JCTC Journal of Chemical Theory and Computation

A Hamiltonian Replica Exchange Approach and Its Application to the Study of Side-Chain Type and Neighbor Effects on Peptide Backbone Conformations

Chao Xu, Jun Wang, and Haiyan Liu*

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, China

Received December 29, 2007

Abstract: We presented a Hamiltonian replica exchange approach and applied it to investigate the effects of various factors on the conformational equilibrium of peptide backbone. In different replicas, biasing potentials of varying strengths are applied to all backbone (φ,ψ) torsional angle pairs to overcome sampling barriers. A general form of constructing biasing potentials based on a reference free energy surface is employed to minimize sampling in physically irrelevant parts of the conformational space. An extension of the weighted histogram analysis formulation allows for conformational free energy surfaces to be computed using all replicas, including those with biased Hamiltonians. This approach can significantly reduce the statistical uncertainties in computed free energies. For the peptide systems considered, it allows for effects of the order of 0.5−1 kJ/mol to be quantified using explicit solvent simulations. We applied this approach to capped peptides of 2−5 peptide units containing Ala, Phe, or Val in explicit water solvent and focused on how the conformational equilibrium of a single pair of backbone angles are influenced by changing the residue types of the same and neighboring residues as well as conformations of neighboring residues. For the effects of changing side-chain types of the same residue, our results consistently showed increased preference of β for Phe and Val relative to Ala. As for neighbor effects, our results not only indicated that they can be as large as the effects of changing the side-chain type of the same residue but also led to several new insights. We found that for the N-terminal neighbors, their conformations seem to have large effects. Relative to the β conformer of an N-terminal neighbor, its α conformer stabilizes the β conformer of its next Ala disregarding the residue type of the neighbor. For C-terminal neighbors, their chemical identities seem to play more important roles. Val as the C-terminal neighbor significantly increases the PII propensity of its previous Ala disregarding its own conformational state. These results are in good accordance with reported statistics of protein coil structure libraries, proving the persistent presence of such effects in short peptides as well as in proteins. We also observed other side-chain identity and neighbor effects which have been consistently reproduced in our simulations of different small peptide systems but not displayed by coil library statistics.

Introduction

Sequence-dependences of intrinsic local conformation propensities of short peptide segments have attracted wide attentions. There has been ample evidence that such propensities may lead to sequence-specific stabilization of local structures in native proteins or nonrandom local conformations in unstructured proteins or peptides.[1–4] One interesting question is to what extent such propensities depend on not

* Corresponding author e-mail: hyliu@ustc.edu.cn.

10.1021/ct7003534 CCC: $40.75
Published on Web 06/21/2008

only the identity of individual residues but also their sequence and conformation contexts. Such context dependences lead to deviations from the "isolated-pair hypothesis" of Flory,[5] which has often been assumed in equilibrium and kinetic treatments of peptide and protein folding.[6]

A number of theoretical and experimental studies indicated that the context effects can be non-negligible. Theoretical investigations included statistical analyses of structural databases, which assume that the effects of nonlocal interactions can be averaged out over a large number of structures if only residues in coil regions are analyzed, so that conformation distributions in structure databases can reflect intrinsic preferences.[7,8] Erman and co-workers analyzed pair wise correlations of backbone torsional angles in native protein structures and found that such correlations favor the choices of native torsional angles.[9,10] Betancourt and Skolnick also showed that dihedral angle distribution for a residue can significantly depend on its sequence and conformation contexts.[11] Sosnick and co-workers performed detailed analyses of backbone conformations in a protein coil library and observed not only preferences for the polyproline II (PII) conformation but also strong influences of both the chemical identity and the conformation of neighboring residues.[12] Experimental data, especially the NMR $J^3_{\alpha HN}$ coupling constants for denatured proteins and short unstructured peptides, have shown good correlations with predictions based on database models.[13,14]

These context dependences may result from steric interactions. Srinivasan and Rose employed Monte Carlo simulations to sample conformations of unfolded peptide chains.[15] They considered only steric interactions and attractive interactions separated by no more than five residues in sequence and observed sequence-specific biases significantly anticipating native secondary structures in the sampling. Pappu et al. found that the conformation of a central residue enveloped by two residues in the helical conformation was sterically restricted by these neighbors.[16] Penkett et al. pointed out that a preceding residue with a β-branched or aromatic side chain could promote the less sterically restricted β conformation of its successive residue.[14] The coupling between solvation and electrostatic interactions has also been proposed to promote context-dependent local conformation propensities. Avbelj and Baldwin surveyed a coil library for the effects of a large or small neighboring residue on the backbone angle φ and identified correlations between neighbor effects and electrostatic free energies.[17]

Small peptides have often been employed to investigate local conformation propensities to avoid complications associated with large protein molecules. Avbelj et al. reported that the intrinsic backbone preferences are present in blocked amino acids.[18] Stenner and co-workers have analyzed structures of XA and AX dipeptides and AXA tripeptides by a range of spectroscopy techniques.[3,4] They also obtained conformational propensities in line with results from host−guest experiments on longer peptides, although they did not observe context effects in the systems they studied. Kallenbach and co-workers have employed temperature dependent far-UV CD spectra and NMR *J*-coupling constants to investigate the neighbor effects on the PII conformation in alaline peptides.[19] They found increased PII-β transitions and *J* values corresponding to more negative φ angles in alanine flanked by isoleucine residues. On the simulation side, Zaman et al. investigated conformational equilibriums in short alanine peptide models, using seven different atomic level force-fields with implicit solvent and Langevin dynamics.[20] They observed that although different force fields yield large variations in the local structure propensities, a residue's conformation was invariably sensitive to the side-chain identity and backbone conformation of its immediate sequence neighbors.

Molecular dynamics (MD) simulations with explicit solvation can be a direct way to investigate the effects of sequence and conformational contexts on backbone conformations. This may not only provide more fundamental insights into when and how such context dependences take effects but also add highly quantitative criteria for the refining and testing of biomolecular force fields. However, existing evidence suggested that context effects are generally subtle, and to quantify them would require conformational equilibriums to be estimated with much lower statistical noises achievable by standard explicit solvent simulations.

Simulations based on generalized canonical ensembles can be employed to reduce statistical errors caused by rare transitions between different local minima on complicated conformational free energy surfaces. Those methods included multicanonical algorithm,[21] simulated tempering,[22,23] and replica exchange method[24–27] (also called replica Monte Carlo,[24] multiple Markov chain,[25] or parallel tempering[26,27]). In particular, using temperature as an extended degree of freedom, temperature replica exchange molecular dynamics (TREMD) have been widely applied to avoid sampling inefficiency in peptide and protein simulations. TREMD, however, meets with difficulty for explicitly solvated systems, because to maintain reasonable exchange acceptance ratios, the temperature difference between neighboring replicas need to be very small and the number of replicas need to be large.[28]Another problem is that although high temperature enhances barrier crossing, it may shift the equilibrium of two states to favor the larger entropy one making the transition less reversible.[29]

A number of alternatives to TREMD have been investigated.[30–33] One approach to avoid the problem of TREMD is Hamiltonian replica exchange (HREMD) in which a series of biasing potentials are used instead of temperatures as the extra degree of freedom. The replica exchange umbrella sampling (REUS) presented by Sugita et al. is one kind of multidimensional replica exchange method (MREM).[34] Fukunishi et al. have proposed scaled hydrophobicity REM and phantom chain REM when studying protein structure prediction problem.[28] Other HREMD schemes include that by Jang et al.[35] who proposed a generalized effective potential to change the effective temperature of the system by modifying the torsional and nonbonded terms of the potential energy function, by Affentranger et al.[36] who proposed to variably scale interactions within the protein and between protein and solvent atoms in the generalized ensemble and leave solvent−solvent interactions unchanged,

and by Kwak et al.[37] who proposed to scale partial local conformation energy.

A critical issue for properly extending a canonical ensemble into a generalized ensemble is to maintain the physical relevance of conformations sampled by the different replicas, and this can be a major difficulty with nondiscriminative, general scaling of certain interactions in HREMD. On the other hand, umbrella sampling is another commonly used technique to overcome sampling inefficiency. An umbrella potential can be designed to compensate the free energy barriers along specific degrees of freedom (DOF) or reaction coordinates (RC), while maintaining the physical relevance of sampling along DOFs perpendicular to the RCs. Equilibrated sampling with umbrella potentials applied on a single or few RCs can still be difficult to achieve if major free energy barriers exist along the remaining DOFs. If a set of coordinates causing major difficulties for sampling could be identified, coordinate-specific modification of the Hamiltonian by umbrella potentials can be integrated with the generalized canonical ensemble scheme so that free energy barriers along a set of coordinates can be simultaneously demolished and truly converged results can be obtained. One example is given by a recent work of Kannan and Zacharias, who proposed a HREMD scheme in which a series of five specifically parametrized potentials on the backbone (φ,ψ) angles were applied to different replicas. Their simulations of peptide systems showed efficiency of this approach.[38]

In this study, we describe a HREMD approach we developed independently which also employed biasing potentials on backbone conformations. The biasing potentials are based on the (φ,ψ) free energy surface of the alanine dipeptide and a simple functional form for boosted sampling proposed by Hamelberga et al.[39] These potentials compensate the free energy barriers to varying extents on the φ-ψ surface in different replicas, allowing the simultaneous sufficient sampling of different (φ,ψ) pairs in a peptides without overextending the sampling into irrelevant parts of the (φ,ψ) spaces. An extension of the weighted histogram analysis method (WHAM) can be employed so that all replicas can be used to compute conformational free energy surfaces. We show that the resulting low statistical uncertainties enable meaningful extraction of residue identity and neighbor effects from simulations of model peptides with explicitly solvent. We will consider such effects on conformational equilibriums in several model systems containing 2−5 peptide units composed of Ala, Val, or Phe. Val has been chosen as the simplest representative for residues with β-branched side chains and Phe for residues with aromatic side chains.[12]We note that systematic deviations of biomolecular force fields from quantum mechanical/molecular mechanical calculations and database statistics for the descriptions of conformational equilibriums in the Ramachandran space have been revealed.[40] Reoptimizing the backbone torsional angle terms in several force fields have led to significantly improved models.[40–43] Because we have not tried to correct such potential systematic deviations in the force field we used, we will not consider comparing the absolute probabilities of different backbone conformers from our simulations with experiments or database statistics. Instead, we will only consider comparisons of the shifts in these probabilities upon changes of side-chain identities or sequence/conformational contexts. These shifts are mainly results of intra- and intermolecular nonbonded interactions and should not be influenced by refining the backbone torsional angle terms in the force field.

Materials and Methods

1. Biasing Potentials for HREMD Simulations of Peptides. For each replica i, a biasing term is applied to each pair of backbone (φ,Ψ) angles, resulting in a total potential energy function of the form

$$V^i_{total} = V_{system} + V^i_{bias}, i = 1, ..., n_r \quad (1)$$

Here V_{system} is the unbiased potential energy function, n_r is the total number of replicas, and

$$V^i_{bias} = \sum_{m=1}^{n_m} \Delta V(\phi_m, \psi_m, \alpha_i) \quad (2)$$

Here the summation is over the total number of n_m backbone ($\varphi\Psi$) pairs.

Different from the work of Kannan and Zacharias who used five fixed biasing potentials for their HREMD simulations,[38] we use a functional form proposed by Hamelberga et al.[39] and the free energy surface $\Delta F_{ref}(\varphi,\psi)$ of alanine dipeptide to design the biasing potential ΔV

$$\Delta V(\phi, \psi, \alpha_i) = \begin{cases} 0, \text{ if} \Delta F_{ref}(\phi, \psi) \geq E_{cut} \\ \dfrac{(E_{cut} - \Delta F_{ref}(\phi, \psi))^2}{\alpha_i + (E_{cut} - \Delta F_{ref}(\phi, \psi))}, \text{ otherwise} \end{cases} \quad (3)$$

In the above formula, the strength of the biasing term depends monotonically on the replica-specific parameter α_i, with $\Delta V \equiv 0$ for $\alpha_i \equiv \infty$ and $\Delta V = E_{cut} - \Delta F_{ref}(\varphi,\psi)$ for $\alpha_i = 0$.

The reason for designing the biasing terms based on $\Delta F_{ref}(\varphi,\psi)$ is that to achieve optimum effects on sampling efficiency, it is preferred that ΔV compensates the free energy barriers separating the major stable conformational states. The same biasing potentials, however, should not result in overextended sampling into irrelevant regions of the conformational space. Here we assume that free energy surface of a model alanine dipeptide (with solvation taken into account) would provide a good first order approximation to the free energy surfaces of (φ,ψ) surfaces in peptides. Such a biasing potential with $\alpha_i = 0$ applied to alanine dipeptide would result in a flat (φ,ψ) free energy surface for regions associated with free energies lower than E_{cut} relative to the lowest free energy minimum. With a properly chosen E_{cut} the free energy barriers separating different conformational minima can be demolished without introducing too much overly extended sampling.

2. Obtaining the Reference Free Energy Surface. The reference free energy surface is numerically represented on a grid spanning the (φ,ψ) space with a bin size of 5° × 5°. Functional values and derivatives at off-grid points have been computed using bicubical interpolation[44] as has been first

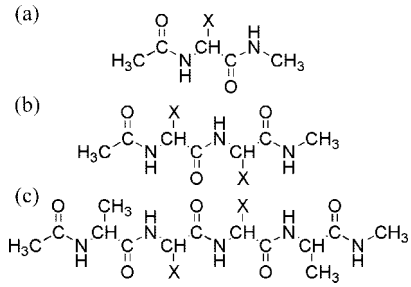

Figure 1. Peptide models containing (a) 2, (b) 3, and (c) 5 peptide units. In (a) X represents the side chain of one of alanine, phenylalanine, and valine and in (b) and (c) one of X can be alanine and the other can be one of alanine, phenylalanine, and valine.

employed in protein simulations by Mackerell et al.[41] For this purpose, a highly smooth and converged free energy surface is needed. We used a multidimensional adaptive umbrella sampling procedure[45] which makes use of WHAM[38,46] to construct this surface. Details of the dipeptide system are given in the computational details section. Repetitive calculations with altered initial configurations indicated that statistical errors can be less than 0.8 kJ/mol over the entire resulting surface and much smaller in lower free energy regions.

Based on the barrier heights on the computed $\Delta F_{ref}(\varphi,\psi)$, we choose $E_{cut} = 23$ kJ/mol. Here we have chosen the series of α, so that the barrier heights on the biased free energy surfaces decreased approximately evenly between neighboring replicas. With the number of replicas $n_r = 8$, the α_is are 10^8, 120, 50, 26.67, 15, 8, 3.33, and 0, respectively, in descending order, with the strengths of the respective biasing potentials in ascending order.

3. Applying WHAM to Trajectories of Replicas to Construct Free Energy Surfaces. Here to construct the free energy surface using a single pair of (φ,ψ) as the reaction coordinates, the effects of the biasing potentials on all (φ,ψ) pairs should also be taken into account for all replicas with a nonzero biasing potentials. In the Appendix we give the corresponding WHAM equations which allows for conformations sampled for all the replicas to be used, not only the unbiased one.

4. Peptide Systems and Simulation Details. The various peptide models considered are given in Figure 1. These are respectively X, X-Ala, Ala-X, Ala-X-Ala-Ala, and Ala-Ala-X-Ala blocked by CH_3-CO- at the N terminus and -NH-CH_3 at the C terminus, respectively. X is one of Ala, Val, or Phe. We will denote these systems by their sequence strings with "nc" attached to the left side and "cc" attached to the right side, representing the N and C terminal capping groups, respectively.

Simulations have been performed with the software package GROMOS96 with necessary modifications to implement the biasing potentials and replica exchange operations. The solute have been described by the GROMOS 53A6 force field.[47] Each peptide in its fully extended conformation ($(\varphi,\psi) = (180,180)$) was solvated in a box of SPC[48] water, size of the cubic periodic box determined by maintaining a minimum distance of 1.4 nm for any solute atom from the box boundary, with the center of mass of the solute at the box center. The weak coupling method[49] has been used to maintain a constant temperature of 300 K and a pressure of 1 atm, with relaxation times of 0.1 and 0.5 ps, respectively. All bond lengths have been constrained using SHAKE.[50] The integration time step was 2 fs. Nonbonded interactions have been treated with a twin-range cutoff method[51] with reaction field corrections for long-range interactions, using a short-range cutoff of 0.8 nm, a long-range cut off pf 1.4 nm, and a dielectric permittivity of 54 for the reaction field corrections, the pair list for short-range interactions updated every 0.02 ps.

Following initial energy minimization, each peptide system, including the dipeptide systems, has been copied into eight replicas, each with a different biasing potential applied to all backbone (φ,Ψ) pairs. Then 30.04 ns REMD simulations have been performed. The initial 40 ps has been used for equilibration and has not been used for analysis. In the HREMD simulations, every 100 steps (0.2 ps), replica exchange was attempted between neighboring replicas. At each exchange step, all the 7 pairs of neighboring replicas were considered in random order, with the constraint that a configuration can be exchanged for at most one time at a given step.

To provide comparisons with the HREMD results, one of the peptide systems, ncAAFAcc, was subjected to 8 independent normal MD simulations at 300 K. Each simulation included 10 ns sampling after 90 ps of equilibration.

Results

1. Convergences and Statistical Errors. As the effects we are interested in translated into small free energy differences, it is critical for the statistical errors produced by the HREMD process to be sufficiently small relative to the magnitudes of these effects. To estimate the statistical errors, we divided each of the 30 ns REMD trajectory sets (each set containing trajectories for all 8 replicas) into blocks of equal lengths. With block lengths of 1, 2, 5, and 10 ns, a REMD trajectory set was divided into 30, 15, 6, and 3 blocks, respectively. Each block was employed separately to construct a free energy surface. Then the averages and standard deviations of free energy surfaces computed using blocks of the same length can be computed. These standard deviations provided estimations of statistical errors and of how such errors decayed with the increasing length of the HREMD trajectories.

Among all the systems considered, ncAFAAcc and ncAAFAcc are the ones associated with the largest statistical uncertainties because of their larger numbers of backbone DOFs and the more difficult sampling of different side-chain conformations for Phe. In addition, as α conformation is the least populated among the three major local backbone conformers, the free energy surfaces for the neighboring residue in the α conformation contained the largest statistical uncertainties as they were computed with the least amount of sampled conformations. Thus to demonstrate the convergence property of the HREMD procedure, as examples for the best and worst situations, respectively, we present results for the free energy surfaces with respect to the (φ,ψ) angles of X in ncXcc and of the second Ala in the ncAAFAcc

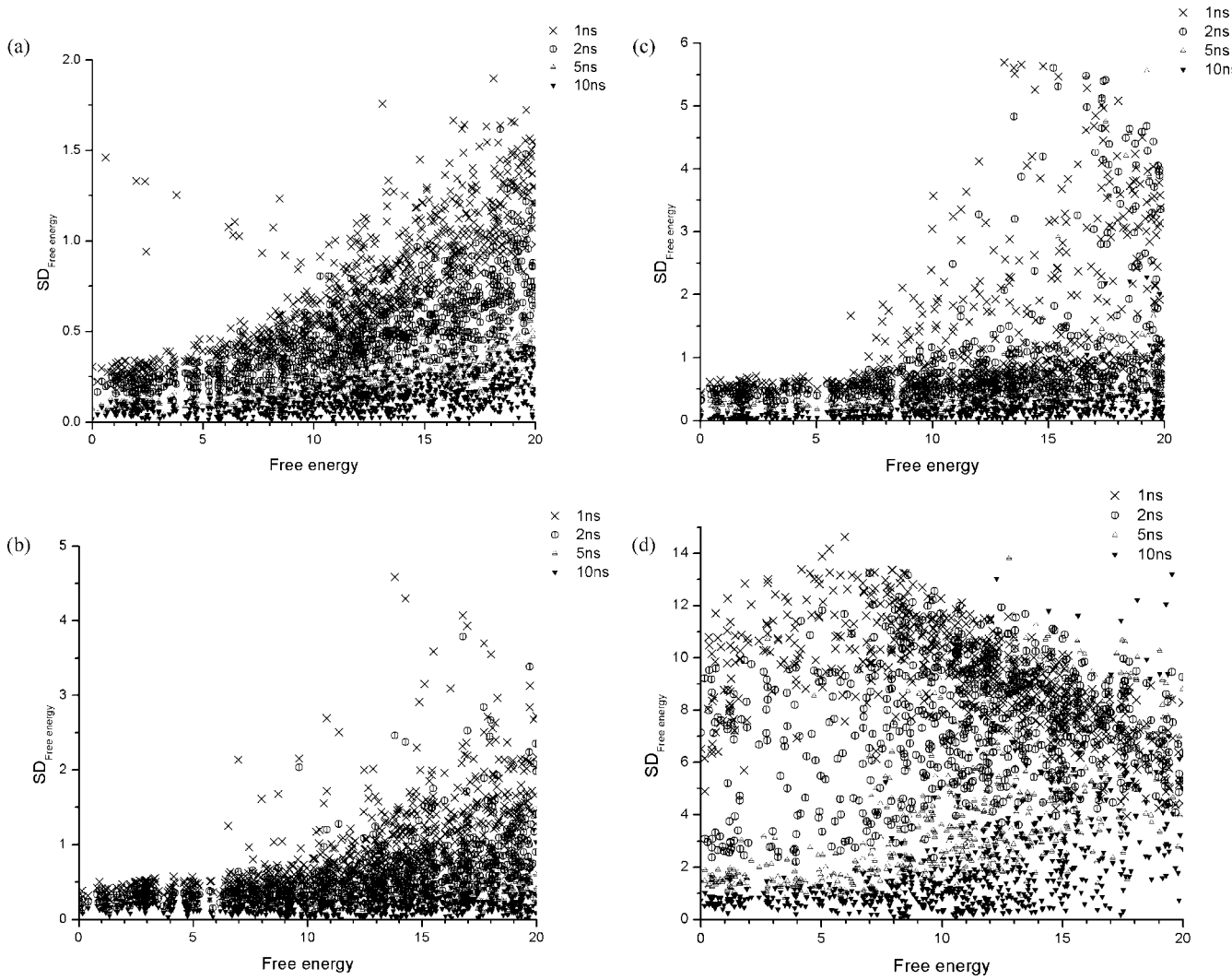

Figure 2. Standard deviations of free energy surfaces determined using different block lengths. Different points on the surfaces have been shown as functions of the free energies relative to the minimum points on the surfaces. (a)−(c) are for the ncXcc systems with X = Ala, Phe, and Val, respectively. (d) is for the second Ala in ncAAFAcc, with F in the α conformation.

system with its neighboring Phe in the α conformation. In Figure 2, we show the standard deviations of the relative free energies, computed using varying block lengths, as functions of the relative free energies themselves at different (φ,ψ) angles. The relative free energies plotted have been computed by treating the entire 30 ns trajectory set as a single block. Before used for averaging and standard deviation computing, each free energy surface computed using a smaller block has been shifted by a constant value so that the root-mean-square deviations from the free energy surface computed using the entire data set were minimized. Only the (φ,ψ) regions with free energies less than 20 kJ/mol relative to the minimum have been considered to determine the constant shifts.

Obviously, the statistical uncertainties depend not only on the block lengths but also on the relative free energies. And for our purpose only the statistical uncertainties in the lower free energy regions are of concern, as only these regions contribute to the conformational equilibriums. When the block length is increased from 1 to 10 ns, the standard deviations for regions with relative free energies below 10 kJ/mol descend from a few kJ/mol to below 0.2 kJ/mol for ncXcc and from above 10 kJ/mol to mostly below 1 kJ/mol for ncAAFAcc.

The standard deviations shown in Figure 2 have been computed for 10^o X 10^o (φ,ψ) bins. To estimate relative free energies between different conformers and the corresponding statistical uncertainties, we first transform a free energy surface into normalized probability densities, then integrated the probability densities within each basins to obtain the total probability P_c for conformer C and used the expressions $G_c = -RT\ln P_c$ to represent the free energy of conformer C. The different basins have been defined as $(-180^o<\varphi<90^o,\ 60^o<\Psi<240^o)$ for β, $(-90^o<\varphi<0^o,\ 60^o<\Psi<240^o)$ for PII, and $(-180^o<\varphi<0^o,\ -120^o<\Psi<60^o)$ for α. As the free energies at the boundaries separating different basins are usually much higher than those close to the minima, the exact choices of these boundaries have negligible influences on the final results, except for the φ value separating β and PII, which are not well separated minima on the dipeptide surfaces. We used the same boundary φ values for different systems for the results to be comparable between each other. The averaged values and standard deviations of G_c have been computed using the

Table 1. Averaged Values and Standard Deviations of Conformational Free Energies (in kJ/mol) for the Second Ala in ncAAFAcc with the Neighboring Phe in Different Conformations, Determined with Different Block Lengths

	block length	1 ns	2 ns	5 ns	10 ns
Phe in	G_β	1.54 ± 0.32	1.54 ± 0.20	1.54 ± 0.15	1.53 ± 0.03
	G_{PII}	3.01 ± 0.44	2.98 ± 0.32	2.96 ± 0.17	2.96 ± 0.12
	G_α	6.12 ± 2.37	5.62 ± 1.34	5.45 ± 0.85	5.36 ± 0.48
Phe in PII	G_β	1.65 ± 0.36	1.68 ± 0.25	1.65 ± 0.18	1.67 ± 0.08
	G_{PII}	2.68 ± 0.48	2.65 ± 0.35	2.64 ± 0.13	2.63 ± 0.02
	G_α	6.75 ± 2.79	6.05 ± 1.89	5.74 ± 0.89	5.55 ± 0.48
Phe in α	G_β	1.84 ± 1.22	1.72 ± 0.62	1.70 ± 0.45	1.62 ± 0.07
	G_{PII}	3.30 ± 1.39	3.10 ± 0.85	2.90 ± 0.41	2.80 ± 0.23
	G_α	6.92 ± 4.38	5.92 ± 2.32	5.70 ± 1.77	5.55 ± 1.02

trajectory partitioning scheme described above. As expected, the statistical uncertainties in G_cs are much smaller than in the free energy surfaces.

Table 1 shows the averaged values and statistical uncertainties for G_c of the second Ala in ncAAFAcc with its neighboring Phe in different backbone conformations, computed using different block lengths. As expected, the statistical uncertainties of G_c are directly correlated with the number of sampled configurations with conformer C. At a given block length, the largest uncertainties are associated with G_αs because of its higher values or less sampling (0.48 to 1.02 kJ/mol for a block length of 10 ns), especially with the neighboring Phe also in the α conformation (1.02 kJ/mol). With a block length of 10 ns, the statistical uncertainties of other G_cs are much smaller (0.02−0.2 kJ/mol). In later sections, we will report the standard deviations computed using a 10 ns block length as the estimated statistical uncertainties of the computed G_cs.

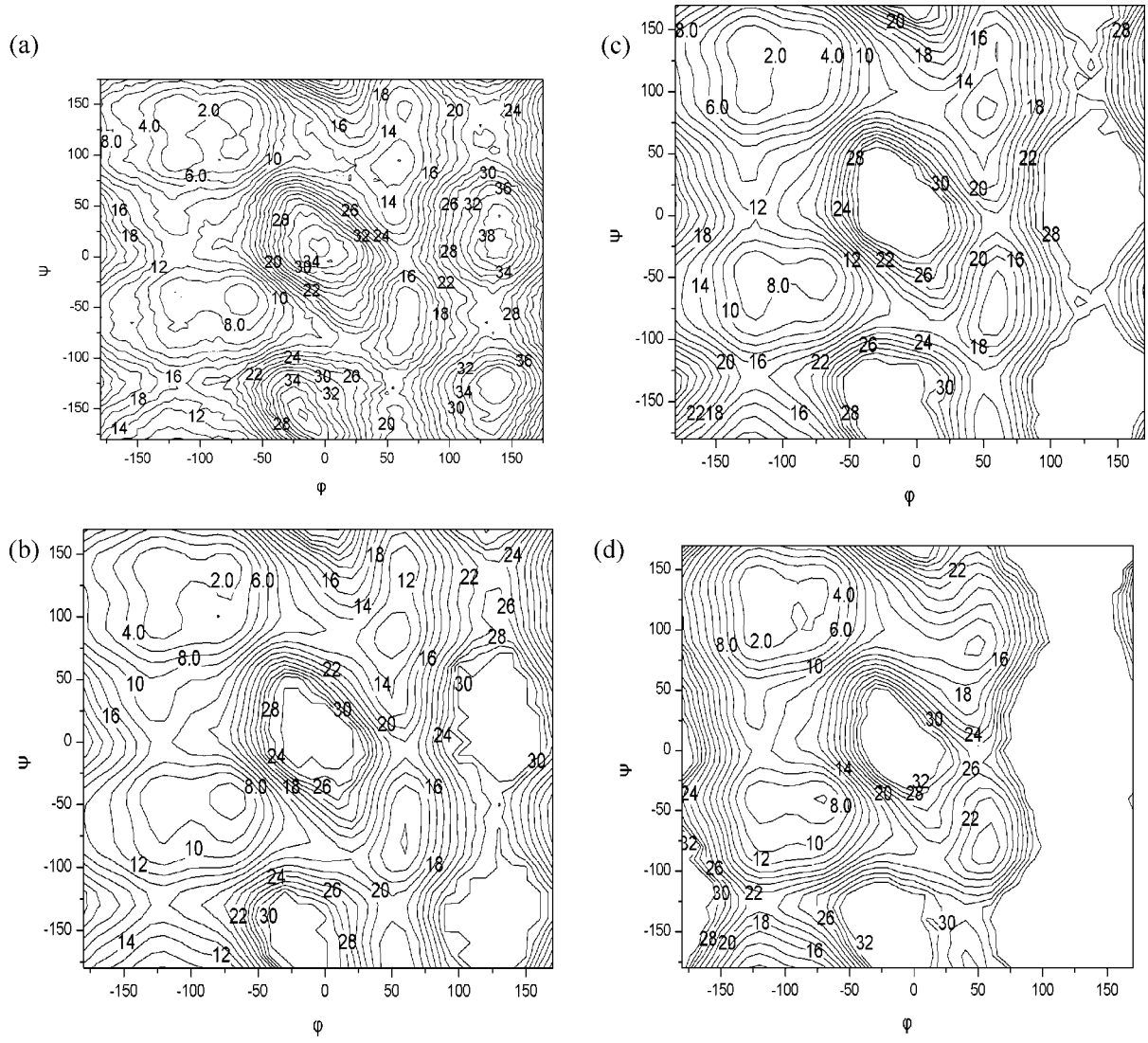

Figure 3. Contour maps of the (φ, Ψ) free energy surfaces in different model dipeptides. The numerical labels are in kJ/mol. (a) is for ncAcc computed using adaptive umbrella sampling. (b)−(d) are for ncXcc with X = Ala, Phe, and Val, respectively, computed using HREMD and the extended WHAM formulations.

Table 2. Free Energies of Different Conformers and Their Differences (in kJ/mol), To Display the Effects of Varying the Side-Chain Identity between Ala, Phe, and Val in Different Model Peptides[a]

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA*cc	1.40 ± 0.01	2.95 ± 0.02	5.74 ± 0.08	1.56 ± 0.02	4.34 ± 0.09
ncF*cc	1.00 ± 0.01	3.70 ± 0.08	6.11 ± 0.20	2.71 ± 0.09	5.12 ± 0.19
ncV*cc	1.27 ± 0.01	2.84 ± 0.01	6.47 ± 0.10	1.57 ± 0.01	5.20 ± 0.11
ncA*Acc	1.57 ± 0.02	2.68 ± 0.02	5.70 ± 0.10	1.11 ± 0.03	4.13 ± 0.11
ncF*Acc	1.14 ± 0.03	3.05 ± 0.04	7.10 ± 0.41	1.91 ± 0.05	5.96 ± 0.43
ncV*Acc	1.40 ± 0.06	2.56 ± 0.02	6.77 ± 0.40	1.16 ± 0.08	5.37 ± 0.46
ncAA*cc	1.46 ± 0.01	2.83 ± 0.06	5.78 ± 0.21	1.38 ± 0.05	4.32 ± 0.22
ncAF*cc	0.90 ± 0.03	4.30 ± 0.08	5.56 ± 0.03	3.41 ± 0.10	4.66 ± 0.05
ncAV*cc	1.24 ± 0.05	2.96 ± 0.04	6.29 ± 0.46	1.71 ± 0.06	5.05 ± 0.51
ncAA*AAcc	1.75 ± 0.06	2.67 ± 0.07	5.01 ± 0.07	0.92 ± 0.13	3.26 ± 0.09
ncAF*AAcc	1.16 ± 0.11	3.42 ± 0.15	5.92 ± 0.90	2.26 ± 0.10	4.76 ± 1.00
ncAV*AAcc	1.46 ± 0.06	2.77 ± 0.14	5.71 ± 0.72	1.31 ± 0.11	4.25 ± 0.77
ncAAA*Acc	1.69 ± 0.08	2.82 ± 0.02	4.94 ± 0.21	1.12 ± 0.09	3.25 ± 0.29
ncAAF*Acc	1.00 ± 0.03	3.52 ± 0.02	6.43 ± 0.20	2.51 ± 0.02	5.43 ± 0.23
ncAAV*Acc	1.27 ± 0.01	2.83 ± 0.01	6.59 ± 0.18	1.57 ± 0.02	5.32 ± 0.18

[a] The residues for which the free energies of their backbone conformations are shown have been marked with the superscript *.

The statistical uncertainties contained in the eight 10 ns normal MD simulations of ncAAFAcc were compared with those in the HREMD results. As expected, samplings by normal MD suffered from insufficient statistics because of the rare transitions between different conformers, and estimations of the relative free energies with statistical uncertainties comparable to that obtained by the HREMD approach cannot be obtained from 8 × 10 ns simulations. Considering the conformation of the second Ala without differentiating conformers of the neighboring Phe, the proportions of the α conformer varied between 2.0% and 34% in different simulations. For the β and PII conformers the values varied between 34.3% and 60.9% and between 22.1% and 33.5%, respectively. Such large variations in the computed conformational probabilities themselves would not allow meaningful extraction of the subtle relative effects of changes in the side-chain identity or in the sequence/conformation contexts.

2. Effects of Side-Chain Identities. Figure 3 shows the free energy surfaces for the model ncXcc dipeptides. Under the given set of force field parameters, all three free energy surfaces show similar qualitative features, with basins centered around (-65^o,-40^o), (-120^o,145^o), and (-70^o,-145^o) corresponding to the α, β, and PII conformers, respectively. Differences associated with different side-chain types are visible. For ncAcc, the results obtained using adaptive umbrella sampling and using HREMD are essentially the same, with the HREMD surface much smoother, testifying to the validity of the extended WHAM formulations given in the Appendix. Table 2 lists the free energies of different conformers associated with a given residue type X in different model peptides and at different positions. We note that the development of the GROMOS 53A6 force field parameters has so far mainly focused on the nonbonded interaction parameters which have been optimized to reproduce condensed phase thermodynamics data of pure and mixtures of small molecules in the liquid phase, and bonded interactions, especially the torsional angle terms associated with peptide backbone conformations have not been changed with respect to the older version. The main purpose of the current study is to look at how various factors such as side-chain identities as well as identities and conformations of neighboring residues would affect the conformational equilibriums. Such effects can be reflected by how the free energy differences between different conformers (ΔG_cs such as G_α-G_β and G_{PII}-G_β) changes across different systems. Unlike the differences themselves, these changes do not depend on the analytical torsional angles terms in the force field.

Table 2 shows that although the effects of side-chain identities on the ΔG_c values (below 1.2 kJ/mol) are quite small, many of them are still significantly larger than the corresponding statistical uncertainties (mostly 0.01−0.2 kJ/mol) and indicate definite and consistent effects of changing side-chain identities. In ncXcc, both Phe and Val have G_α-G_β values larger (by ca. 0.8 kJ/mol) than Ala. Interestingly, the G_{PII}-G_β values for Phe are significantly larger (by ca. 1.1 kJ/mol) than for Val and Ala, indicating that Phe does not prefer PII, at least in such small peptides. The effects of side-chain identities are consistently reproduced in longer peptides if we compare the G_α-G_β and G_{PII}-G_β values between Ala, Phe, and Val at the same position in a peptide of the same length (see Table 2).

It is also interesting to compare the ΔG_cs of the same residue type at different positions and in different peptide systems. For the second and third Ala in the pentapeptide model ncAAAAcc, the G_α-G_β values are ca. 3.2 kJ/mol, more than 1 kJ/mol lower than the corresponding differences in the shorter peptides ncAcc and ncAAcc. For Phe and Val, going from the shorter to the longer peptides does not result in such large changes in G_α-G_β. These combined produced much larger increases in G_α-G_β upon changing the central residues from Ala to Phe or Val in the model pentapeptides than in the shorter peptides. Effects of changing peptide length on G_{PII}-G_β are smaller.

In summary, our HREMD simulations show that Phe and Val prefer β over α relative to Ala. In addition, Phe does not prefer the PII conformation relative to β as compared with either Val or Ala. For Ala, the preference for α relative to β increases significantly when the peptide is extended into pentapeptides. For Phe and Val, such peptide length dependences are not strong.

3. N-Terminal Side Neighbor Effects. Tables 3 and 4 list the free energies G_c of different conformers of the second

Table 3. Effects of Chemical Identity and Conformation (Noted by Subscripts) of the N-Terminal Side Neighbor on Conformational Equilibriums of the Second Ala (Marked with the Superscript *) in ncXAcc Systems[a]

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA_βA*cc	1.49 ± 0.02	2.77 ± 0.05	5.76 ± 0.24	1.28 ± 0.04	4.27 ± 0.26
ncA_{PII}A*cc	1.42 ± 0.02	2.88 ± 0.09	5.80 ± 0.16	1.46 ± 0.10	4.38 ± 0.16
ncA_αA*cc	1.26 ± 0.04	3.05 ± 0.11	6.16 ± 0.23	1.80 ± 0.14	4.90 ± 0.23
ncF_βA*cc	1.49 ± 0.03	3.08 ± 0.09	4.88 ± 0.20	1.60 ± 0.12	3.39 ± 0.19
ncF_{PII}A*cc	1.43 ± 0.02	3.28 ± 0.07	4.66 ± 0.11	1.84 ± 0.09	3.22 ± 0.08
ncF_αA*cc	1.14 ± 0.04	3.19 ± 0.04	6.77 ± 0.34	2.05 ± 0.04	5.63 ± 0.39
ncV_βA*cc	1.69 ± 0.04	2.49 ± 0.04	5.61 ± 0.15	0.80 ± 0.08	3.92 ± 0.19
ncV_{PII}A*cc	1.63 ± 0.02	2.52 ± 0.01	5.74 ± 0.17	0.89 ± 0.01	4.12 ± 0.18
ncV_αA*cc	1.17 ± 0.05	3.08 ± 0.07	6.66 ± 0.20	1.91 ± 0.11	5.49 ± 0.25

[a] Free energies and differences in kJ/mol are given.

Table 4. As in Table *3* but for the ncAXAAcc Systems

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncAA_βA*Acc	1.76 ± 0.11	2.72 ± 0.04	4.89 ± 0.26	0.95 ± 0.13	3.13 ± 0.36
ncAA_{PII}A*Acc	1.74 ± 0.09	2.83 ± 0.06	4.84 ± 0.33	1.08 ± 0.02	3.10 ± 0.42
ncAA_αA*Acc	1.37 ± 0.04	3.08 ± 0.12	5.27 ± 0.34	1.71 ± 0.12	3.90 ± 0.37
ncAF_βA*Acc	1.69 ± 0.13	2.75 ± 0.09	4.85 ± 0.26	1.06 ± 0.22	3.15 ± 0.38
ncAF_{PII}A*Acc	1.55 ± 0.05	2.98 ± 0.07	4.95 ± 0.24	1.44 ± 0.06	3.40 ± 0.27
ncAF_αA*Acc	1.28 ± 0.14	3.12 ± 0.04	6.20 ± 1.27	1.84 ± 0.10	4.92 ± 1.41
ncAV_βA*Acc	1.97 ± 0.05	2.53 ± 0.09	4.60 ± 0.17	0.56 ± 0.14	2.63 ± 0.18
ncAV_{PII}A*Acc	1.96 ± 0.04	2.53 ± 0.12	4.81 ± 0.34	0.57 ± 0.13	2.84 ± 0.36
ncAV_αA*Acc	1.30 ± 0.15	3.21 ± 0.23	5.50 ± 0.33	1.91 ± 0.36	4.20 ± 0.46

Table 5. Effects of Chemical Identity and Conformation (Noted by Subscripts) of the C-Terminal Side Neighbor on Conformational Equilibriums of the First Ala (Marked with the Superscript *) in ncAXcc Systems[a]

Peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncA^*A_βcc	1.61 ± 0.03	2.66 ± 0.03	5.58 ± 0.08	1.05 ± 0.06	3.97 ± 0.11
ncA^*A_{PII}cc	1.48 ± 0.02	2.70 ± 0.03	6.07 ± 0.18	1.22 ± 0.04	4.59 ± 0.20
ncA^*A_αcc	1.57 ± 0.06	2.60 ± 0.06	5.96 ± 0.22	1.03 ± 0.11	4.39 ± 0.26
ncA^*F_βcc	1.37 ± 0.02	2.96 ± 0.09	5.98 ± 0.33	1.58 ± 0.09	4.61 ± 0.34
ncA^*F_{PII}cc	1.58 ± 0.06	2.59 ± 0.05	5.99 ± 0.15	1.00 ± 0.11	4.40 ± 0.20
ncA^*F_αcc	1.36 ± 0.06	2.75 ± 0.03	6.68 ± 0.42	1.39 ± 0.08	5.33 ± 0.48
ncA^*V_βcc	1.98 ± 0.03	2.09 ± 0.05	6.11 ± 0.27	0.11 ± 0.05	4.13 ± 0.28
ncA^*V_{PII}cc	1.85 ± 0.04	2.14 ± 0.03	6.51 ± 0.24	0.29 ± 0.07	4.66 ± 0.25
ncA^*V_αcc	1.91 ± 0.11	2.00 ± 0.02	6.87 ± 0.64	0.09 ± 0.12	4.97 ± 0.75

[a] Free energies and differences in kJ/mol are given.

Ala in the ncXAcc and the third Ala in the ncAXAAcc systems. Results for X in different conformations have been listed separately.

Interestingly, the conformations of X have much larger and better-defined effects on the conformational equilibrium of its next Ala than the identity of X. Compared with X in the β conformation, X in the α conformation results in stabilization of β of the next Ala relative to α by 0.6−0.7 kJ/mol when X is Ala, and 1.2−1.8 kJ/mol when X is Phe or Val, and relative to PII by 0.5−0.7 kJ/mol when X is Ala or Phe and 1.1−1.3 kJ/mol when X is Val. Also compared with X in the β conformation, X in PII conformation does not result in much change in the conformational equilibrium of the next Ala, with the computed ΔG_cs less than 0.4 kJ/mol of varying signs.

The effects of the identity of X are less certain, and the computed changes are dependent on both the lengths of the model peptides and the conformations of X. Although the data show the trend that changing X in the α conformation from Ala to Phe or Val would further stabilize the β and PII conformers relative to the α conformer of the next Ala, the computed differences are small, and some of them suffer from large statistical uncertainties.

4. C-Terminal Side Neighbor Effects. Tables 5 and 6 list the free energies G_c of different conformers of the first Ala in the ncAXcc and the second Ala in the ncAAXAcc systems.

Unlike the N-terminal side neighbor effects, the chemical identity of X as the C-terminal neighbor produced consistent effects on the conformational equilibrium on its previous Ala. The G_α-G_β values for Ala with Phe or Val as its next residue in either α or β conformation are consistently larger than with Ala as its next residue, mostly by 0.5−1.0 kJ/mol except for ncAVcc. X in the PII conformation does not have much identity-specific effects on the α-β equilibrium of its previous Ala. Another consistent observation is that Val can significantly stabilize the PII conformer of its previous Ala as compared with Ala or Phe, reducing G_{PII}-G_β values from 0.8 to 1.5 kJ/mol to 0.0 to 0.3 kJ/mol.

As compared with the conformation effects of the N-terminal side neighbors, the conformation effects of X on its previous Ala are less obvious and less consistent across different residue types and peptide lengths. There seems to be the trend that X in either PII or α would slightly stabilize the β and PII conformers relative to α of the previous Ala.

Table 6. As Table 5*5*, but for the ncAAXAcc Systems

peptides	G_β	G_{PII}	G_α	G_{PII}-G_β	G_α-G_β
ncAA*A_βAcc	1.82 ± 0.08	2.68 ± 0.09	4.71 ± 0.04	0.86 ± 0.17	2.89 ± 0.10
ncAA*A_{PII}Acc	1.65 ± 0.07	2.65 ± 0.08	5.33 ± 0.07	1.01 ± 0.15	3.68 ± 0.14
ncAA*A_αAcc	1.76 ± 0.03	2.56 ± 0.18	5.35 ± 0.57	0.80 ± 0.17	3.59 ± 0.59
ncAA*F_βAcc	1.53 ± 0.02	2.96 ± 0.12	5.36 ± 0.48	1.44 ± 0.10	3.84 ± 0.50
ncAA*F_{PII}Acc	1.67 ± 0.08	2.63 ± 0.02	5.55 ± 0.48	0.96 ± 0.06	3.88 ± 0.56
ncAA*F_αAcc	1.62 ± 0.07	2.80 ± 0.23	5.55 ± 1.02	1.19 ± 0.19	3.93 ± 1.09
ncAA*V_βAcc	2.07 ± 0.08	2.19 ± 0.01	5.84 ± 0.22	0.12 ± 0.09	3.76 ± 0.30
ncAA*V_{PII}Acc	2.14 ± 0.17	2.11 ± 0.03	5.80 ± 0.58	−0.04 ± 0.15	3.65 ± 0.75
ncAA*V_αAcc	2.15 ± 0.15	1.93 ± 0.16	6.71 ± 0.08	−0.22 ± 0.31	4.56 ± 0.11

5. Correlations with Experiments and Database Statistics. The effects of side-chain identity on conformation preferences in short peptides and in coil libraries have been well documented. The increased β propensity of Phe and Val with respect to α and PII relative to Ala are consistent with experimental studies and database statistics such as refs 3, 12, and 18. The decreased PII preferences of Phe indicated by our simulations are not displayed in reported statistical analyses of databases (for example, ref 12). It is difficult to verify this particular result with reported experimental data because of the dominance of the β conformer for Phe in short peptides[3] and also because the NMR *J*-coupling constant cannot differentiate between the PII and α conformers. Whether this would indicate limitations of the force field or of the statistical analyses may deserve further investigations.

The neighbor effects from our HREMD analysis may not only be correlated with previous experimental and statistical results but also lead to additional insights into them. For the identity changing of the N-terminal neighbor between Ala, Phe, and Val, our HREMD results indicated that conformation-specific neighbor effects dominate. The N-terminal neighbor in α significantly stabilize the β conformer of the next Ala. We found this in excellent agreement with the statistics presented as Figure 6 in ref 12, which showed separately the (φ, Ψ) distributions of Ala following residues of different types and in different conformations in a restricted coil library. We found that as the C-terminal side neighbor, Val may significantly stabilize the PII conformer of its previous Ala. We also found this in consistence with Figure 7b of ref 12, which showed clearly that Ala right before residues with β-branched side chains in a restricted coil library have higher fractions in the PII conformations than those before residues with aromatic or alanine-like side chains. It is interesting to note that for this particular result, both our HREMD simulations and the database analyses showed very low uncertainties.

However, there are also some effects implicated by our results but not displayed in the coil library analyses and vice versa. Our simulations indicated that relative to Ala as the C-terminal neighbor, Val or Phe as the C-terminal neighbor in the α or β conformation would stabilize the β conformer with respect to α of its previous Ala, but this was not indicated by Figure 7b of ref 12. On the other hand this same figure showed that alanine residues with an aromatic side C-terminal neighbor would have quite different conformation preferences with the aromatic residue in different conformations, especially that when the aromatic residue is in the PII conformation, the fraction of the α conformers of the Ala drops significantly. While our results did implicate some stabilization effects of the PII conformer of Ala with the next Phe in PII, the G_α-G_β of the Ala was essentially unchanged compared with the case with Phe in the β conformation. It is yet difficult to judge whether these differences should be attributed to the limitations of either analyses (for examples, crude classifications of residues by their chemical identities and insufficient amount of data in the statistical analyses or inaccuracies of the force field) or to different environmental contexts of the model peptides and the coil library.

Conclusions

We have presented a Hamiltonian replica exchange approach and applied it to investigate the conformational equilibriums of short peptides. A general form for constructing biasing potentials of varying strengths in different replicas has been employed. For peptide systems, the biasing potentials are based on a reference free energy surface of model alanine dipeptide and applied to all backbone (φ,Ψ) pairs to overcome major sampling barriers. Free energy surfaces using one of the biased pairs as the reaction coordinates can be constructed using conformations sampled in all replicas, including the biased ones, using an extension of the WHAM formulations. This HREMD approach can significantly reduce the statistical uncertainties in computed conformational free energies, thus allowing the subtle but important effects of factors such as side-chain identities and conformations of neighboring residues to be quantified using explicit solvent simulations.

Using the GROMOS 53A6 force field, we apply this HREMD approach to investigate the effects of these factors in model peptides containing two to five peptide units. In this work, we have considered Ala, Phe, and Val as simplest representatives of non β-branched, aromatic, and β-branched side chains, respectively. As for side-chain identity effects, our results consistently showed an increased preference of β by Phe and Val relative to Ala, in agreement with other analyses. Our results not only indicate that neighbor effects can be as large as effects of the side-chain identity but also lead to additional insights into such effects. For N-terminal neighbors, the effects of their conformations seem to be larger than those of their chemical identities. Any of Ala, Phe, and Val as the N-terminal neighbor of an Ala significantly stabilizes the β conformer of the Ala when the neighbor is in the α conformation. In the case of C-terminal neighbors, their chemical identities seem to play more important roles

in modifying the conformational preferences of its previous Ala. We found that Val may significantly increase the PII propensity of its previous Ala. Beside these results which are in accordance with reported statistics of protein coil structure libraries, our simulations also indicated that the PII conformer of Phe is less preferred compared with Ala and Val and that Phe and Val as the C-terminal neighbor in either α or β conformation would stabilize the β conformer of its previous Ala.

Previous simulations have shown that different force fields did not produce consistent results for backbone conformational equilibriums of the same system. One important reason for these is that the analytical backbone torsional angle terms, a major determinant of the backbone conformational equilibriums, have been determined using different reference data in different force fields. We have focused not on the free energy differences between different conformers themselves but on the changes in these free energy differences associated with the changes of side-chain identity of the same residue and of side-chain identity and backbone conformation of neighboring residues. These changes do not depend on the analytical backbone torsional angle terms but mainly result from complicated changes in intra- and intermolecular nonbonded interactions. Parameters for such interactions in different biomolecular force fields have been shown to reproduce a more convergent set of reference data (for example, free energies of transfer between different solvents). The HREMD approach is essential for meaningful analyses of these small changes because of sufficiently lowered statistical uncertainties in the computed conformational free energies. Given the reasonable agreements between the current simulations and other analyses, we expect that accurate computations of conformational free energies using explicit solvent simulations can help not only to delineate the relationships between local sequences and conformations but also to identify their physical origins through future well-designed computer experiments. From a computational chemistry point of view, such simulations may also be employed to test and refine a given force field model, to compare different models, and to investigate when and how the inclusion of higher order effects (for example, polarization) in force fields would be important for the description of conformational equilibriums.

Acknowledgment. Financial support from the Chinese Natural Science Foundation (grant number 30670485) and from the Chinese Ministry of Science and Technology (grant number 2006AA02Z303) are acknowledged.

Appendix

In this Appendix we present the WHAM equations[38,46] for the estimation of a probability density for a reaction coordinate using HREMD. In different replicas biasing potentials have been applied to DOFs other than the reaction coordinate. We can first consider the unbiased probability density for a certain (φ_t,ψ_t) pair from a single simulated replica j

$$P_{unbiased}(\phi_t, \psi_t) = P_{unbiased}(Q) = \frac{Z^j_{biased}}{Z_{unbiased}} P^j_{biased}(Q)\Lambda_{j,Q} \tag{A1}$$

with

$$\Lambda_{j,Q} = <\exp(\beta V^j_{bias})>_{biased,j,Q} \tag{A2}$$

Here Q refers to the reaction coordinate (φ_t,ψ_t), $P^{unbiased}(Q)$ is the unbiased probability density, $Pbiased^j(Q)$ is the probability density of replica j with biasing potential V^j_{bias}, and $<...>_{biased,j,Q}$ refers to average over members of subensemble j with (φ_t,ψ_t) at Q. $Zbiased^j$ and $Z_{unbiased}$ are the configurational partition functions of the biased and original system, respectively

$$Z_{unbiased} = \int \exp(-\beta V_{system}) d\Omega \tag{A3}$$

and

$$Z^j_{biased} = \int \exp[-\beta(V_{system} + V^j_{bias})] d\Omega \tag{A4}$$

The $P^{unbiased}(Q)$ can be estimated using each of the individual replicas based on eq A1. We denote an estimation using replica j as $\tilde{P}unbiased^j$. Here we use $\tilde{X}$ to represent estimation for X from the REMD samples. The estimation considering all n_r replicas is the weighted summation of individual estimations

$$\tilde{P}_{unbiased}(Q) = \sum_{j=1}^{n_r} w_j(Q)\tilde{P}^j_{unbiased}(Q) \tag{A5}$$

The weighting factors $w_j(Q)$ should be chosen to minimize the statistical uncertainty in $\tilde{P}_{unbiased}(Q)$ subject to a normalization constraint.[38,46] If we assume that the uncertainties in $\tilde{P}_{biased}{}^j(Q)$and in $\tilde{\Lambda}_{j,Q}$ contribute independently to the uncertainties in $\tilde{P}_{unbiased}{}^j(Q)$, we have the following squared standard error

$$\begin{aligned} <\tilde{P}_{unbiased}(Q)^2> - <\tilde{P}_{unbiased}(Q)>^2 &= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \times \\ &\left\{<\left(\tilde{P}^j_{biased}(Q)\tilde{\Lambda}_{j,Q}\right)^2> - <\tilde{P}^j_{biased}(Q)\tilde{\Lambda}_{j,Q}>^2\right\} \\ \approx \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 &\left[\Lambda_{j,Q}{}^2 \delta^2 \tilde{P}^j_{biased}(Q) + P^j_{biased}(Q)^2 \delta^2 \tilde{\Lambda}_{j,Q}\right] \end{aligned} \tag{A6}$$

As in standard WHAM the uncertainties in $\tilde{P}_{biased}{}^j(Q)$ is given by[38,46]

$$\delta^2 \tilde{P}^j_{biased}(Q) = \frac{\delta^2 \tilde{H}^j(Q)}{\Delta Q^2 M^{j2}} = \frac{<H^j(Q)>_{biased,j}}{\Delta Q^2 M^{j2}} = \frac{P^j_{biased}(Q)}{M^j \Delta Q} \tag{A7}$$

Here $H^j(Q)$is the number of independent conformations sampled with reaction coordinate Q in replica j, M^j is the total number of independent conformations sampled in replica j, and ΔQ is the bin size of the reaction coordinates.

We assume that the uncertainties in $\tilde{\Lambda}_{j,Q}$ is proportional to $\tilde{\Lambda}_{j,Q}$, the fluctuations in exp (βV^j_{bias}), and inversely the square root of number of independent samples for the averaging,

$$\delta^2\widetilde{\Lambda}_{j,Q} = \delta^2 < \exp(\beta V^j_{bias}) >_{j,Q}$$
$$= \Lambda^2_{j,Q}\delta^2 < \exp(\beta V^j_{bias} - \ln\Lambda_{j,Q}) >_{j,Q}$$
$$\approx \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{<H^j(Q)>}$$
$$= \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{P^j_{biased}(Q)M^j\Delta Q} \quad \text{(A8)}$$

with

$$\delta_{\Lambda_{j,Q}}{}^2 = < [\exp(\beta V^j_{bias} - \ln\Lambda_{j,Q})]^2 >_{biased,j,Q} - < [\exp(\beta V^j_{bias} - \ln\Lambda_{j,Q})] >^2_{biased,j,Q} \quad \text{(A9)}$$

then the overall uncertainty is approximately

$$< \widetilde{P}(Q)^2 > - < \widetilde{P}(Q) >^2 \approx \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \times$$
$$\left[\Lambda_{j,Q}{}^2 \frac{P^j_{biased}(Q)}{M^j\Delta Q} + P^j_{biased}(Q)^2 \Lambda_{j,Q}{}^2 \frac{\delta_{\Lambda_{j,Q}}{}^2}{P^j_{biased}(Q)M^j\Delta Q}\right]$$
$$= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right)^2 \Lambda_{j,Q}{}^2 \frac{P^j_{biased}(Q)}{M^j\Delta Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right)$$
$$= \sum_j w_j(Q)^2 \left(\frac{Z^j_{biased}}{Z_{unbiased}}\right) \Lambda_{j,Q} \frac{P_{unbiased}(Q)}{M^j\Delta Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right) \quad \text{(A10)}$$

Minimizing the above uncertainty with respect to $w_j(Q)$ under the constraint $\sum_j w_j(Q) = 1$ results in

$$w_j(Q) = \lambda \frac{1}{Z^j_{biased}\Lambda_{j,Q}\left(1 + \delta_{\Lambda_{j,Q}}{}^2\right)} \quad \text{(A11)}$$

with λ determined by the normalization constraint

$$\lambda = \frac{1}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\Lambda_{k,Q}\left(1 + \delta_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A12)}$$

Combining eqs A1, A5, A11, and A12 gives

$$\widetilde{P}_{unbiased}(Q) =$$
$$\sum_j \frac{\frac{1}{Z^j_{biased}} \frac{M^j}{\widetilde{\Lambda}_{j,Q}\left(1 + \widetilde{\delta}_{\Lambda_{j,Q}}{}^2\right)}}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \widetilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \frac{Z^j_{biased}}{Z_{unbiased}} \frac{\widetilde{H}^j(Q)}{M^j\Delta Q} \widetilde{\Lambda}_{j,Q}$$
$$= \frac{1}{\Delta Q} \frac{\sum_j \frac{\widetilde{H}^j(Q)}{\left(1 + \widetilde{\delta}_{\Lambda_{j,Q}}{}^2\right)}}{\sum_k \frac{Z_{unbiased}}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \widetilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A13)}$$

Applying the normalization condition

$$\sum_Q \Delta Q \widetilde{P}^j_{biased}(Q) = \sum_Q \Delta Q \frac{Z_{unbiased}}{Z^j_{biased}} \widetilde{P}_{unbiased}(Q) \frac{1}{\widetilde{\Lambda}_{j,Q}} = 1 \quad \text{(A14)}$$

to all the biased probability densities produces the following set of self-consistent equations which can be solved iteratively to obtain $Z_{unbiased}/Z^k_{biased}$.

$$Z^j_{biased} = Z_{unbiased} \sum_Q \Delta Q \frac{1}{\widetilde{\Lambda}_{j,Q}} \frac{1}{\Delta Q} \frac{\sum_l \frac{\widetilde{H}^l(Q)}{\left(1 + \widetilde{\delta}_{\Lambda_{l,Q}}{}^2\right)}}{\sum_k \frac{Z_{unbiased}}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \widetilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}}$$

$$= \sum_Q \frac{\frac{1}{\widetilde{\Lambda}_{j,Q}} \sum_l \frac{\widetilde{H}^l(Q)}{\left(1 + \widetilde{\delta}_{\Lambda_{l,Q}}{}^2\right)}}{\sum_k \frac{1}{Z^k_{biased}} \frac{M^k}{\widetilde{\Lambda}_{k,Q}\left(1 + \widetilde{\delta}_{\Lambda_{k,Q}}{}^2\right)}} \quad \text{(A15)}$$

References

(1) Zimm, B. H.; Bragg, J. K. *J. Chem. Phys.* **1959**, *31*, 526–535.

(2) Serrano, L. *J. Mol. Biol.* **1995**, *254*, 322–333.

(3) Eker, F.; Griebenow, K.; Cao, X. L.; Nafie, L. A.; Schweitzer-Stenner, R. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10054–10059.

(4) Hagarman, A.; Measey, T.; Doddasomayajula, R. S.; Dragomir, I.; Eker, F.; Griebenow, K.; Schweitzer-Stenner, R. *J. Phys. Chem. B* **2006**, *110*, 6979–6986.

(5) Flory, P. J. In *Statistical Mechanics of Chain Molecules*; Wiley Interscience: New York, 1969; pp 252−256.

(6) Ohkubo, Y. Z.; Brooks, C. L. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13916–13921.

(7) Munoz, V.; Serrano, L. *Proteins: Struct., Funct., Genet.* **1994**, *20*, 301–311.

(8) Swindells, M. B.; Macarthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596–603.

(9) Keskin, O.; Yuret, D.; Gursoy, A.; Turkay, M.; Erman, B. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 992–998.

(10) Ormeci, L.; Gursoy, A.; Tunca, G.; Erman, B. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 29–40.

(11) Betancourt, M. R.; Skolnick, J. *J. Mol. Biol.* **2004**, *342*, 635–649.

(12) Jha, A. K.; Colubri, A.; Zaman, M. H.; Koide, S.; Sosnick, T. R.; Freed, K. F. *Biochemistry* **2005**, *44*, 9691–9702.

(13) Smith, L. J.; Bolin, K. A.; Schwalbe, H.; MacArthur, M. W.; Thornton, J. M.; Dobson, C. M. *J. Mol. Biol.* **1996**, *255*, 494–506.

(14) Penkett, C. J.; Redfield, C.; Dodd, I.; Hubbard, J.; McBay, D. L.; Mossakowska, D. E.; Smith, R. A. G.; Dobson, C. M.; Smith, L. J. *J. Mol. Biol.* **1997**, *274*, 152–159.

(15) Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 14258–14263.

(16) Pappu, R. V.; Srinivasan, R.; Rose, G. D. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 12565–12570.

(17) Avbelj, F.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 10967–10972.

(18) Avbelj, F.; Grdadolnik, S. G.; Grdadolnik, J.; Baldwin, R. L. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 1272–1277.

(19) Chen, K.; Liu, Z. G.; Zhou, C. H.; Shi, Z. S.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2005**, *127*, 10146–10147.

(20) Zaman, M. H.; Shen, M. Y.; Berry, R. S.; Freed, K. F.; Sosnick, T. R. *J. Mol. Biol.* **2003**, *331*, 693–711.

(21) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249–253.

(22) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451–458.

(23) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsovvelyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776–1783.

(24) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.

(25) Tesi, M. C.; vanRensburg, E. J. J.; Orlandini, E.; Whittington, S. G. *J. Stat. Phys.* **1996**, *82*, 155–181.

(26) Hansmann, U. H. E. *Chem. Phys. Lett.* **1997**, *281*, 140–150.

(27) Falcioni, M.; Deem, M. W. *J. Chem. Phys.* **1999**, *110*, 1754–1766.

(28) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058–9067.

(29) Faraldo-Gomez, J. D.; Roux, B. *J. Comput. Chem.* **2007**, *28*, 1634–1647.

(30) Cheng, X. L.; Cui, G. L.; Hornak, V.; Sinnnerling, C. *J. Phys. Chem. B* **2005**, *109*, 8220–8230.

(31) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.

(32) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.

(33) Christen, M.; van Gunsteren, W. F. *J. Chem. Phys.* **2006**, *124*, 154106.

(34) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042–6051.

(35) Jang, S. M.; Shin, S.; Pak, Y. *Phys. Rev. Lett.* **2003**, *91*, 058305.

(36) Affentranger, R.; Tavernelli, I.; Di Iorio, E. E. *J. Chem. Theory Comput.* **2006**, *2*, 217–228.

(37) Kwak, W.; Hansmann, U. H. E. *Phys. Rev. Lett.* **2005**, *95*, 138102.

(38) Kannan, S.; Zacharias, M. *Proteins: Struct., Funct., Bioinf.* **2007**, *66*, 697–706.

(39) Hamelberg, D.; Mongan, J.; McCammon, J. A. *J. Chem. Phys.* **2004**, *120*, 11919–11929.

(40) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Am. Chem. Soc.* **2004**, *126*, 698–699.

(41) Mackerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–1415.

(42) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G. M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J. M.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.

(43) Kaminski, G. A.; Friesner, R. A.; Tirado-Rives, J.; Jorgensen, W. L. *J. Phys. Chem. B* **2001**, *105*, 6474–6487.

(44) Akima, H. *ACM Trans. Math. Soft.* **1996**, *22*, 357–361.

(45) Wang, J.; Gu, Y.; Liu, H. Y. *J. Chem. Phys.* **2006**, *125*, 094907.

(46) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.

(47) Oostenbrink, C.; Villa, A.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2004**, *25*, 1656–1676.

(48) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In *Intermolecular Forces*; Pullman, B., Ed.; D. Reidel Publishing Company: Dordrecht, The Netherlands, 1981; pp 331−342.

(49) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Dinola, A.; Haak, J. R. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(50) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Chem. Phys.* **1977**, *23*, 327–341.

(51) Tironi, I. G.; Sperb, R.; Smith, P. E.; van Gunsteren, W. F. *J. Chem. Phys.* **1995**, *102*, 5451–5459.

CT7003534