm for the near-native state, applying Eqs. (2) and (3) produced similar results.

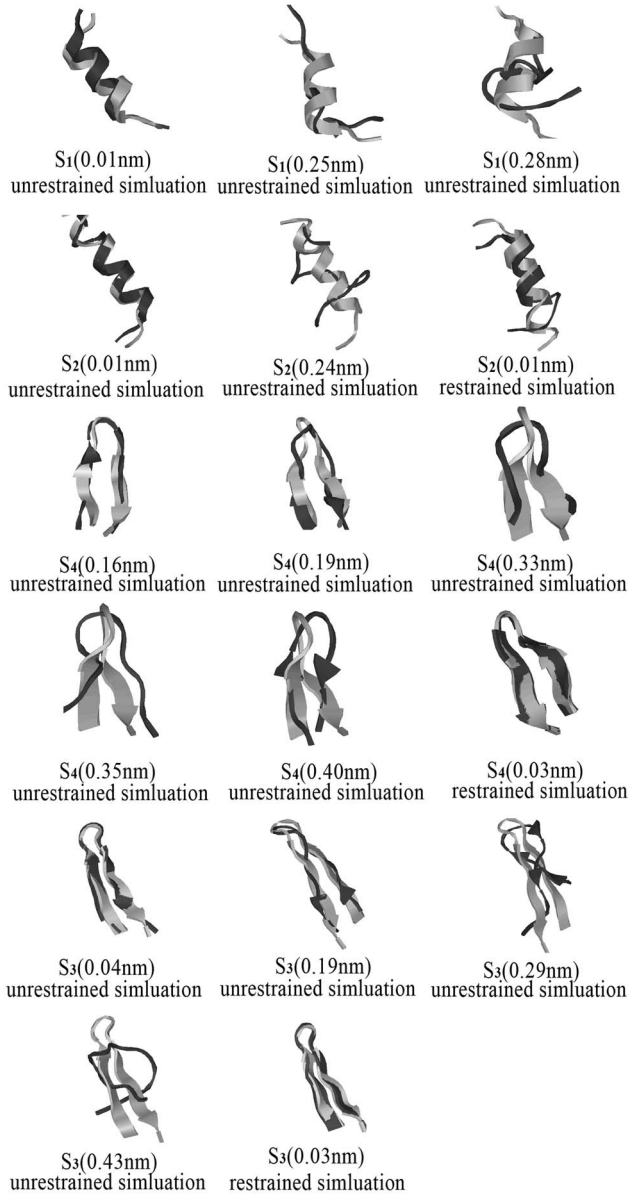

FIG. 5. The conformations with the minimum $\mathrm{RMSD}_{C\alpha}$ from respective native structures in different conformation clusters. The corresponding peptide systems and minimum $\mathrm{RMSD}_{C\alpha}$ are labeled below the images. The native structures are shown in gray: the conformations with the minimum $\mathrm{RMSD}_{C\alpha}$ in different conformation clusters are shown in black. Only large clusters with more than 500 conformations have been shown.

For S_2, S_3, and S_4, the restraining potentials may have a potential effect of distorting the distributions, especially for the near-native states. We considered an alternative partitioning of conformational states, in which only two conformational states are considered for each peptide, with the native one defined as before and the non-native one containing all remaining conformations. Conformations sampled in the unrestrained simulations have been used for averaging over the non-native states. Applying Eqs. (2) and (3) still predicts that increasing the absolute partial charges on the peptide backbone atoms would stabilize the native states relative to the

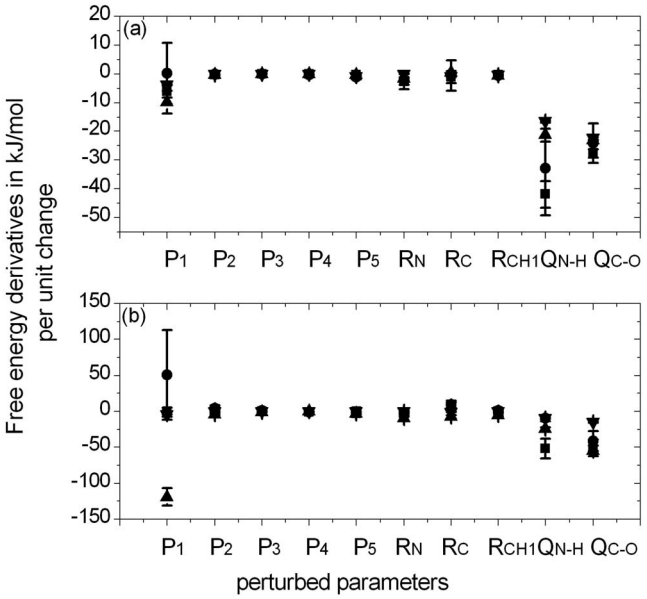

FIG. 6. Derivatives of relative conformational free energies with respect to perturbations of different parameters computed using Eq. (3); the native, near-native, and non-native states have been defined using the RC ranges given in Table I. (a) Derivatives of the free energies of native states relative to near-native states; (b) derivatives of the free energies of near-native states relative to non-native states. Different symbols represent different peptides, with squares for S_1, circles for S_2, up triangles for S_3, and down triangles for S_4. Error bars represent standard deviations between results obtained using different 4 ns trajectory blocks.

redefined non-native states of all four peptides, while adjusting other parameters would have either small or mutually inconsistent effects.

E. Predicted shifts in conformational equilibriums according to conformation clusters

For each peptide, we treated each cluster containing at least 500 sampled conformations as an independent conformation state and predicted how the equilibriums between them would be shifted. For each cluster, the derivatives of free energy differences in Eq. (2) with respect to the ten model parameters have been computed. As we have more than two states, we have considered a virtual C_1 state with a minimum ΔF in computing the free energy derivatives. Still, only the relative values between different states are of physical meaning. In Fig. 7, we show the free energy derivatives of different clusters versus the minimum $\mathrm{RMSD}_{C\alpha}$ of their constituent conformations from the respective native structures. For S_2, S_3 and S_4, data corresponding to the single cluster from the restrained simulations are also shown.

The results in Fig. 7 are consistent with the predictions for conformational states defined based on RCs. Figures 7(a)–7(h) indicate that according to FEP, perturbing any of the GB or van der Waals parameters would not produce consistent effects on the equilibriums between clusters containing nativelike members (i.e., with small $\mathrm{RMSD}_{C\alpha}$ from the respective native structures) and clusters containing no nativelike members (i.e., with large $\mathrm{RMSD}_{C\alpha}$ from the respective native structures). For different peptides, either types of conformational clusters can be relatively stabilized or destabilized by such perturbations.

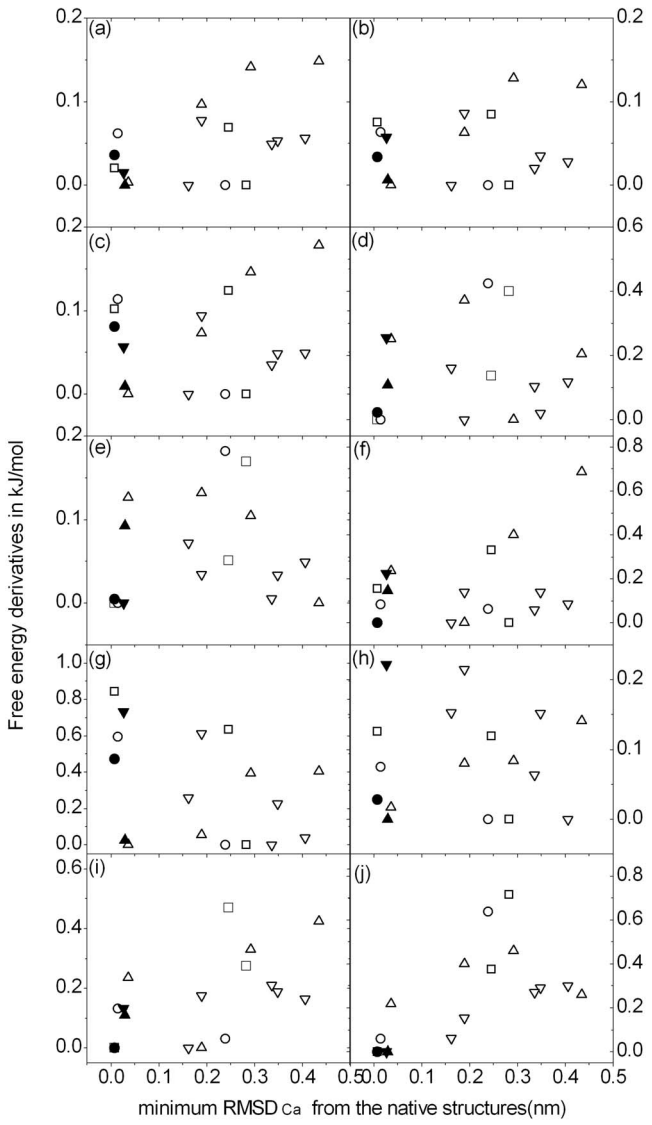

FIG. 7. Free energy derivatives of conformational states cooresponding to conformation clusters vs the minimum $\mathrm{RMSD}_{C\alpha}$ (in nm) of the constituent conformations from the respective native structures. Each point corresponds to a cluster. Open symbols: Simulations using unrestrained reference model. Filled symbols: Simulations using restrained reference model. As in Fig. 6, different symbols represent different peptides. (a)–(j) correspond to results with respect to the ten parameters P_1, P_2, P_3, P_4, P_5, R_N, R_C, R_{CH1}, $Q_{N—H}$ and $Q_{C—O}$, respectively.

From Fig. 7(i) we can predict that increasing the absolute partial charges on backbone N—H would tend to stabilize clusters containing nativelike members relative to other clusters. For S_1, the large cluster having the lowest minimum $\mathrm{RMSD}_{C\alpha}$ from its native structure would be strongly stabilized relative to the other two large clusters, both having minimum $\mathrm{RMSD}_{C\alpha}$ between 0.2 and 0.3 nm. Other clusters contain less than 500 conformations. For S_2, there are two large clusters having minimum $\mathrm{RMSD}_{C\alpha}$ less than 0.1 nm from its native structure. The remaining large cluster has a minimum $\mathrm{RMSD}_{C\alpha}$ of 0.24 nm. The equilibriums would be shifted in favor of the cluster having the lowest minimum $\mathrm{RMSD}_{C\alpha}$, although the other cluster having a minimum $\mathrm{RMSD}_{C\alpha}$ less than 0.1 nm would be disfavored relative to the cluster of 0.24 nm minimum $\mathrm{RMSD}_{C\alpha}$. For both S_3 and

S_4, the equilibriums would be shifted in the most favor of two larger clusters, both having minimum $\mathrm{RMSD}_{C\alpha}$ below 0.2 nm. The clusters of the lowest minimum $\mathrm{RMSD}_{C\alpha}$ would be destabilized relative to these two clusters, but would still be stabilized relative to other large clusters having larger minimum $\mathrm{RMSD}_{C\alpha}$.

To estimate the overall effects of perturbing the N—H charges, we predicted how the equilibriums between all clusters having less than 0.2 nm minimum $\mathrm{RMSD}_{C\alpha}$ from native structure and all other clusters would be shifted. For the four peptides S_1–S_4, the predicted $\Delta\Delta F/\Delta p$ are −72, −22, −38, and −3 kJ mol^{-1}, respectively, all in favor of the aggregated clusters containing nativelike structures.

Figure 7(j) shows that for all four peptides, increasing the absolute charge on the C—O atoms would consistently stabilize clusters containing nativelike members relative to other clusters. If the equilibriums between all clusters having less than 0.2 nm minimum $\mathrm{RMSD}_{C\alpha}$ from native and all other clusters are considered, the predicted $\Delta\Delta F/\Delta p$ are −81, −80, −50, and −21 kJ mol^{-1} for the four peptides S_1–S_4, respectively, all in favor of the aggregated clusters containing nativelike structures.

F. Test the predictions by simulations with perturbed parameters

We reperformed the T-REMD simulations with an adjusted model, in which the absolute charges on backbone N, H, C, and O atoms have been increased by 20%. The new conformational equilibriums relative to the original model verified the FEP predictions. For S_1, the native state minimum on the new FES [Fig. 2(b)] is now −7.7 kJ mol^{-1} relative to the non-native minimum, lower than the −1.1 kJ mol^{-1} obtained using the reference model [Fig. 2(a)]. For S_2, the native minimum was shifted to a larger N_{HB}, with a free energy of 0.1 kJ mol^{-1} relative to the non-native minimum [Fig. 2(d)], as compared with the 5.2 kJ mol^{-1} obtained using the reference model [Fig. 2(c)]. For S_3, the perturbed model produced the native minimum, although it is still 3.0 kJ mol^{-1} higher than the non-native minimum [Fig. 2(f)], while the unperturbed reference model still generated no minimum corresponding to the native state [Fig. 2(e)].

Restrained simulations have also been performed using the perturbed model on S_2, S_3, and S_4. For S_2, the minimum corresponding to the native state is −6.4 kJ mol^{-1} relative to the non-native minimum [Fig. 3(b)] as compared with the 0.8 kJ mol^{-1} in the unperturbed FES [Fig. 3(a)]. The single minima on the FESs for S_3 [Fig. 3(d)] and S_4 [Fig. 3(f)] are shifted to larger N_{HB} values and deeper as compared with the reference FESs [Figs. 3(c) and 3(e), respectively], indicating more stable native states under the perturbed model.

We note that although the FEP predictions based on simulations using $H_{\mathrm{reference}}$ alone and the above changes of the FESs are of the same sign, the actual values may differ significantly. The following thermodynamic cycle is implicated in comparisons between the FEP predictions and results of actual simulations. It may provide a starting point to explore possible origins of these differences,

$$\begin{array}{ccc} C_{1,H_{\mathrm{reference}}} & \overset{\Delta F_{12,H_{\mathrm{reference}}}}{\rightleftharpoons} & C_{2,H_{\mathrm{reference}}} \\ \Delta F_{C_1}\downarrow\uparrow & & \Delta F_{C_2}\downarrow\uparrow \\ C_{1,H_{\mathrm{perturbed}}} & \overset{\Delta F_{12,H_{\mathrm{perturbed}}}}{\rightleftharpoons} & C_{2,H_{\mathrm{perturbed}}} \end{array} \tag{4}$$

in which the free energy differences associated with the vertical changes are those defined in Eq. (1). The free energy differences associated with the horizontal changes represent equilibriums between difference conformations under different Hamiltonians. Ideally, we should have

$$\Delta\Delta F_{C_1,C_2} \equiv \Delta F_{C_2} - \Delta F_{C_1} \equiv \Delta F_{12,\mathrm{perturbed}} - \Delta F_{12,\mathrm{reference}}. \tag{5}$$

We note that two possibilities, either separately or jointly, can lead to deviations from the above equality. The first is that the changes in the Hamiltonian from $H_{\mathrm{reference}}$ to $H_{\mathrm{perturbed}}$ might have been too large to allow the meaningful use of the FEP equation [Eq. (2)], which is well known to have convergence problems when the change in the Hamiltonian is large. The second is that there are too large sampling errors in $\Delta F_{12,\mathrm{perturbed}}$ and $\Delta F_{12,\mathrm{reference}}$.

To test whether perturbing the backbone charges by 20% has exceeded the convergence limit of Eq. (2), we recomputed the vertical free energy changes from simulations using the perturbed Hamiltonian by applying FEP in the backward direction,

$$(\Delta\Delta F_{C_1,C_2})_{\mathrm{backward}} = -k_B T \ln \frac{\langle \exp^{-(H_{\mathrm{reference}}-H_{\mathrm{perturbed}})/k_B T}\rangle_{C_1,\mathrm{perturbed}}}{\langle \exp^{-(H_{\mathrm{reference}}-H_{\mathrm{perturbed}})/k_B T}\rangle_{C_2,\mathrm{perturbed}}}. \tag{6}$$

These backward FEP results together with the estimations using $\Delta F_{12,\mathrm{perturbed}}$ and $\Delta F_{12,\mathrm{reference}}$ have been plotted against the forward FEP results in Fig. 8.

We note that the forward and backward FEP results have been separately obtained using two mutually independent sets of conformation samples; if there had been convergence problems in either the forward or the backward perturbation, there would be no agreement between the forward and the backward perturbation results. Figure 8 shows that the forward and backward FEP results agree reasonably well, with a mean square difference of 1.1 kJ/mol, and are much better than their comparisons with $\Delta F_{12,\mathrm{perturbed}} - \Delta F_{12,\mathrm{reference}}$.

In fact, the accurate determination of the free energies $\Delta F_{12,\mathrm{perturbed}} - \Delta F_{12,\mathrm{reference}}$ requires sufficient sampling of transitions between different conformational states (here the native, near-native, and non-native states) under each Hamiltonian, which are very difficult to achieve even with the T-REMD extended sampling protocol. This is especially true if one of the states should be of very high relative free energy, as it would be rarely sampled. On the other hand, convergence of the vertical changes mainly depends on how well equilibrium sampling within different conformational states has been achieved and to what extent the within-state distributions associated with the reference and the perturbed Hamiltonians are similar.

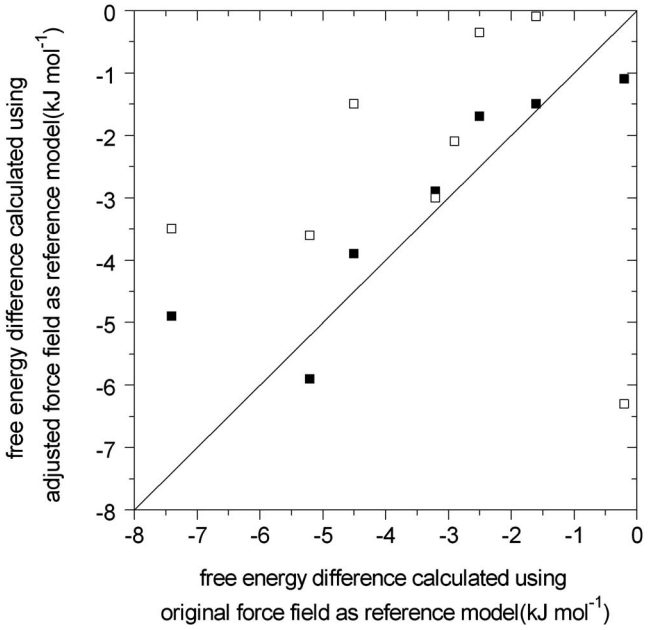

FIG. 8. The $\Delta\Delta F_{C_1,C_2}$ obtained by applying the backward FEP to simulations using the perturbed Hamiltonian (filled symbols) and estimations of the same free energy changes from the FESs of the reference and the perturbed models (open symbols) are plotted against the forward FEP results. Different points correspond to different peptides and different pairing of conformational states.

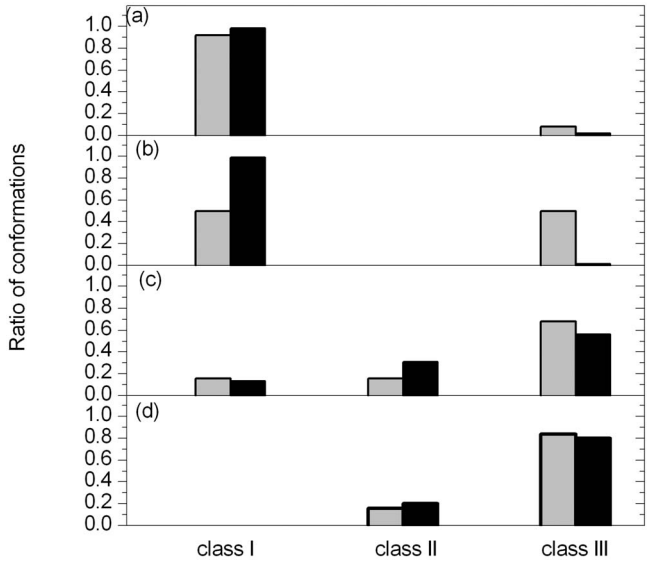

FIG. 9. Percentages of conformations contained in different classes of clusters. Gray: Simulations using the unrestrained reference model. Black: Simulations using unrestrained perturbed model. Class I includes all clusters having less than 0.1 nm minimum $\mathrm{RMSD}_{C\alpha}$ from the respective native structures. Class II includes all clusters having minimum $\mathrm{RMSD}_{C\alpha}$ from 0.1 to 0.2 nm. Class III includes all clusters having minimum $\mathrm{RMSD}_{C\alpha}$ larger than 0.2 nm. (a)–(d) correspond to results for peptides S_1, S_2, S_3, and S_4, respectively.

Other quantitative indicators of shifts in conformational equilibriums are the ratios of conformations in native and near-native states in the sampled ensembles. Table II compared results from simulations using the reference and perturbed models. The results are again fully consistent with the FEP predictions.

For conformational states defined based on clustering, it is not possible to establish one-to-one correspondences between clusters sampled using the reference model and using the perturbed models. Thus straightforward comparisons between the test simulations and the predictions in Figs. 7(i) and 7(j) are difficult. To facilitate comparisons without establishing correspondences between conformation clusters sampled by different simulations, we divide the conformational clusters into three classes: the first class includes all clusters having less than 0.1 nm minimum $\mathrm{RMSD}_{C\alpha}$ from the native structure, the second class includes all clusters having minimum $\mathrm{RMSD}_{C\alpha}$ from the native structure between 0.1 and 0.2 nm, and the third class includes all clusters having larger than 0.2 nm minimum $\mathrm{RMSD}_{C\alpha}$ from the native structure. Similar to the clustering analyses on simulations using the reference model, conformations sampled by unrestrained simulations using the perturbed model have been clustered. Figure 9 compared the percentages of conformations contained in each of the three classes of clusters in simulations using the unrestrained reference model and using the unrestrained perturbed model.

Figures 9(a) and 9(b) show that for S_1 and S_2, the perturbation significantly increases the ratios between conformations in class I clusters and in class III clusters. These are consistent with the predicted effects on these two peptides shown in Figs. 7(i) and 7(j). For S_3, Fig. 9(c) shows a large increase in the fraction of conformations in class II clusters. Figure 7(i) predicts that increasing the charge on N—H would be in favor of a cluster of such class relative to other classes. Figure 9(d) indicates a slight increase in the fraction of class II clusters for S_4. Thus results in Fig. 9 are in general consistent with FEP predictions in Figs. 7(i) and 7(j).

IV. CONCLUSIONS

Force field developments have relied on calibrations based on small molecule systems and repetitious trial-and-error test simulations of larger molecules. We have proposed an approach that combines FEP with improved sampling techniques, which may allow for experimental data on larger systems, especially the large amount of data on conformational equilibriums of different peptides, to enter the calibration phase in force field development as an additional dimension of restraints.

Although the FEP approach may have convergence problems when changes in the Hamiltonian are large, our results show that for all our test systems, changing the backbone charges by 20% does not exceed the convergence limits of Eq. (2) in combination with the extended sampling by T-REMD. Thus when combined with extended sampling techniques, the FEP predictions can cover a large range of changes of the force field parameters. At least for refining existing models rather than *de novo* construction of new models, we expect that combining extensive sampling using a single reference Hamiltonian and FEP would be much more efficient than multiple repetitions of expensive sampling using different perturbed Hamiltonians. When FEP is used, such repetition may be seldom needed in contexts of force field recalibrations, except for being used as final tests.

Although we have only tested the approach on implicit solvent models and force field parameters involving backbone atoms, in principle the method is not limited to them. Besides force field development, the approach may also be employed to identify roles of different physical interactions in conformational equilibriums of peptide and protein systems.

ACKNOWLEDGMENTS

We thank Professor Wilfred van Gunsteren for providing us with the GROMOS96 package. We thank Dr. Yuedong Yang for valuable discussions. Financial support from the Chinese Natural Science Foundation (Grant No. 30670485) and from the Chinese Ministry of Science and Technology (Grant No. 2006AA02Z303) is acknowledged.

[1] W. F. van Gunsteren, D. Bakowies, R. Baron, I. Chandrasekhar, M. Christen, X. Daura, P. Gee, D. P. Geerke, A. Glattli, P. H. Hunenberger, M. A. Kastenholz, C. Oostenbrink, M. Schenk, D. Trzesniak, N. F. van der Vegt, and H. B. Yu, Angew. Chem., Int. Ed. **45**, 4064 (2006).
[2] A. D. Mackerell, Jr., J. Comput. Chem. **25**, 1584 (2004).
[3] D. L. Beveridge and F. M. DiCapua, Annu. Rev. Biophys. Biophys. Chem. **18**, 431 (1989).
[4] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, J. Phys. Chem. A **101**, 3005 (1997).
[5] W. F. van Gunsteren, S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi, *Biomolecular Simulation, The GROMOS96 Manual and User Guide* (Vdf Hochschulverlag AG an der ETH Zürich, Zürich, Switzerland, 1996).
[6] A. P. Lyubartsev, A. A. Martinovski, S. V. Shevkunov, and P. N. Vorontsov-Velyaminov, J. Chem. Phys. **96**, 1776 (1992).
[7] E. Marinari and G. Parisi, Europhys. Lett. **19**, 451 (1992).
[8] Y. Sugita and Y. Okamoto, Chem. Phys. Lett. **314**, 141 (1999).
[9] J. Tirado-Rives and W. L. Jorgensen, Biochemistry **30**, 3864 (1991).
[10] B. Cui, M. Y. Shen, and K. F. Freed, Proc. Natl. Acad. Sci. U.S.A. **100**, 7087 (2003).
[11] V. Munoz, P. A. Thompson, J. Hofrichter, and W. A. Eaton, Nature (London) **390**, 196 (1997).
[12] A. G. Cochran, N. J. Skelton, and M. A. Starovasnik, Proc. Natl. Acad. Sci. U.S.A. **98**, 5578 (2001).
[13] J. Zhu, H. Y. Liu, and Y. Y. Shi, J. Phys. Chem. B **106**, 4844 (2002).
[14] W. F. van Gunsteren, H. J. C. Berendsen, and J. A. C. Rullmann, Mol. Phys. **44**, 69 (1981).
[15] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[16] E. Gallicchio, M. Andrec, A. K. Felts, and R. M. Levy, J. Phys. Chem. B **109**, 6722 (2005).
[17] R. Zhou and B. J. Berne, Proc. Natl. Acad. Sci. U.S.A. **99**, 12777 (2002).
[18] W. Y. Yang, J. W. Pitera, W. C. Swope, and M. Gruebele, J. Mol. Biol. **336**, 241 (2004).