Peptide Conformational Equilibria Computed via a Single-Stage Shifting Protocol

F. Marty Ytreberg* and Daniel M. Zuckerman†

Department of Computational Biology and the Department of Environmental and Occupational Health, Graduate School of Public Health, University of Pittsburgh, 200 Lothrop Street, Pittsburgh, Pennsylvania 15261

Received: March 2, 2005

We study the conformational equilibria of two peptides using a novel statistical mechanics approach designed for calculating free energy differences between highly dissimilar conformational states. Our results elucidate the contrasting roles of entropy in implicitly solvated leucine dipeptide and decaglycine. The method extends earlier work by Voter and overcomes the notorious "overlap" problem in free energy computations by constructing a mathematically equivalent calculation with high conformational similarity. The approach requires only equilibrium simulations of the two states of interest, without the need for sampling transition states. We discuss possible extensions and optimizations of the approach.

1. Introduction

Although free energy differences (ΔG) are fundamental to the description of every molecular process, computer-based estimation of ΔG remains among the most difficult and time-consuming tasks in computational chemistry and biology. Despite the obstacles, molecular mechanics ΔG calculations are presently applied for protein engineering,[1,2] drug design,[3,4] toxicology studies,[5] solubility estimation,[6,7] and determining binding affinities of ligands to proteins.[8] Although many ad hoc and approximate methods are in wide use,[9] including docking protocols,[10−14] there is a widely recognized need to more rigorously include molecular flexibility and entropic effects.[10−14] Additionally, the development of improved, polarizable molecular mechanics force fields[6,15−17] suggests rigorous, ensemble-based ΔG estimates will play an increasing role in elucidating subtle molecular effects, aided by the decreasing cost of computer resources.

Our concern here is to calculate the conformational free energy differences ΔG and entropy differences ΔS for peptides with a method based on statistical mechanics, which fully accounts for molecular flexibility. Because such techniques attempt a full accounting of entropic and enthalpic effects, they are computationally demanding, and it is the need for greater efficiency which we address here. Our ideas are framed and tested for molecular mechanics calculations, but they should also prove applicable with quantum mechanical computation.[18]

Present statistical mechanics methods can be classified into equilibrium and nonequilibrium approaches. Equilibrium ΔG calculations, such as thermodynamic integration and multistage free energy perturbation, have been used successfully for many years.[19−22] Equilibrium methods can be very accurate, but have a large computational cost. On the other hand, Jarzynski's equality[23] has recently has opened up a host of *non*equilibrium ΔG methods.[24−28] Nonequilibrium methods can provide rapid estimates for ΔG, but typically suffer from bias without careful convergence testing.[29−32] Single-stage approaches[7,33] may be considered nonequilibrium calculations which are based upon simulating only one or both end states of interest.

Poor "overlap", the lack of conformational similarity between the molecular states of interest, is a key cause of computational expense in statistical mechanics methods. This phenomenon is illustrated schematically in Figure 1. Unfortunately, in many cases of interest, poor overlap is the rule rather than the exception. Such dissimilarity is implicit in the critical problem of conformational equilibria, for instance.

The most common approach to improve overlap in free energy calculations is that used in thermodynamic integration, namely, simulating the system at multiple hybrid, intermediate stages (e.g., refs 6, 19−22, 34, 35). An alternative strategy was taken by van Gunsteren and collaborators[36] and McCammon and collaborators[33] who built nonphysical intermediate reference states that increased overlap between the two end states. A third noteworthy solution to the overlap problem is the computation of absolute free energies for each of the end states, avoiding any need for configurational overlap.[37−40]

Here, we address the overlap problem using a two-decade-old, elegant idea from solid-state systems.[41] In ref 41, Voter pointed out that shifting potential energy functions (or coordinates) by a constant vector in configuration space can dramatically improve the overlap of the states while maintaining the value of ΔG *exactly*; he applied this idea to tungsten dimers on a crystal surface. A similar approach was adopted for lattice systems by Bruce and collaborators.[42] In their "metric scaling" approach, Reinhardt and collaborators also built on Voter's idea for slow-growth ΔG estimates for crystal lattice changes.[43] Most recently, in a generalization of Voter's approach, Jarzynski introduced "targeted free energy perturbation" and applied the method to a Lennard−Jones fluid.[44] However, ref 44 also indicates that constructing the mapping is likely to be as hard as estimating the free energy itself.

Here, we introduce the first generalization of Voter's shifting strategy specifically tailored to molecular calculations. Because no intermediate stages are required, the approach has the potential to dramatically reduce computation times. Overlap is obtained by shifting *internal coordinates* of the molecule. The shifting strategy is further combined with Bennett's iterative

* E-mail: fmy1@pitt.edu.
† E-mail: dmz@ccbb.pitt.edu.

10.1021/jp0510692 CCC: $30.25
Published on Web 04/02/2005

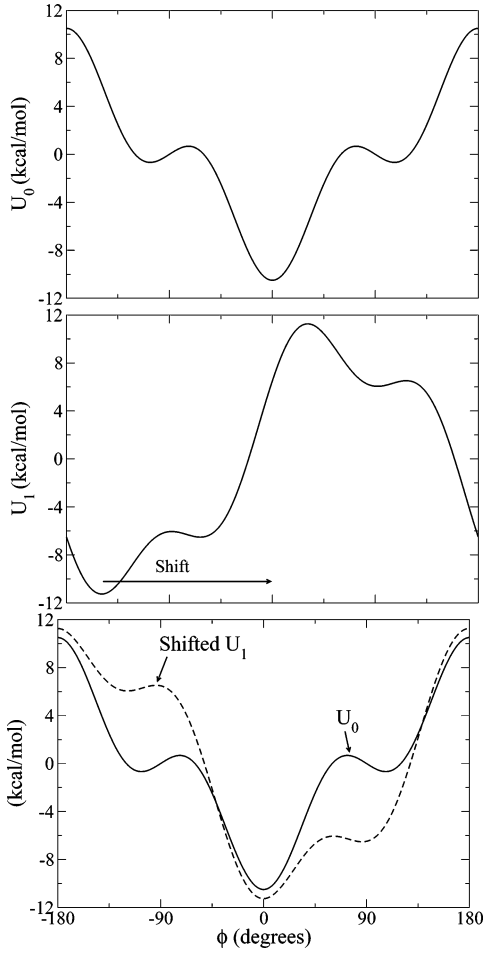

Figure 1. Illustration of the shifting idea using two idealized torsional potentials (a) $U_0(\phi)$ and (b) $U_1(\phi)$. There is minimal configurational overlap between these two potentials, since simulations will mainly sample the deep, dissimilar minima of the potentials. (c) With the use of an appropriate shifting constant, it is possible to construct overlap between the states without altering the numerical value of the conformational free energy difference. Note that for peptides, we shift internal coordinates rather than potentials (see Section 2.3); however, these two procedures are equivalent.

approach[45] to efficiently utilize the data. These two ideas, (i) shifting internal coordinates and (ii) the use of Bennett's iterative method, provide the backbone of the single-stage shifting approach presented here.

In outline, this report first builds the necessary mathematical framework for the single-stage shifting method in Section 2. Then in Section 3, the method is tested on leucine dipeptide (ACE-(leu)$_2$-NME) in GBSA implicit solvent, correctly calculating the free energy difference ΔG and entropy difference ΔS between the alpha and beta conformations. The leucine dipeptide system is also used to demonstrate various shifting approaches, and show the importance of using Bennett's iterative method. Finally, the single-stage shifting method is used to predict ΔG and ΔS between the alpha and extended conformations of decaglycine (ACE-(gly)$_{10}$-NME) in GBSA solvent.

While our results already demonstrate notable efficiency (the decaglycine calculations, for instance, would be very costly by conventional staging methods), we have not yet pursued a number of fairly clear avenues for optimization; these are discussed in Section 4. Finally, the potential extension of the single-stage shifting method to explicitly solvated systems, and for "alchemical" mutations, is explored in Section 5.

2. The Single-Stage Shifting Method

In this section we introduce a single-stage method which improves the overlap between the end states by construction. In essence, a mathematically equivalent calculation, with high overlap, is performed instead of addressing the original problem.

2.1. Constructing an Equivalent Shifted Calculation. Consider two end states defined by potential energy functions $U_0(\vec{x})$ and $U_1(\vec{x})$, where $\vec{x}$ is a set of configurational coordinates. The two states could be two different conformations, or the bound and unbound states of a protein. For problems in conformational equilibria, as studied below, U_0 and U_1 can be the same potential function restricted to different regions of configurational space (i.e., two conformational states). The free energy difference between these two states is given by

$$e^{-\beta\Delta G} = \frac{Z[U_1(\vec{x})]}{Z[U_0(\vec{x})]} = \frac{\int d\vec{x}\; e^{-\beta U_1(\vec{x})}}{\int d\vec{x}\; e^{-\beta U_0(\vec{x})}}, \quad (1)$$

where $\beta = 1/k_BT$, and Z indicates a partition function, and integration is to be performed over the indicated potential function. Note that for implicit solvation, as is used in this study, the Gibbs free energy is equivalent to the Helmholtz free energy. Using the formalism of free energy perturbation,[46] we can rewrite eq 1 as

$$e^{-\beta\Delta G} = \frac{\int d\vec{x}\; (e^{-\beta[U_1(\vec{x})-U_0(\vec{x})]})\, e^{-\beta U_0(\vec{x})}}{\int d\vec{x}\; e^{-\beta U_0(\vec{x})}} = \langle e^{-\beta[U_1(\vec{x})-U_0(\vec{x})]}\rangle_0, \quad (2)$$

where the $\langle ...\rangle_0$ indicates an equilibrium average over U_0 configurations.

Voter's shifting strategy[41] improves the overlap between states (potential functions) without altering ΔG. Imagine shifting the origin of the configurational coordinates $\vec{x}$ in U_1 by a constant vector $\vec{C}$. This corresponds to a simple change of variables, $\vec{x} \rightarrow \vec{x} + \vec{C}$, and $Z[U_1(\vec{x} + \vec{C})] = Z[U_1(\vec{x})]$ remains unchanged because integration is performed over all space. Equation 2 can now be written as

$$e^{-\beta\Delta G} = \frac{\int d\vec{x}\; (e^{-\beta[U_1(\vec{x}+\vec{C})-U_0(\vec{x})]})\, e^{-\beta U_0(\vec{x})}}{\int d\vec{x}\; e^{-\beta U_0(\vec{x})}} = \langle e^{-\beta[U_1(\vec{x}+\vec{C})-U_0(\vec{x})]}\rangle_0. \quad (3)$$

In principle, eq 3 implies that one can arbitrarily shift the coordinates for the U_0 configurations needed for the average. (Shifting U_0 rather than U_1 produces an analogous result.) In practice, the shifting should be done to maximize the overlap between configurations in U_0 and U_1. Unfortunately, no *Cartesian* shift vector can bring two distinct molecular configurations into overlap. Therefore, we apply the Voter idea in *internal coordinates*, and we term the approach single-stage shifting.

To demonstrate the reasoning behind a shift in internal coordinates, consider the schematic torsional potentials shown in Figure 1a,b. Simulations for U_1 will mainly sample the large "trans" well at $\phi = 0$, while simulations for U_0 will mainly sample the large "gauche" well at $\phi \approx -150$. Shifting U_1 by a constant corresponding to the difference between the minima of the two potentials (i.e., roughly 150° for this example) produces Figure 1c, where the overlap between ϕ configurations is excellent. Such a shift does *not* alter the U_1 partition function

and thus ΔG is unaffected. Note that, below, we shift internal coordinates rather than potentials, but these are equivalent procedures.

In practice, the dimensionality of the system will be high and the shifting constant will not be as obvious as that in Figure 1. Reasonable methods for determining the shifting constant will be discussed in Section 2.3.

2.2. Utilizing Bidirectional Data. Bennett's Methods. In single-stage free energy perturbation (eq 3), only configurations from U_0 are evaluated using U_1. However, if simulations are performed in both states of interest, one could just as readily evaluate configurations from U_1 using U_0. Bennett showed that the "bidirectional" ($U_0 \rightarrow U_1$ and $U_1 \rightarrow U_0$) evaluations could be combined to minimize the uncertainty in ΔG.[45]

Bennett introduced both "iterative" and "acceptance-ratio" methods to utilize bidirectional data.[45] Generalizing the acceptance-ratio formulation to the shifting approach yields[41,45]

$$e^{-\beta\Delta G} = \frac{\langle \min(1.0,\, e^{-\beta[U_1(\vec{x}+\vec{C})-U_0(\vec{x})]}) \rangle_0}{\langle \min(1.0,\, e^{-\beta[U_0(\vec{x}-\vec{C})-U_1(\vec{x})]}) \rangle_1}. \quad (4)$$

Similarly, generalizing to the iterative method to shifted coordinates leads to the following relation[45]

$$\langle (1 + e^{-\beta[U_0(\vec{x})-U_1(\vec{x}+\vec{C})+\Delta G]})^{-1} \rangle_0 = \langle (1 + e^{-\beta[U_1(\vec{x})-U_0(\vec{x}-\vec{C})-\Delta G]})^{-1} \rangle_1. \quad (5)$$

Since ΔG is on both sides of eq 5, it must be solved in an iterative fashion. Equation 5, in its unshifted form, has been shown to be the optimal use of bidirectional data.[35,45,47,48]

Below we use the single-stage shifting method to calculate ΔG using eqs 3−5 with internal coordinates shifted by a constant $\vec{C}$, which is chosen to maximize the overlap between U_0 and U_1 states. For the systems studied below, leucine dipeptide and decaglycine, we find the iterative method of eq 5 to be the most efficient use of the raw data.

2.3. Practical Implementation in Molecular Systems. Determining a shifting constant $\vec{C}$ in eq 5 involves making a decision about the subset of coordinates to shift. There is generally a minimum number of coordinates needed. For example, for leucine dipeptide, we define the alpha and beta conformations using two backbone torsion angles; thus these two torsions *must* be shifted. In practice, it is also possible to shift too many coordinates, leading to steric clashes. This will be shown below for peptides.

Once the subset of shifted internal coordinates has been determined, the shift constant $\vec{C}$ must be calculated for all coordinates in the subset. In Figure 1, we chose to shift according to the minimum of U_0 and U_1. This leads to our first approach: after equilibrium simulation in each potential, find the lowest energy frame (snapshot) for both U_0 and U_1; then choose the constant vector $\vec{C}$ which aligns the two lowest-energy frames,

$$\vec{C} = \vec{y}_0^{\,\min} - \vec{y}_1^{\,\min}, \quad (6)$$

where $\vec{y}$ represents the subset of coordinates (e.g., only torsions) of the configuration $\vec{x}$. Strictly speaking, $\vec{C}$ is a vector of the same dimensionality as $\vec{x}$, with zero for every component not present in $\vec{y}$. Another reasonable choice for a shift constant is realized by generating a histogram for each coordinate in $\vec{y}$ for both U_0 and U_1. The shift constant is then chosen to align the peaks of these histograms,

$$\vec{C} = \vec{y}_0^{\,\text{mode}} - \vec{y}_1^{\,\text{mode}}, \quad (7)$$

Below, we shift using both of these choices for various internal coordinate subsets, and compare the results.

Procedurally, an estimate of ΔG is generated using the following steps:

1. Generate an equilibrium trajectory in both states of interest, i.e., using U_0 and U_1. Determine, as discussed above, the subset of coordinates to shift $\vec{y}$, and the shifting vector $\vec{C}$ using eq 6 or 7.

2. For each frame in the U_0 trajectory, shift the internal coordinates by $\vec{C}$ (e.g., $\phi_2 \rightarrow \phi_2 + 90.0$). Then, for each U_0 frame, $\vec{x}$, evaluate and record $U_1(\vec{x} + \vec{C}) - U_0(\vec{x})$.

3. Repeat step 2 for each frame in the U_1 ensemble with one important difference: the frames must be shifted in the opposite direction (e.g., $\phi_2 \rightarrow \phi_2 - 90.0$), yielding $U_0(\vec{x} - \vec{C}) - U_1(\vec{x})$.

4. Solve eq 3, 4, and/or 5 to determine a estimate.

Below, we utilize this process to generate multiple ΔG estimates, after which the mean ΔG and standard deviation are calculated.

When shifting internal coordinates care must be taken to shift the coordinates so that the partition function remains unchanged. Using the standard internal coordinates for bond length r, bond angle θ, and dihedrals ω, a differential volume element in configuration space is given by

$$d\vec{x}_1 = r_1^2 dr_1 \sin\theta_1 d\theta_1\, d\omega_1 = d(r_1^3/3)\, d(\cos\theta_1)\, d\omega_1, \quad (8)$$

where $\vec{x}_1$ represents one possible set of internal coordinates. Equation 8 holds for every possible set of internal coordinates and thus implies that the shifting of internal coordinates must be done according to simple rules. Bond lengths must be shifted according to the cube of the length: $r^3 \rightarrow r^3 - r_c^3$, where r_c is the bond length shifting constant (i.e., one component of the vector $\vec{C}$) found either by the peaks from histograms of r^3 and using eq 7, or by comparing minimum energy frames and using eq 6. Similarly, bond angles must be shifted using $\cos\theta \rightarrow \cos\theta - \cos\theta_c$, and dihedrals are shifted via $\omega \rightarrow \omega - \omega_c$.

3 Results for Peptides

To test the effectiveness of the single-stage shifting method, we performed all-atom simulations of leucine dipeptide (ACE-$(leu)_2$-NME) and decaglycine (ACE-$(gly)_{10}$-NME). Both peptides were simulated using the TINKER Version 4.2 molecular dynamics package.[49] The peptides were solvated implicitly using the generalized Born surface area (GBSA) approach,[50] and Langevin dynamics were utilized with the friction coefficient set to that of water (91.0 ps^{-1}). A time step of 1.0 fs was chosen for all simulations. Leucine dipeptide was maintained at 500.0 K (to enable independent verification of our result) and utilized the CHARMM27 force field parameter set.[51] Decaglycine was maintained at 300.0 K and used the AMBER96 force field parameter set[52] for comparison with Ref 37.

3.1. Leucine Dipeptide. Leucine dipeptide (ACE-$(leu)_2$-NME) was chosen as a test system because (i) it possesses some of the complexity of a large molecule: four backbone torsions and eight side-chain dihedrals; and (ii) it is small enough to allow for very long simulation times. Long simulation times are necessary for an unbiased, independent determination of the conformational population (hence ΔG), providing a strict test of the single-stage shifting method.

TABLE 1: Conformational Free Energy Differences ΔG for Alpha → Beta Conformations for Leucine Dipeptide in GBSA Solvent Using Several Variations of the Single-stage Shifting Approach[a]

shifting approach	shifted coordinates	iterative ΔG (kcal/mol)	accept. ratio ΔG (kcal/mol)
peak of histogram (eq 7)	backbone torsions only	1.08 (0.15)	1.02 (0.13)
	all torsions	0.56 (0.87)	0.55 (0.86)
	all torsions and bond angles	0.76 (2.64)	1.01 (3.25)
	all internal coordinates	0.78 (2.63)	1.09 (3.26)
lowest energy frames (eq 6)	backbone torsions only	1.29 (0.35)	1.30 (0.39)
	all torsions	0.99 (0.66)	0.98 (0.66)
	all torsions and bond angles	3.36 (4.75)	6.48 (9.45)
	all internal coordinates	4.50 (7.16)	8.80 (14.32)

[a] Each shifting ΔG result is an average of 16 independent 2.0 ns estimates of ΔG with standard deviation shown in parentheses. The iterative results use eq 5 and the acceptance-ratio results use eq 4. For comparison, the value of ΔG obtained from long simulation (4.0 μs) is 0.95 (0.05) kcal/mol. The entropy difference ΔS, found by single-stage shifting, was zero within uncertainty, consistent with the value obtained by long simulation. This table also demonstrates that shifting too many coordinates (i.e., bond angles and lengths) leads to steric clashes, and thus a larger uncertainty for ΔG.

For leucine dipeptide we calculated the free energy difference ΔG and entropy difference ΔS for the alpha → beta conformational change. We defined the alpha and beta conformations using two internal backbone torsions, namely, for alpha: $-145 < \phi_2 < -25$ and $-125 < \psi_1 < -5$; and for beta: $-160 < \phi_2 < -40$ and $70 < \psi_1 < -170$. A temperature of 500 K was chosen to enable repeated crossing of the free energy barrier between the alpha and beta conformations. At 500 K, leucine dipeptide switches between alpha and beta conformations at a rate of around 2.5 transitions per nanosecond with GBSA solvation. Note that this high temperature is required only to obtain an unbiased estimate of ΔG; our single-stage shifting approach works equally well at lower temperatures.

To obtain an independent and unbiased value of ΔG, four 1.0 μs simulations were performed yielding around 2500 transition events per trajectory. The four trajectories were then used to calculate ΔG via the definition

$$\Delta G \equiv -\frac{1}{\beta}\ln\left(\frac{N_{\text{beta}}}{N_{\text{alpha}}}\right), \tag{9}$$

where N_{alpha} and N_{beta} are, respectively, the number of dynamics steps the system was in the alpha and beta conformations. The unbiased value of ΔG for alpha → beta was found to be $\Delta G = 0.95 \pm 0.05$ kcal/mol, where the value of ΔG given is the average of the four 1.0 μs estimates with uncertainty given by the standard deviation.

The unbiased entropy difference was calculated using eq 9 via $T\Delta S = \Delta U_{\text{avg}} - \Delta G$ where ΔU_{avg} is the average of $U_{\text{beta}} - U_{\text{alpha}}$. We found that $T\Delta S$ was zero within uncertainty. This is consistent with our observation that the fluctuating degrees of freedom for leucine dipeptide are mainly the side-chain torsions and thus do not change dramatically between the alpha and beta conformations. Given the overall ΔG value, the alpha conformation is favored due to intramolecular attractions.

With an unbiased value of ΔG from eq 9, it is possible to test various implementations of the single-stage shifting method. To this end, we simulated leucine dipeptide in the alpha and beta conformations, generating four 1.0 ns trajectories for each conformation. Each of the trajectories were obtained by constraining the backbone torsions to stay within the defined ranges for alpha and beta, given above. The constraining force was zero if the torsion angles were within the allowed range and harmonic otherwise. Frames were saved every 0.1 ps yielding 10 000 frames per trajectory.

The four trajectories provide sixteen independent single-stage shifting estimates of ΔG and ΔS using all possible pairings. The results for ΔG are summarized in Table 1 where we used both the lowest energy frame and histogram peak shifting approaches (Section 2.3) to estimate ΔG for leucine dipeptide. The value of ΔS was found to be zero within uncertainty, consistent with the value found using eq 9. For each shifting approach, we tested four sets of shifting coordinates: backbone torsions only, all torsions, all torsions and bond angles, or all internal coordinates (torsions, angles, and bond lengths). In addition, we also show results using both the iterative method of eq 5 and the acceptance-ratio method of eq 4. The values for ΔG are averages with standard deviations shown in parentheses. Our single-stage shifting results in Table 1 agree well with the unbiased estimate given above. The table also show that shifting torsions results in a small uncertainty, while including bond angles and lengths makes the ΔG estimate less certain.

In all of our simulations, we have found that shifting torsions (either backbone only, or all torsions), and using the iterative method of eq 5, has consistently provided accurate results. Also, as demonstrated in Table 1, lower uncertainty is obtained by shifting according to histogram peaks and using eq 7 rather than shifting by the lowest energy frames.

We also stress the importance of using bidirectional data to determine ΔG for conformational equilibria. To this end, in Figure 2, we employed the single-stage shifting method using three data analysis protocols: single-stage free energy perturbation in both directions (eq 3),[46] Bennett's acceptance-ratio method (eq 4), and Bennett's iterative method (eq 5).[45] The superior convergence properties of the iterative method can be seen in Figure 2. The solid horizontal black line represents the independent, 4.0 μs ΔG value obtained by using eq 9. The data are ΔG estimates using the single-stage shifting method, where backbone torsions were shifted according to the histogram peaks. The data for the figure was generated using a single trajectory in each of the alpha and beta conformations. The red curve was generated from Bennett's acceptance-ratio method, the green dashed curve is free energy perturbation in the forward direction (i.e., $\Delta G_{\text{alpha}\to\text{beta}}$), and the green solid curve is free energy perturbation in the reverse direction (i.e., $-\Delta G_{\text{beta}\to\text{alpha}}$). Finally the solid blue curve is Bennett's iterative method. It is clear from the figure that using bidirectional data is very important to the success of our single-stage shifting method. Further, Bennett's iterative method is shown to converge more quickly than the other methods.

Figure 3 demonstrates the efficiency of the single-stage shifting method compared to long simulation and use of eq 9. The horizontal black line is the unbiased ΔG from long (4.0 μs) simulation. The curve shows the average (blue squares) and standard deviation (errorbars) of the single-stage shifting method where backbone torsions were shifted by the histogram peaks, and the iterative method of eq 5 was used to analyze the data.

TABLE 2: Conformational Free Energy Differences ΔG and Entropy Differences $T\Delta S$ for Alpha → Extended Conformations of Decaglycine in GBSA Solvent Using Several Shifting Approaches[a]

shifting approach	shifted coordinates	iterative ΔG (kcal/mol)	iterative $T\Delta S$ (kcal/mol)
peak of histogram (eq 7)	backbone torsions only	−12.39 (0.47)	9.75 (0.66)
	all torsions	−12.59 (0.82)	9.95 (0.93)
	all torsions and bond angles	−12.77 (2.94)	10.12 (2.97)
	all internal coordinates	−12.75 (3.11)	10.10 (3.13)
lowest energy frames (eq 6)	backbone torsions only	−11.97 (1.59)	9.34 (1.71)
	all torsions	−12.12 (2.57)	9.48 (2.76)
	all torsions and bond angles	−18.53 (11.14)	15.89 (10.93)
	all internal coordinates	−24.79 (14.03)	22.15 (13.86)

[a] Each shifting result is an average of 16 independent 2.0 ns estimates with standard deviation shown in parentheses, using the iterative method of eq 5. Acceptance ratio results (data not shown) had a much larger uncertainty than the iterative method. This data clearly shows that lower uncertainty is obtained by shifting via histogram peaks rather than by the lowest energy frame. Also, shifting nontorsional coordinates (i.e., bond angles and lengths) dramatically increases uncertainty

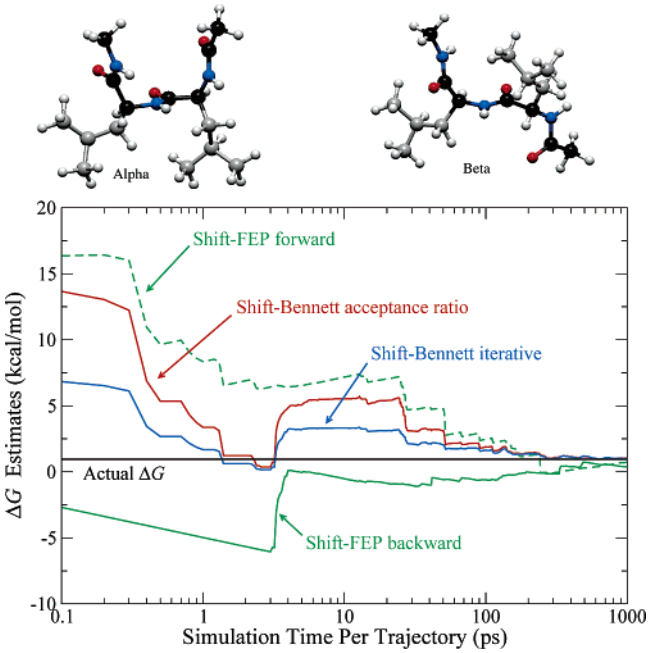

Figure 2. Comparison between several shifting methods for the conformational change in free energy for alpha → beta of leucine dipeptide in GBSA solvent. The alpha (top left, "cis-like") and beta (top right, "trans-like") conformations are also shown, with the backbone carbons colored black. Two trajectories (one in each conformation) were analyzed using several protocols of the single-stage shift technique. The solid horizontal black line represents the 4.0 μs value of ΔG obtained by using eq 9. The curves are obtained using the single-stage shifting method where backbone torsions were shifted according to the peak of the histogram (i.e., lowest uncertainty ΔG estimate in Table 1). The red curve was generated from Bennett's acceptance-ratio method, the green dashed curve is free energy perturbation in the forward direction (i.e., $\Delta G_{\text{alpha}\rightarrow\text{beta}}$), and the green solid curve is free energy perturbation in the reverse direction (i.e., $-\Delta G_{\text{beta}\rightarrow\text{alpha}}$). The solid blue curve is Bennett's iterative method. The figure clearly suggests the superior convergence properties of the iterative method for our single-stage shifting data.

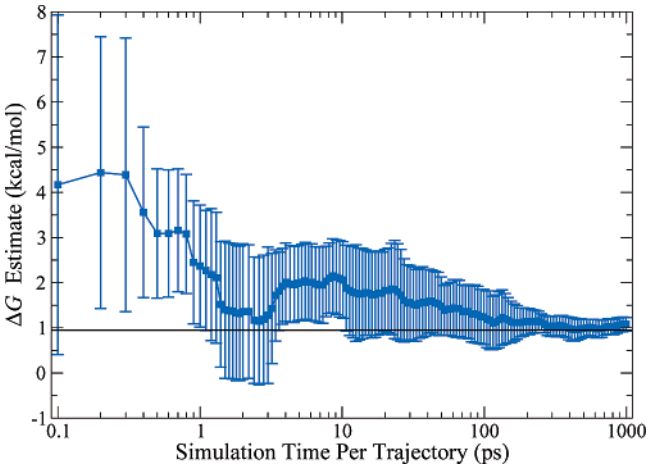

Figure 3. Conformational free energy difference ΔG for alpha → beta conformations of leucine dipeptide in GBSA solvent, shown as a function of the simulation time. The horizontal black line shows the known ΔG obtained using 4.0 μs of simulation and eq 9. The data were obtained using the single-stage shifting method where backbone torsions were shifted according to the peak of the histogram, and the iterative method of eq 5 was used to analyze the data. The average ΔG estimates (blue squares) and standard deviations (errorbars) were calculated using sixteen independent estimates of ΔG. Using the single-stage shifting method, with only 30 ps of simulation (15 ps in alpha and 15 ps in beta), the average estimate of ΔG is within 1.0 kcal/mol of the known value with a standard deviation of less than 1.0 kcal/mol. This may be compared with unconstrained simulations, in which transitions between alpha and beta conformations occurred, on average, every 400 ps.

In our unconstrained simulations, leucine dipeptide switched between the alpha and beta conformations, on average, once every 400 ps. Using our single-stage shifting method, with only 30 ps of simulation (15 ps in alpha and 15 ps in alpha), a reasonably accurate and precise value of ΔG can be obtained.

3.2. Decaglycine. We also applied the single-stage shifting method to decaglycine (ACE-$(\text{gly})_{10}$-NME), predicting the conformational ΔG and entropy difference ΔS for alpha → extended conformations. We again defined the alpha and extended conformations by the internal backbone torsions (i.e., excluding ϕ_1 and ψ_{10}), namely, for alpha: $-115 < \phi < 5$ and $-115 < \psi < 5$; and for extended: $120 < \phi < -120$ and $120 < \psi < -120$. Previous ΔG and ΔS calculations of decaglycine in a vacuum were performed by Karplus and Kushick[39] and quite recently by Cheluvaraja and Meirovitch.[37] Apparently, decaglycine's conformational ΔG and ΔS have not previously been computed in implicit solvent.

To calculate ΔG and ΔS, four trajectories in each conformation were generated using the simulation parameters defined previously. Thus, 16 independent ΔG and ΔS estimates can be calculated. Each trajectory was 1.0 ns in length with a frame saved every 0.1 ps yielding 10 000 frames per trajectory, although this may be quite suboptimal; see Section 4. As with leucine dipeptide, each of the 16 ΔG and ΔS estimates was generated using various shifting approaches.

Table 2 shows the results of our calculation of ΔG and ΔS for the alpha → extended conformational change using various shifting approaches. The entropy change was estimated via $T\Delta S = \Delta U_{\text{avg}} - \Delta G$ where ΔU_{avg} is the average of $U_{\text{ext}} - U_{\text{alpha}}$. Acceptance ratio estimates had a much larger uncertainty than the iterative method estimates and thus are not included in the table. The results clearly demonstrate that shifting by histogram

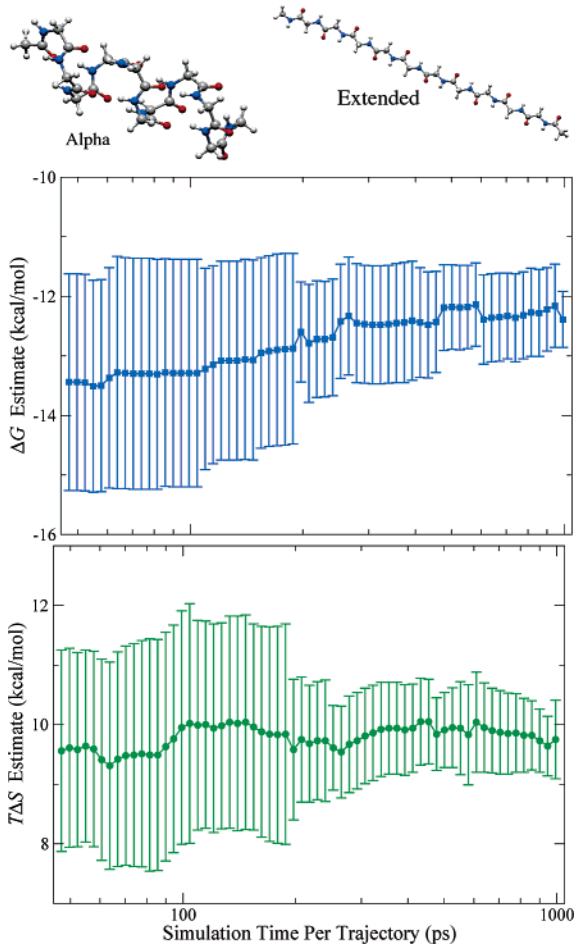

Figure 4. Conformational free energy difference ΔG and entropy difference $T\Delta S$ for the alpha → extended conformational change in decaglycine with GBSA solvent. The alpha (top left) and beta (top right) conformations are also shown. The data were obtained from eight independent simulations: four constrained to be in alpha and four in extended. From the eight simulations, 16 independent estimates of ΔG were obtained, and the averages (blue squares) and standard deviations (error bars) were calculated as a function of simulation time. Each ΔG estimate was generated using the end point shifting method by shifting backbone torsions via histogram peaks and using the iterative method of eq 5.

peaks provides a higher level of precision than shifting by the lowest energy frame. Also, as with leucine dipeptide, lower uncertainty is obtained when shifting torsions only (i.e., not bond angles and lengths).

The results in Table 2 suggest that the "compensating" role of the entropy is vital for an accurate ΔG calculation in GBSA solvent. Our studies of decaglycine in a vacuum (data not shown), as well those of other groups[37,39] show that the alpha conformation is more stable than extended, due mainly to energetics. However, our results suggest that, with the addition of (implicit) solvent, the energy difference between the two conformations becomes small enough that the entropy term dominates ΔG, to the degree that the extended conformation is more stable than alpha.

Figure 4 shows ΔG and $T\Delta S$ as a function simulation time. The data points are the average of the sixteen independent estimates with standard deviation given by the error bars for both ΔG (blue squares) and $T\Delta S$ (green circles). These estimates were obtained using the iterative method, and shifting the backbone torsions by the histogram peak. The figure demonstrates the apparent convergence of the ΔG and $T\Delta S$ estimates. Note that it would be impractical to obtain an independent estimate for decaglycine (as we did with leucine dipeptide) because of the required simulation times.

In the current implementation, a total of 8 ns of simulation was required to obtain a reasonably accurate and precise estimate of ΔG. To our knowledge, no multistage calculation has ever been attempted on this system, undoubtedly due to the computational expense. Nevertheless, we believe additional optimization of our current shifting protocol will be possible, as we now discuss.

Finally, we note that, as in any ΔG computation, our results reflect the definitions chosen for the alpha and extended states.

4. Further Optimization of the Single-Stage Shifting Approach

While the present results indicate that peptide conformational equilibria can be determined by sub-nanosecond simulations, several promising avenues for optimization have not been explored. We briefly sketch several possible approaches for improving efficiency, including combining the shifting approach with traditional staged calculations.

It is useful to consider the upper limit for the computational cost of the single-stage shifting procedure. The maximum is essentially twice that of equilibrium simulation, provided the method is hard-wired into the molecular dynamics program, due to the extra energy call that must be made for each shifted frame. In the current study, the shifting procedure was scripted external to the simulation program (TINKER), and thus the cost for trajectory analysis was high, limiting the number of frames per trajectory to 10 000. We found that, for this fixed number of frames per trajectory, the simulation time between frames had very little affect on the ΔG estimate. Thus, we feel that substantial increases in efficiency could be realized by utilizing frames more frequently in the ΔG calculation, i.e., fewer time steps between frames. (It is worthwhile to recall that interactions change enough over a single time step to require recalculation of forces.) Ultimately, then, reliable and accurate ΔG calculations should take no more time than required to sample the equilibrium ensemble in a given state (the minimum time for any ΔG method).

Additionally, the shifting procedures explored in this report ignore correlation between coordinates, such as those known from Ramachandran plots, where backbone torsions ϕ and ψ do not vary independently. Ramachandran plots, moreover, average over many residues, which individually are likely more correlated. To motivate more general shifts, consider one state where a certain pair of ϕ,ψ angles inhabit a predominantly vertical region of Ramachandran space, while the other state populates a region with a very different orientation. In such a case, simple shifts alone (e.g., those in eqs 6 and 7) will not maximize overlap to the extent that a combined shift and (partition-function-preserving) rotation would. Further, if one oblong well is very narrow and the other well is very broad, then coordinate scaling (contraction/expansion) should also be performed (see also Refs 43, 44). In general the coordinates can be linearly scaled, rotated and/or translated using a constant matrix $\mathbf{A}$ (i.e., $\vec{x} \rightarrow \mathbf{A}\vec{x} + \vec{C}$), and eqs 3−5 must be generalized to account for the matrix $\mathbf{A}$. More complex, nonlinear transformations are also possible, but may not be practical.

In larger systems than those considered here (e.g., whole proteins), the gain in overlap due to the internal coordinate shift may prove insufficient to permit reliable single-stage computation of ΔG values. In such cases, it may be advantageous to combine the single-stage shifting approach with multistage methodology. To do so, a path connecting the two states of

interest can be defined (e.g., refs 53−55), and independent trajectories can be generated at intermediate stages along the path. Then, between each successive intermediate stage, the incremental free energy difference (δG) is estimated using the shifting protocol outlined in Section 2.3. The δG estimates can then be summed to obtain the full ΔG for the complete path.

5. Extension to "Alchemical" Calculations and Explicitly Solvated Systems

It is possible to extend the formalism of the single-stage shifting method beyond conformational ΔG calculation for implicitly solvated molecules. In this section we outline the potential for the single-stage shifting method to be used for "alchemical" mutations, which are the basis for relative binding affinities and solubilities,[56] and on explicitly solvated systems.

For alchemical mutation, two distinct potential energy functions U_0 and U_1, one for each molecule, are used in eqs 3−5. While the mathematical formalism is unchanged from conformational ΔG calculations, the difficulty in alchemical mutations lies in determining the shifting vector $\vec{C}$, since the number of degrees of freedom for U_0 and U_1 are different, in general. (For conformational ΔG calculations, such as those above, the number of degrees of freedom for U_0 and U_1 are always the same.) This difficulty can be overcome by introducing "dummy" coordinates as in ref 54. Although dummy coordinates will change the absolute free energy values, use of a thermodynamic cycle guarantees that the free energy difference ΔG will remain unchanged. Thus, accurate *relative* binding affinities and solubilities can be obtained.[54]

It may also be possible, though likely more challenging, to extend the single-stage shifting approach to explicitly solvated molecular systems. The method must then be generalized to include nonstandard *inter*molecular "external" coordinates. This can, in principle, be accomplished by introducing auxiliary vectors to describe the location and orientation (e.g., Euler angles) of each solvent or solute molecule. The necessary shifting vector $\vec{C}$ will now include the full set of intra- and intermolecular coordinates. Correlations between molecular motions can be included as sketched in Section 4.

6. Conclusion

In a study of peptide equilibria, we have demonstrated a simple method for substantially overcoming the overlap problem in calculating free energy differences (ΔG) and entropy difference (ΔS) between conformational states. The new single-stage shifting method utilizes a shift in internal coordinates to improve the overlap between configurations, motivated by Voter's study in ref 41. The approach requires only simulation in the two states of interest without the need for "staged" intermediate calculations. Bennett's iterative method[45] is used to efficiently calculate a ΔG value from the raw data.

We tested the single-stage shifting approach on two peptides, obtaining excellent results with sub-nanosecond simulation times. First, for leucine dipeptide in implicit solvent, we accurately calculated the conformational ΔG for alpha → beta conformations, judging by nearly perfect agreement with a 4.0 μs simulation. The entropy difference ΔS was found to be nearly zero, also consistent with long simulation. The single-stage shifting method was then used to predict the conformational ΔG and ΔS for alpha → extended conformations of decaglycine in implicit solvent, apparently for the first time. We find that, with implicit solvent, the extended conformation of decaglycine is favored over alpha, due mainly to the entropy gain in the extended state. By contrast, in a vacuum, the alpha conformation is preferred due mainly to the strongly favorable intramolecular interactions. It must be borne in mind that our quantitative results necessarily depend on our state definitions (see Section 3).

While the present report describes a single type of application of the single-stage shifting approach (to conformational equilibria), we believe the idea will find quite broad applications, in part due to the substantial potential for further optimization. We have therefore discussed optimization of the method and application to "alchemical" mutations (for relative binding affinities). Inclusion of explicit solvent is also possible, in principle, but may represent a practical challenge. The single-stage shifting approach may also be combined with multistage simulation, allowing further optimization. We are currently exploring these ideas.

Acknowledgment. We would like to thank Carlos Camacho, Ronald White, Srinath Cheluvaraja, and Edward Lyman for many fruitful discussions. Funding for this research was provided by the Department of Environmental and Occupational Health and the Department of Computational Biology at the University of Pittsburgh, and the National Institutes of Health (Grants ES007318 and GM070987).

References and Notes

(1) DeGrado, W. F.; Nilsson, B. O. *Curr. Opin. Struct. Biol.* **1997**, *7*, 455−456.
(2) Lazar, G. A.; Marshall, S. A.; Plecs, J. J.; Mayo, S. L.; Desjarlais, J. R. *Curr. Opin. Struct. Biol.* **2003**, *13*, 513−518.
(3) Jorgensen, W. L. *Science* **2004**, *303*, 1813−1818.
(4) Sotriffer, C.; Klebe, G.; Stahl, M.; Bohm, H.-J. *Burger's Medicinal Chemistry and Drug Discovery*, 6th ed.; Wiley: New York, 2003; volume 1.
(5) Oostenbrink, B. C.; Pitera, J. W.; van Lipzip, M. M.; Meerman, J. H. N.; van Gunsteren, W. F. *J. Med. Chem.* **2000**, *43*, 4594−4605.
(6) Grossfield, A.; Ren, P.; Ponder, J. W. *J. Am. Chem. Soc.* **2003**, *125*, 15671−15682.
(7) Pitera, J. W.; van Gunsteren, W. F. *J. Phys. Chem. B* **2001**, *105*, 11264−11274.
(8) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. *Proc. Nat. Acad. Sci. U.S.A.* **1994**, *91*, 7673−7677.
(9) Kolmodin, K.; Aqvist, J. *FEBS Lett.* **2000**, *465*, 8−11.
(10) Shoichet, B. K. *Nature* **2004**, *432*, 862−865.
(11) Wei, B. Q.; Weaver, L. H.; Ferrari, A. M.; Matthews, B. W.; Shoichet, B. K. *J. Mol. Biol.* **2004**, *337*, 1161−1182.
(12) Halperin, I.; Wolfson, H.; Nussinov, R. *Proteins* **2002**, *47*, 409−443.
(13) Brooijmans, N.; Kuntz, I. D. *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 335−373.
(14) Trosset, J. Y.; Scheraga, H. A. *J. Comput. Chem.* **1999**, *20*, 412−427.
(15) Patel, S.; Mackerell, A. D., Jr.; Brooks, C. L., III. *J. Comput. Chem.* **2004**, *25*, 1504−1514.
(16) Kaminski, G. A.; Stern, H. A.; Berne, B. J.; Friesner, R. A.; Cao, Y. X.; Murphy, R. B.; Zhou, R.; Halgren, T. A. *J. Comput. Chem.* **2002**, *23*, 1515−1531.
(17) Lamoureux, G.; B., R. *J. Chem. Phys.* **2003**, *119*, 3025−3039.
(18) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 9450−9453.
(19) Yang, W.; Bitetti-Putzer, R.; Karplus, M. *J. Chem. Phys.* **2004**, *120*, 2618−2628.
(20) McCammon, J. A. *Curr. Opin. Struct. Biol.* **1991**, *2*, 96−200.
(21) Jorgensen, W. L.; Ravimohan, C. *J. Chem. Phys.* **1985**, *83*, 3050−3054.
(22) Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V. S. *J. Chem. Phys.* **2003**, *119*, 5740−5761.
(23) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690−2693.
(24) Zuckerman, D. M.; Woolf, T. B. *Phys. Rev. Lett.* **2002**, *89*, 180602.
(25) Hummer, G.; Szabo, A. *Proc. Nat. Acad. Sci. U.S.A.* **2001**, *98*, 3658−3661.
(26) Liphardt, J.; Dumont, S.; Smith, S. B.; Tinoco, I.; Bustamante, C. *Science* **2002**, *296*, 1832−1835.
(27) Park, S.; Schulten, K. *J. Chem. Phys.* **2004**, *120*, 5946−5961.
(28) Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. *J. Chem. Phys.* **2003**, *119*, 3559−3566.
(29) Zuckerman, D. M.; Woolf, T. B. *Chem. Phys. Lett.* **2002**, *351*, 445−453.

(30) Ytreberg, F. M.; Zuckerman, D. M. *J. Chem. Phys.* **2004**, *120*, 10876−10879; **2004**, *121*, 5022−5023.
(31) Ytreberg, F. M.; Zuckerman, D. M. *J. Comput. Chem.* **2004**, *25*, 1749−1759.
(32) Hummer, G.; Szabo, A. *J. Chem. Phys.* **1996**, *105*, 2004−2010.
(33) Mordasini, T. Z.; McCammon, J. A. *J. Phys. Chem. B* **2000**, *104*, 360−367.
(34) Beveridge, D.; DiCapua, F. *Annu. Rev. Biophys. Biophys. Chem.* **1989**, *18*, 431−492.
(35) Lu, N.; Kofke, D. A.; Woolf, T. B. *J. Comput. Chem.* **2004**, *25*, 28−40.
(36) Oostenbrink, C.; van Gunsteren, W. F. *J. Comput. Chem.* **2003**, *24*, 1730−1739.
(37) Cheluvaraja, S.; Meirovitch, H. *Proc. Nat. Acad. Sci.* **2004**, *101*, 9241−9246.
(38) Chang, C. E.; Gilson, M. K. *J. Am. Chem. Soc.* **2004**, *126*, 13156−13164.
(39) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325−332.
(40) Stoessel, J. P.; Nowak, P. *Macromolecules* **1990**, *23*, 1961−1965.
(41) Voter, A. F. *J. Chem. Phys.* **1985**, *82*, 1890−1899.
(42) Bruce, A. D.; Wilding, N. B.; Ackland, G. J. *Phys. Rev. Lett.* **1997**, *79*, 3002−3005.
(43) Miller, M. A.; Reinhardt, W. P. *J. Chem. Phys.* **2000**, *113*, 7035−7046.
(44) Jarzynski, C. *Phys. Rev. E* **2002**, *65*, 046122.
(45) Bennett, C. H. *J. Comput. Phys.* **1976**, *22*, 245−268.
(46) Zwanzig, R. W. *J. Chem. Phys.* **1954**, *22*, 1420−1426.
(47) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361−2366.
(48) Shirts, M. R.; Bair, E.; Hooker, G.; Pande, V. S. *Phys. Rev. Lett.* **2003**, *91*, 140601.
(49) Ponder, J. W.; Richard, F. M. *J. Comput. Chem.* **1987**, *8*, 1016−1024, http://dasher.wustl.edu/tinker/.
(50) Still, W. C.; Tempczyk, A.; Hawley, R. C. *J. Am. Chem. Soc.* **1990**, *112*, 6127−6129.
(51) Foloppe, N.; MacKerell, A. D. *J. Comput. Chem.* **2000**, *21*, 86−104.
(52) Kollman, P.; Dixon, R.; Cornell, W.; Fox, T.; Chipot, C.; Pohorille, A. *Comput. Sim. Biomol. Syst.* **1997**, *3*, 83−96.
(53) Tobias, D. J.; Brooks, C. L. *J. Phys. Chem.* **1992**, *96*, 3864−3870.
(54) Shobana, S.; Roux, B.; Andersen, O. S. *J. Phys. Chem. B* **2000**, *104*, 5179−5190.
(55) Kong, X.; Brooks, C. L. *J. Chem. Phys.* **1996**, *105*, 2414−2423.
(56) Tembe, B. L.; McCammon, J. A. *Comput. Chem. (Oxford)* **1984**, *8*, 281−283.