Calculation of conformational free energies with the focused confinement method

Arjan van der Vaart,* Paul B. Orndorff, and Sang T. Le Phan

Department of Chemistry, University of South Florida, Tampa, FL 33620

E-mail: avandervaart@usf.edu

Phone: +1-813-974-8762. Fax: +1-813-974-3203

Abstract

We introduce the focused confinement method, a reaction coordinate-free simulation approach for the calculation of conformational free energies. These are obtained in a series of restrained simulations that transform part of the molecule of interest to independent harmonic oscillators resulting in mixed harmonic-anharmonic states. It is shown that the free energy difference between these mixed states can be readily calculated through the construction of chimeric trajectories. By focusing the confinement to the conformationally active region, the method requires fewer restrained simulations than the traditional confinement method, which eases the treatment of large systems. The accuracy and efficiency of the method is demonstrated for implicitly and explicitly solvated systems.

1 Introduction

By providing a quantitative measure of the likelihood of states, conformational free energies are key to the understanding of protein stability, allostery and dynamics. It is often not possible to calculate conformational free energy differences from unbiased simulations

due to limited sampling. Instead, free energy differences are generally obtained from enhanced sampling simulations, in which the Hamiltonian and/or temperature is modified in order to enforce multiple transitions between states.[1–11] Since this biasing can be accounted for by reweighting techniques, such as the multistate Bennett acceptance ratio estimator (MBAR),[12] accurate free energy differences can be computed from these simulations.

While many enhanced sampling techniques exist,[1–11] the confinement method (CM)[13–19] seems particularly appealing for the calculation of conformational free energy differences. Rather than connecting the conformations of interest by a geometrical path, CM connects them to harmonic oscillator (HO) states for which the free energies are known analytically. By foregoing a geometrical pathway, CM does not require knowledge of the reaction coordinate or order parameters that separate the conformations of interest; collective variables that are typically difficult to identify for protein conformational changes. Moreover, along geometrical pathways the conformations of interest are generally separated by free energy barriers, which slow down sampling, but these barriers are not encountered in CM. CM does not waste time on areas outside the basins of interest; in addition, by foregoing a geometrical path, free energy differences between highly dissimilar states can be readily obtained. While the method was originally developed for vacuum and implicit solvent simulations, it now also works in explicit solvent.[20,21] Comparisons show excellent agreement with umbrella sampling for a series of systems in implicit and explicit water.[19–21]

Despite these attractive features, treatment of large systems is costly and very few CM applications to proteins have been reported;[22,23] and none in explicit solvent. In CM all protein atoms need to be transformed to independent three-dimensional HOs. For large systems this is problematic for two reasons. First, the free energy cost of this transformation will grow with system size, since more atoms need to be transformed in larger systems. This means that larger systems will require simulations at more intermediate restraint strengths to accurately bridge the free energy gap. Moreover, each protein degree of freedom needs to be fully equilibrated, which means that larger systems typically require longer simulations,

especially at low restraint strengths. While the latter can be mitigated by applying mild temperature replica exchange,[19] the unfavorable scaling of CM severely impedes treatment of proteins. Here we introduce the focused confinement method to solve this deficiency and significantly ease the treatment of large systems.

2 Theory

Focused confinement exploits the fact that conformational changes are generally restricted to a (small) portion of the molecule. This is exemplified by hydrophobin I, a protein for which the conformational motion is limited to a loop while the rest of the molecule remains in the same conformation (Fig. 1).[24] Such loop motions are commonly observed in proteins. Even for this relatively simple conformational change, the reaction coordinate or order parameters are not easily identified *a priori*. In proteins that display larger conformational changes like domain motions, significant parts of the protein also remain in the same conformation. This observation suggests that computation time can be saved by focusing the confinement to the conformationally active region.

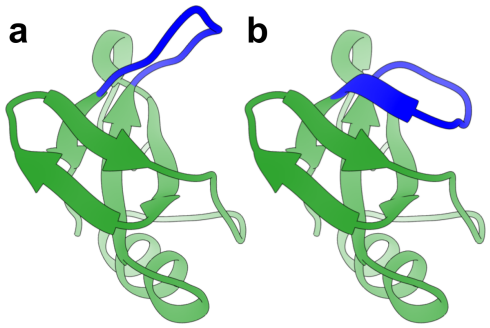

Figure 1: Conformational states of hydrophobin I.[24] a) Open. b) Closed.

This approach is illustrated and contrasted with traditional CM in Fig. 2. Solute AB has two conformations: A_aB_x and A_aB_y. The structural difference between these conformations is limited to the B region of the molecule, which is either in the x or y conformation, while region A has the same a conformation in both states. The conformationally active (B) or inactive (A) regions are not necessarily contiguous, and could in principle be interspersed

throughout the entire molecule. In addition, the conformationally inactive region is not necessarily rigid, and could be floppy as well; it just samples the same configurational space in the A_aB_x and A_aB_y conformations.

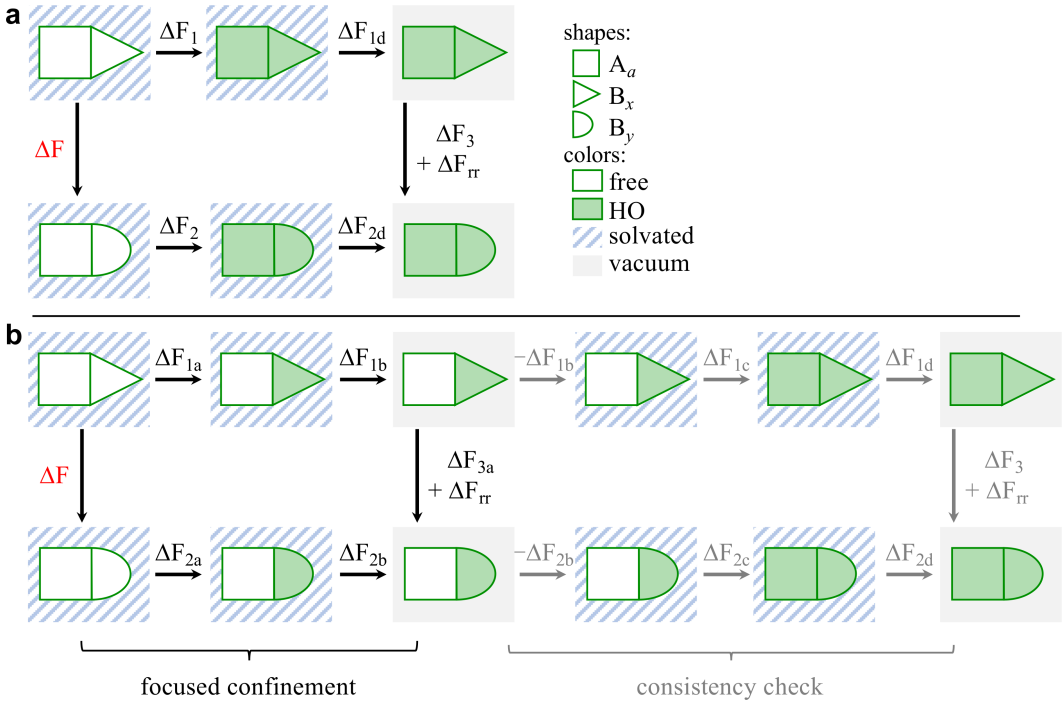

Figure 2: Confinement and focused confinement methods. a) Thermodynamic cycle of the confinement method. b) Thermodynamic cycle of the focused confinement method (black arrows). Grey arrows indicate the extension of the thermodynamic cycle to allow for a consistency check with CM; this consistency check is normally not needed. The meaning of the free energy components is explained in the text.

Confinement method

In CM[13–19] the free energy difference between the A_aB_y and A_aB_x conformations follows from the thermodynamic cycle of Fig. 2a and is given by

$$\Delta F = \Delta F_1 - \Delta F_2 + \Delta F_{1d} - \Delta F_{2d} + \Delta F_3 + \Delta F_{rr}. \tag{1}$$

ΔF_1 is the free energy of confinement: the free energy of transforming A_aB_x to the HO state $(A_aB_x)^{HO} = A_a^{HO}B_x^{HO}$. Similarly, ΔF_2 is the free energy of confining A_aB_y. These free

energies are obtained in a series of molecular dynamics (MD) simulations by augmenting the normal, unbiased potential by increasingly large restraining potentials. Denoting the total mass of the solute by M, the solute atomic masses by m_i and solute instantaneous atomic positions by r_i, the restraining potential for the j$^{\text{th}}$ simulation is given by $U_{\text{conf}} = 2(\pi\nu_j)^2 \sum_i m_i |r_i - r_i^0|^2 = 2M(\pi\nu_j)^2\rho_m^2$, where the sum is over all solute atoms and the strength of the restraint is given by frequency ν_j. The restraints are centered at atomic positions r_i^0 of reference structure $(\text{A}_a\text{B}_x)^0 = a^0x^0$ or $(\text{A}_a\text{B}_y)^0 = a^0y^0$, respectively, which is generally taken from a geometry optimization or short unrestrained simulation. The mass-weighted root mean square deviation of the solute with the reference structure is indicated by ρ_m. To speed up convergence, the overall translation and rotation of the system are removed during confinement by performing a mass-weighted best-fit alignment of the system onto the reference structure at each simulation step.[17] The contribution of these motions to the free energy (ΔF_{rr}) is calculated from the partition function of the rigid rotor;[15,19] the translational contribution cancels when calculating the difference ΔF. In explicit solvent simulations the restraints are active on the solute only and the solvent moves freely without restraints.[20,21]

The maximum value of the restraining frequency ($\nu_{\max}$) must be chosen high enough that the solute reaches the HO state. This can be readily checked since for a purely harmonic system $\langle U_{\text{conf}} \rangle = \frac{1}{2} N_{\text{DOF}} k_{\text{B}} T$, where N_{DOF} is the number of degrees of freedom, k_{B} the Boltzmann constant, T the temperature, and $\langle . \rangle$ indicates an average.[17] By accumulating U_{conf}, ΔF_1 and ΔF_2 can be readily obtained from thermodynamic integration[13–17] or from MBAR.[20] To minimize the computational cost, the number of restraining frequencies (n_ν) and their magnitude ν_i are chosen based on a distribution overlap criterion that allows for larger spacings at high frequencies.[19] This dictates a spacing such that the free energy difference between neighboring frequencies is typically a few kcal/mol at the low frequency side, and ten or more kcal/mol on the high frequency side.[19] In practice, the proper spacing can be deduced by monitoring the statistical error of the confinement free energy as a function

of frequency, and inserting extra frequencies whenever this error jumps unacceptably (*i.e.* resulting in an overall confinement error larger than $k_{\mathrm{B}}T$). Care must also be taken to adjust the time step as a function of ν_j in order to sufficiently sample each oscillation period.

ΔF_{1d} and ΔF_{2d} are desolvation free energies of the HO states. Since the HO state is so heavily restrained, these correspond to the desolvation free energies of the reference states.[20] For implicitly solvated systems these follow directly from the implicit solvent model, while for explicitly solvated systems these can be assessed in two ways. It can be obtained from free energy perturbation (FEP) simulations,[21] using the Weeks, Chandler and Andersen decomposition of the desolvation free energy into repulsive, dispersive and electrostatic components.[25,26] The electrostatic contributions of the desolvation free energies can also be obtained from Poisson-Boltzmann (PB) calculations, and their non-polar contributions from the solute's solvent accessible surface area (SASA).[20] In either case, calculated desolvation free energies are for highly restrained, artificial states and can therefore not be compared to experimental values.

ΔF_3 is the free energy difference between the HO states; since the partition function of the harmonic oscillator is known analytically, this free energy difference can be readily calculated. To simplify the treatment of CM and focused confinement, we will use the same $\nu_{\max}$ for each leg. ΔF_3 then equals the energy difference between the reference states.

Focused confinement method

The thermodynamic cycle used for focused confinement is shown by black arrows in Fig. 2b. In focused confinement:

$$\Delta F = \Delta F_{1a} - \Delta F_{2a} + \Delta F_{1b} - \Delta F_{2b} + \Delta F_{3a} + \Delta F_{rr}. \tag{2}$$

ΔF_{1a} (ΔF_{2a}) is the free energy of confining the conformationally active region B to x^0 (y^0) while region A freely samples the a conformation without restraints; that is, the free energy

of transforming region B to the HO state while region A can freely move. This free energy is obtained in a series of restrained simulations using the confinement protocol described above. ΔF_{1b} and ΔF_{2b} are the desolvation free energies of the $A_aB_x^{HO}$ and $A_aB_y^{HO}$ states, respectively; the desolvation free energies of the solute in which region A can move freely while region B is confined. These are obtained from averaging since, in contrast to CM, region A can freely move.

In CM, the vertical free energy difference (ΔF_3) can be readily obtained from analytical theory. In contrast, the vertical in focused confinement (ΔF_{3a}) cannot be calculated analytically, since the partition function of a mixed harmonic $-$ anharmonic system cannot be readily evaluated. Formally ΔF_{3a} is given as:

$$\Delta F_{3a} = -k_B T \ln \frac{\int\limits_{r_A \in a,\, r_B \in y^0} e^{-U(r_A,\, r_B)/k_B T} dr_A dr_B}{\int\limits_{r_A \in a,\, r_B \in x^0} e^{-U(r_A,\, r_B)/k_B T} dr_A dr_B}, \tag{3}$$

where r_A and r_B are the coordinates of the A and B region, respectively, and U the potential energy. The integrals are over the configurational space of the $A_aB_x^{HO}$ and $A_aB_y^{HO}$ states. Region A is in the a conformation in both species and free to move; the r_A portion of the integral is therefore over the $r_A \in a$ region. Region B is purely harmonic. It samples around the reference structure (which is x^0 for $A_aB_x^{HO}$ and y^0 for $A_aB_y^{HO}$) in a highly restrained manner. In fact, the restraints on B in the $A_aB_x^{HO}$ and $A_aB_y^{HO}$ states are so high, that its configurational space effectively collapsed into a single point, which is the reference structure. This means that the integral in Eq. 3 can be replaced by:

$$\Delta F_{3a} = -k_B T \ln \frac{\int\limits_{r_A \in a} e^{-U(r_A,\, r_B)/k_B T} \delta(r_B - y^0) dr_A dr_B}{\int\limits_{r_A \in a} e^{-U(r_A,\, r_B)/k_B T} \delta(r_B - x^0) dr_A dr_B} = -k_B T \ln \frac{\int\limits_{r_A \in a} e^{-U(r_A,\, y^0)/k_B T} dr_A}{\int\limits_{r_A \in a} e^{-U(r_A,\, x^0)/k_B T} dr_A}, \tag{4}$$

where δ indicates the delta function. The ratio of integrals now corresponds to a ratio of partition functions of A sampling a while B is held fixed at x^0 or y^0. This suggests that

ΔF_{3a} can be readily obtained by performing two vacuum simulations: one in which B is fixed to x^0 while A freely samples a (the X simulation) and one in which B is fixed at y^0 while A freely samples a (the Y simulation). By collecting values of $U(r_A^X, y^0)$, that is, energies of snapshots of the X simulation in which the coordinates of B are replaced by y^0, and $U(r_A^Y, x^0)$, that is, energies of snapshots of the Y simulation in which the coordinates of B are replaced by x^0, the free energy difference ΔF_{3a} can be calculated from either Bennett's overlapping distribution method[27] or MBAR.[12] This is feasible, since A samples the same conformation in both systems. That means that the A configurations found in the X (or Y) simulation will be relevant for both $\mathrm{A}_a\mathrm{B}_x^{\mathrm{HO}}$ and $\mathrm{A}_a\mathrm{B}_y^{\mathrm{HO}}$, ensuring large overlap of the probability density functions. Few, if any, steric clashes will be observed when swapping the B coordinates between the simulations when constructing the "chimeric" trajectories needed for Bennett's overlapping distribution method or MBAR.

Eq. 4 is fully consistent with CM: in the limit that both regions A and B are fully harmonic and restrained to the reference state, both regions collapse onto the reference structures and $\Delta F_{3a} \rightarrow U(a^0y^0) - U(a^0x^0)$, the energy difference between the reference states. This equals the CM value for the vertical free energy when using the same $\nu_{\max}$ for each leg (as done here). The correctness of the focused confinement method can be readily tested by extending the thermodynamic cycle to encompass the end points of CM, the $\mathrm{A}_a^{\mathrm{HO}}\mathrm{B}_x^{\mathrm{HO}}$ and $\mathrm{A}_a^{\mathrm{HO}}\mathrm{B}_y^{\mathrm{HO}}$ states (grey arrows of Fig. 2b). ΔF_{1c} (ΔF_{2c}) is the free energy of transforming A to the HO state while B is already fully harmonic; this free energy is calculated by focused confinement. Since the configurational space of B has virtually collapsed to the reference state, these focused confinement simulations are performed by treating B as fixed. This allows for a larger time step in the low frequency regime. A comparison of Fig. 2a with 2b shows that:

$$\Delta F_1 - \Delta F_{1a} - \Delta F_{1c} = 0, \tag{5}$$

$$\Delta F_2 - \Delta F_{2a} - \Delta F_{2c} = 0, \tag{6}$$

and

$$\Delta F_3 - \Delta F_{2d} - \Delta F_{2c} + \Delta F_{2b} - \Delta F_{3a} - \Delta F_{1b} + \Delta F_{1c} + \Delta F_{1d} = 0. \quad (7)$$

These equalities, together with the calculated values of ΔF, serve as stringent tests of the method.

The computational advantage of focused confinement over traditional CM is apparent from Fig. 2 and Eq. 5-6. Clearly, $\Delta F_{1a} < \Delta F_1$ and $\Delta F_{2a} < \Delta F_2$. Since the free energy differences that need to be bridged are smaller in focused confinement, fewer restraint frequencies need to be used, leading to significant savings in computer time.

3 Methods

The free energy differences between two conformers of sucrose and the capped Val-Ala-Pro-Ala peptide were calculated by traditional CM and focused confinement. The reference states were obtained from a clustering analysis of unbiased MD trajectories, followed by restrained energy minimizations in which the conformationally inactive part of the molecule was successively made identical across conformers by increasing the force constant of the rmsd restraint. For sucrose the conformationally active region of the molecule corresponded to the glucose unit (Fig. 3a). In the reference states, this glucose unit was either in the chair or boat conformation. The conformationally inactive region corresponded to the fructose unit, which had the same conformation in both reference states. For the VAPA peptide, the conformationally inactive region corresponded to the first residue and the N, H, and C^{α} atoms of the second residue, while the conformationally active region corresponded to the remainder of the molecule (Fig. 3b). The prolyl peptide bond of the conformationally active region was either in the *cis* or *trans* conformation in the reference states, while the conformationally inactive region was identical.

The free energy differences between these conformations were calculated in vacuum, in implicit solvent, and in explicit solvent. Confinement simulations for the consistency checks

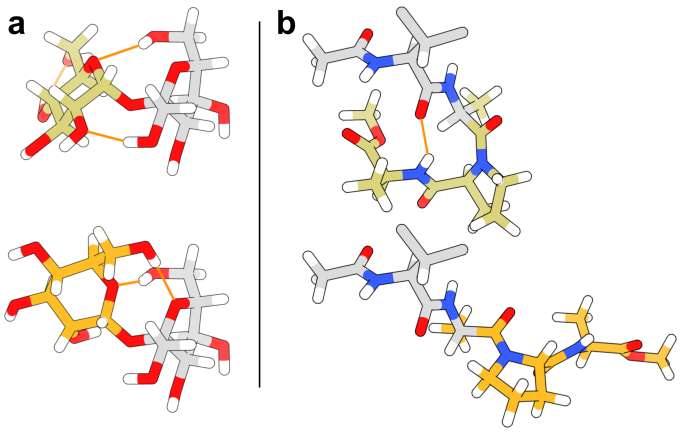

Figure 3: Reference states of sucrose (a) and the VAPA peptide (b). For sucrose the conformationally active region corresponds to the glucose unit, with carbons shown in tan for the chair (top) and orange for the boat conformation (bottom). Carbon atoms of the conformationally active region of the VAPA peptide are shown in tan for the *cis* (top) and orange for the *trans* conformation (bottom). Carbon atoms of the conformationally inactive region are shown in grey. Intramolecular hydrogen bonds are shown as orange lines.

of Fig. 2 and Eq. 5-7 were performed as well. The vacuum and explicit solvent simulations used the CHARMM 36 force field.[28–30] The explicit solvent simulations were performed in the NVT ensemble, after heating and equilibration in the NPT ensemble at 1 bar. Cubic boxes with a water layer of at least 10 Å around the solute were used, resulting in 737 TIP3P[31] water molecules for the sucrose chair simulations, 754 for the sucrose boat, 865 for the VAPA *cis*, and 1948 water molecules for the VAPA *trans* simulations. The particle-mesh Ewald method was used for long-range electrostatic interactions[32] and periodic boundary conditions were in effect.

The desolvation free energies of the explicitly solvated systems were calculated in two ways. The FEP method[21] followed the protocol of Ref.[26] as implemented in the PERT module of the CHARMM program, with λ values of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0 for the electrostatic contribution to the desolvation free energy, ξ values of 0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9 and 1.0 for the dispersive contribution, and s values of 0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, and 1.0 for the repulsive contribution. In the FEP calculations the entire solute was kept frozen for CM; for focused confinement the conformationally active region was

kept frozen while the conformationally inactive region was unrestrained and free to move. All windows were equilibrated for 200 ps and simulated for 100 ps in the NPT ensemble with periodic boundary conditions and the particle-mesh Ewald method in effect. All free energy components were calculated with MBAR. For the PB+SASA method,[20] the polar contribution to the desolvation free energies was calculated with the Poisson-Boltzmann solver of the PBEQ module of the CHARMM program,[33] using a grid spacing of 0.1 Å, a dielectric constant of 80 for water, and an internal dielectric constant of 1. The nonpolar contribution was calculated with GBMV.[34] Bootstrapping was used to calculate average PB+SASA solvation free energies and their standard deviations.

Since there is no desolvation for the vacuum systems, their desolvation free energies are zero. The implicit solvent simulations used the CHARMM 22 force field[35,36] and GBSW.[37] The Born radii for sucrose were set to 1.02 times the van der Waals radii,[38] while optimized radii and adjusted backbone torsional energies were used for the peptide.[39]

All MD simulations were performed at 300 K using Langevin dynamics. To break correlation times,[19] simulations for the lowest three frequencies were performed using temperature replica exchange,[40] at temperatures of 300, 310, 320 and 330 K for the vacuum and implicit solvent simulations, and 300, 305, 310, and 315 K for the explicit solvent simulations. Only the lowest temperature replicas were used for calculating the free energies. The restraining frequencies were between 0.001 and 30 AKMA (0.02 and 613.5 ps^{-1}). Time steps were chosen such that there were at least 30 time steps per harmonic oscillator period. Systems were equilibrated for 500,000 steps at each frequency before production. The same trajectories could be used to calculate ΔF_{3a} for the vacuum and explicit solvent systems, since this free energy is calculated in vacuum and the force fields for the vacuum and explicit solvent treatment were the same. Separate vacuum simulations were performed to calculate ΔF_{3a} for the implicit solvent treatment. All simulations were performed with the CHARMM program;[33] MBAR and decorrelation analyses were performed with pymbar.[12] Free energies were calculated with MBAR using 500 uncorrelated frames per frequency for the vacuum

and explicitly solvated systems and 250 uncorrelated frames per frequency for the implicitly solvated systems. Conformational free energy differences for sucrose and the VAPA peptide were also calculated by umbrella sampling.[41] This involved restrained sampling of the θ puckering angle[42] and glucose-fructose O_6-O_6 and O_5-O_3 distances of sucrose, and restrained sampling of the prolyl peptide bond of the VAPA peptide. Settings and conditions were as in the confinement simulations.

The free energy difference between the closed and open state of hydrophobin I was calculated by focused confinement in explicit water. The initial protein coordinates were taken from protein data bank entry 2FZ6,[24] chains B (open) and C (closed). Missing residues 5 and 6 of chain B were built using the coordinates of chain C. Protonation states were calculated with the H++ server[43] at pH 6.5, and reference states were obtained from restrained energy minimizations. The setup and analysis was the same as for the sucrose and VAPA peptide explicit solvent simulations described above, except that $n_\nu = 71$ and temperature replica exchange at 300, 304, 308, and 312 K was used for the lowest 14 frequencies. Desolvation free energies were calculated with PB+SASA and FEP. The same protocols were followed as in the calculations of the sugar and peptide, except that two additional λ values of 0.033 and 0.067 were used for the electrostatic component, and 3 additional s values of 0.167, 0.183, and 0.225 for the repulsive contribution; moreover, all windows were equilibrated for 400 ps. The closed and open state systems contained 3981 and 4925 water molecules, respectively. Free energies were calculated using 250 uncorrelated frames per frequency.

4 Results

The free energy differences between the boat and chair conformations of sucrose and the *trans* and *cis* conformations of the VAPA peptide (Fig. 3) were calculated by CM and focused confinement in vacuum, implicit and explicit solvent. As illustrated for the explicit solvent VAPA simulations in Fig. 4, the HO states were reached at $\nu_{\max}$. The free energy differences

(ΔF), as well as their free energy components are listed in Table 1. ΔF was calculated with Eq. 1 for CM and Eq. 2 for focused confinement; $\Delta\Delta F$ indicates the difference between the CM and focused confinement values. In all cases, focused confinement free energies were identical within error to the CM results, with $|\Delta\Delta F|$ ranging between 0.1 and 0.5 kcal/mol.

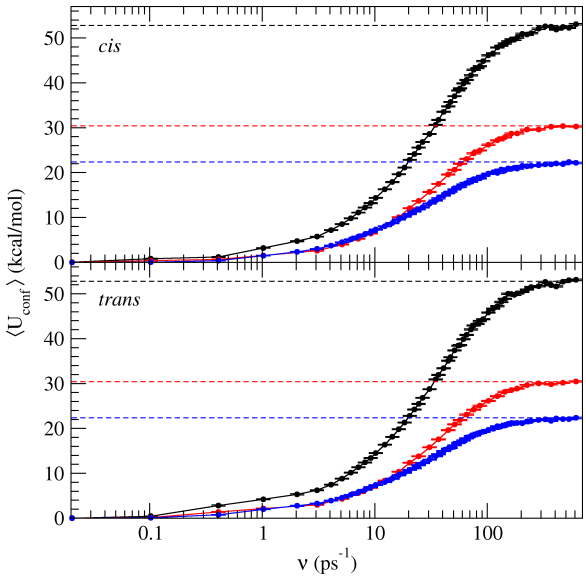

Figure 4: HO state convergence for the VAPA peptide in explicit water. Shown is the value of $\langle U_{\rm conf}\rangle$ as a function of the restraint frequency. CM in black, focused confinement in red, consistency check in blue; these correspond to the horizontal legs 1 and 2 of the thermodynamic cycle of Fig. 2a ($n_\nu = 56$), legs 1a and 2a of Fig. 2b ($n_\nu = 30$), and legs 1c and 1d of Fig. 2b ($n_\nu = 56$), respectively. The HO value of $\frac{1}{2}N_{\rm DOF}k_{\rm B}T$ is indicated by the dashed lines; $N_{\rm DOF}$ is the number of degrees of freedom of the region that is confined, which differs between CM, focused confinement and the consistency checks. Statistical errors are less than 0.2 kcal/mol and shown by vertical bars.

Table 1: Confinement (CM) and focused confinement (FCM) results for the chair to boat conversion of sucrose and the *cis* to *trans* conversion of the VAPA peptide. The conformational free energy difference ΔF was calculated using Eq. 1 for CM and with Eq. 2 for FCM; $\Delta\Delta F$ denotes the difference between the CM and FCM conformational free energy differences. PB+SASA and FEP indicate the different ways in which the desolvation free energies were calculated for the explicitly solvated systems; for reference the values of $\Delta\Delta F_{\text{desolv}} = \Delta F_{2d} - \Delta F_{1d}$ (for CM) and $\Delta\Delta F_{\text{desolv}} = \Delta F_{2b} - \Delta F_{1b}$ (for FCM) are shown as well. ΔF_{umb} indicates conformational free energy differences calculated by umbrella sampling. All energies in kcal/mol.

		Sucrose				VAPA			
				Explicit				Explicit	
		Vacuum	Implicit	PB+SASA	FEP	Vacuum	Implicit	PB+SASA	FEP
	ΔF_{rr}	0.0	0.0	0.0	0.0	-0.3	-0.1	-0.1	-0.1
CM	ΔF_1	255.5 ± 0.2	262.9 ± 0.4	266.0 ± 0.3	266.0 ± 0.3	357.4 ± 0.3	362.8 ± 0.4	360.7 ± 0.3	360.7 ± 0.3
	ΔF_2	262.5 ± 0.3	264.5 ± 0.4	267.6 ± 0.3	267.6 ± 0.3	363.8 ± 0.3	366.9 ± 0.5	367.5 ± 0.3	367.5 ± 0.3
	ΔF_{1d}	0.0 ± 0.0	11.8 ± 0.0	28.1 ± 0.0	17.1 ± 0.0	0.0 ± 0.0	5.9 ± 0.0	18.3 ± 0.0	17.7 ± 0.0
	ΔF_{2d}	0.0 ± 0.0	20.5 ± 0.0	38.5 ± 0.0	25.3 ± 0.0	0.0 ± 0.0	8.4 ± 0.0	21.4 ± 0.0	21.4 ± 0.0
	ΔF_3	15.5 ± 0.0	15.5 ± 0.0	15.5 ± 0.0	15.5 ± 0.0	8.5 ± 0.0	6.1 ± 0.0	8.5 ± 0.0	8.5 ± 0.0
	ΔF	8.5 ± 0.4	5.2 ± 0.3	3.6 ± 0.4	5.8 ± 0.4	1.9 ± 0.4	-0.6 ± 0.6	-1.5 ± 0.5	-2.1 ± 0.5
	n_ν	48	48	48	48	56	56	56	56
	$\Delta\Delta F_{\text{desolv}}$	0.0 ± 0.0	8.8 ± 0.0	10.4 ± 0.0	8.1 ± 0.0	0.0 ± 0.0	2.5 ± 0.0	3.1 ± 0.0	3.7 ± 0.0
FCM	ΔF_{1a}	118.6 ± 0.2	122.2 ± 0.4	125.3 ± 0.3	125.3 ± 0.3	197.5 ± 0.3	203.0 ± 0.4	201.0 ± 0.3	201.0 ± 0.3
	ΔF_{2a}	120.8 ± 0.2	123.6 ± 0.4	126.7 ± 0.4	126.7 ± 0.4	202.2 ± 0.3	205.7 ± 0.4	204.5 ± 0.4	204.5 ± 0.4
	ΔF_{1b}	0.0 ± 0.0	18.9 ± 0.1	36.3 ± 0.1	19.5 ± 0.0	0.0 ± 0.0	6.7 ± 0.0	19.6 ± 0.0	17.2 ± 0.0
	ΔF_{2b}	0.0 ± 0.0	22.0 ± 0.1	41.6 ± 0.1	22.9 ± 0.0	0.0 ± 0.0	9.2 ± 0.0	24.1 ± 0.0	21.9 ± 0.0
	ΔF_{3a}	10.3 ± 0.0	10.3 ± 0.0	10.3 ± 0.0	10.3 ± 0.0	6.5 ± 0.0	4.4 ± 0.0	6.5 ± 0.0	6.5 ± 0.0
	ΔF	8.2 ± 0.3	5.8 ± 0.5	3.7 ± 0.5	5.6 ± 0.5	1.5 ± 0.3	-0.8 ± 0.6	-1.6 ± 0.5	-1.8 ± 0.5
	n_ν	25	25	25	25	30	30	30	30
	$\Delta\Delta F_{\text{desolv}}$	0.0 ± 0.0	3.1 ± 0.1	5.3 ± 0.1	3.4 ± 0.0	0.0 ± 0.0	2.4 ± 0.0	4.5 ± 0.0	4.7 ± 0.0
Check	ΔF_{1c}	136.7 ± 0.2	140.7 ± 0.3	140.9 ± 0.2	140.9 ± 0.2	160.2 ± 0.2	160.0 ± 0.3	159.3 ± 0.2	159.3 ± 0.2
	ΔF_{2c}	141.6 ± 0.2	140.3 ± 0.3	141.3 ± 0.2	141.3 ± 0.2	161.9 ± 0.2	161.1 ± 0.3	162.8 ± 0.2	162.8 ± 0.2
	Eq. 5	0.2 ± 0.3	0.0 ± 0.6	-0.2 ± 0.5	-0.2 ± 0.5	-0.2 ± 0.5	-0.1 ± 0.6	0.4 ± 0.5	0.4 ± 0.5
	Eq. 6	0.0 ± 0.4	0.5 ± 0.6	-0.4 ± 0.5	-0.4 ± 0.5	-0.2 ± 0.5	0.1 ± 0.7	0.2 ± 0.5	0.2 ± 0.5
	Eq. 7	0.3 ± 0.3	0.0 ± 0.4	-0.2 ± 0.3	0.0 ± 0.3	0.3 ± 0.3	0.4 ± 0.4	-0.1 ± 0.3	-0.5 ± 0.3
	$\Delta\Delta F$	0.3 ± 0.5	-0.5 ± 0.6	-0.1 ± 0.6	0.2 ± 0.6	0.4 ± 0.5	0.2 ± 0.9	0.1 ± 0.7	-0.3 ± 0.7
ΔF_{umb}		8.6 ± 0.3	5.7 ± 0.4	5.6 ± 0.4	5.6 ± 0.4	1.4 ± 0.3	-0.7 ± 0.4	-1.9 ± 0.5	-1.9 ± 0.5

Comparison to umbrella sampling results (Table 1) showed close agreement within a fraction of a kcal/mol for all cases, except for sucrose in explicit solvent when the desolvation free energies were calculated by PB+SASA. A comparison of $\Delta\Delta F_{\text{desolv}}$, where $\Delta\Delta F_{\text{desolv}} = \Delta F_{2d} - \Delta F_{1d}$ for CM and $\Delta\Delta F_{\text{desolv}} = \Delta F_{2b} - \Delta F_{1b}$ for focused confinement, showed that this disagreement was due to the desolvation free energy. $\Delta\Delta F_{\text{desolv}}$ is listed in Table 1 with its constituent components in Table 2. For sucrose PB+SASA overestimates $\Delta\Delta F_{\text{desolv}}$ by about 2 kcal/mol in both CM and focused confinement, which equals the discrepancies of ΔF with umbrella sampling. In contrast, for the peptide PB+SASA and FEP closely agreed, and the umbrella sampling, CM and focused confinement values for ΔF closely agreed. While focused confinement and CM results were identical within error for all cases, demonstrating correctness of focused confinement, the results indicated that PB+SASA should not be used for the sugar.

Table 2: Desolvation energy components for explicitly solvated systems. $\Delta F_{\text{desolv}} = \Delta F_{\text{polar}} + \Delta F_{\text{nonpolar}}$ for PB+SASA, and $\Delta F_{\text{desolv}} = \Delta F_{\text{elec}} + \Delta F_{\text{disp}} + \Delta F_{\text{rep}}$ for FEP. All energies in kcal/mol.

			PB+SASA		FEP		
System	State	Method	ΔF_{polar}	$\Delta F_{\text{nonpolar}}$	ΔF_{elec}	ΔF_{disp}	ΔF_{rep}
Sucrose	Chair	CM	30.6 ± 0.0	-2.5 ± 0.0	18.0 ± 0.0	27.3 ± 0.0	-28.1 ± 0.0
Sucrose	Boat	CM	41.1 ± 0.0	-2.6 ± 0.0	27.2 ± 0.0	26.8 ± 0.0	-28.7 ± 0.0
Sucrose	Chair	FCM	38.9 ± 0.1	-2.6 ± 0.0	22.4 ± 0.0	27.1 ± 0.0	-29.9 ± 0.0
Sucrose	Boat	FCM	44.2 ± 0.1	-2.6 ± 0.0	25.7 ± 0.0	26.8 ± 0.0	-29.6 ± 0.0
VAPA	*cis*	CM	21.8 ± 0.0	-3.5 ± 0.0	21.0 ± 0.0	34.6 ± 0.0	-37.9 ± 0.0
VAPA	*trans*	CM	25.3 ± 0.0	-3.9 ± 0.0	24.6 ± 0.0	39.4 ± 0.0	-42.6 ± 0.0
VAPA	*cis*	FCM	23.0 ± 0.0	-3.4 ± 0.0	20.6 ± 0.0	34.5 ± 0.0	-37.9 ± 0.0
VAPA	*trans*	FCM	28.1 ± 0.0	-3.9 ± 0.0	24.9 ± 0.0	39.5 ± 0.0	-42.5 ± 0.0
Hydrophobin I	Open	FCM	785.5 ± 0.4	-22.1 ± 0.0	502.8 ± 0.1	219.3 ± 0.0	-146.0 ± 0.0
Hydrophobin I	Closed	FCM	760.3 ± 0.4	-20.9 ± 0.0	472.1 ± 0.1	207.0 ± 0.0	-127.2 ± 0.0

Critical to CM and focused confinement is the ability to calculate the "vertical" free energy ΔF_3 or ΔF_{3a} (Fig. 2). ΔF_3 corresponds to the free energy difference of two purely harmonic states, which is known analytically, while ΔF_{3a} corresponds to the free energy difference of two mixed harmonic-anharmonic states and is not known analytically. By using the same $\nu_{\max}$ for each leg (as was done here), ΔF_3 equals the energy difference of the reference states and therefore has no statistical error. In contrast, ΔF_{3a} is obtained from sampling (Eq. 4) and will therefore have a statistical error. However, Table 1 shows that in all cases, this error was much less than 0.1 kcal/mol. The reason why ΔF_{3a} can be readily calculated at high accuracy is illustrated in Fig. 5. As explained in the theory section, calculation of ΔF_{3a} requires the construction of "chimeric" trajectories. These are constructed from vacuum trajectories in which the conformationally active region is frozen while the conformationally inactive region is free to move. In the chimeras, the coordinates of the active region are swapped. Since by definition, the conformationally inactive regions sample the same conformational space in each of the original trajectories, large distribution overlaps and no steric clashes are to be expected for the chimeras. Fig. 5 shows that these expectations were indeed observed in the simulations. Fig. 5a and b show the sucrose chimeric trajectories in which no steric clashes and sampling of the same space was observed. Fig. 5c shows the application of Bennett's overlapping distribution method[27] to the VAPA peptide, which demonstrates large overlap of the energy distributions (black and red lines). Consequently the estimated free energy difference (blue line) is constant over a large range of energy values, and agrees with the MBAR calculated value for ΔF_{3a}.

Table 1 shows that all consistency checks (Eq. 5-7) were passed well within $k_{\mathrm{B}}T$. Together with the $\Delta\Delta F$ values these checks clearly demonstrate the correctness and accuracy of the focused confinement method. Table 1 also demonstrates the computational advantage of focused confinement. The conformationally active region of sucrose consisted of 23 out of 45 atoms, or 51% of the total. The focused confinement free energies closely matched this number, with values between 46 and 47% of their CM counterparts, while using 48% fewer

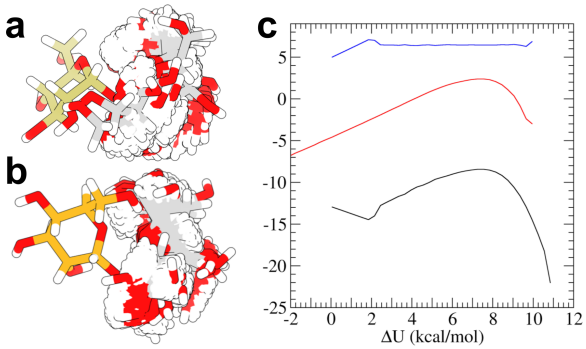

Figure 5: Overlap of mixed harmonic-anharmonic states. a) Chimeric trajectory obtained by swapping glucose to the chair conformation in the fixed boat - free to move fructose trajectory. b) Chimeric trajectory obtained by swapping glucose to the boat conformation in the fixed chair - free to move fructose trajectory. c) Bennett's overlapping distribution method[27] for the VAPA peptide. In red the value of $\ln p_{trans}(\Delta U) + (\Delta U/2k_{\mathrm{B}}T)$ (unitless), in black $\ln p_{cis}(\Delta U) + (\Delta U/2k_{\mathrm{B}}T)$ (unitless), and in blue ΔF_{3a} (in kcal/mol) as calculated from these two curves. Here $\Delta U = U(r_A, y^0) - U(r_A, x^0)$, and p indicates the normalized probability. Data is shown for CHARMM 36, used for both the vacuum and explicit solvent treatments.

frequencies. The conformationally active region of the VAPA peptide consisted of 36 out of 61 atoms, or 59% of the total. ΔF_{1a} and ΔF_{2a} were 56% of ΔF_1 and ΔF_2, while using 46% fewer frequencies. Despite using significantly fewer frequencies, the statistical errors in the focused confinement free energies were the same as their CM counterparts. Reduction of the number of frequencies in CM on the other hand significantly increased the statistical error: when using the same frequencies as in focused confinement the CM confinement errors increased between two and fourfold with values well beyond $k_{\mathrm{B}}T$. Overall, these observations suggest that focused confinement saves computer time in roughly 1 to 1 proportion of the fraction of conformationally inactive atoms while maintaining the same accuracy.

While sucrose and the VAPA peptide were merely used as test systems, their calculated free energies show some interesting trends. Vacuum favors the *cis* conformer of VAPA, while in solution the *trans* conformer is favored. This is likely due to two effects. The *cis* conformer has an intramolecular hydrogen bond which is absent in the *trans* conformer (Fig. 3). In vacuum this extra interaction favors the *cis* conformer, but less so in solution since the *trans* conformer can form hydrogen bonds with the solvent. In addition, the dipole moment across

the prolyl peptide bond is large in the *trans* and nearly absent in the *cis* conformer, hence water favors the *trans* and vacuum the *cis* conformer. While the chair conformer is clearly favored for sucrose, the free energy difference with the boat conformer is larger in vacuum than in solution. This is likely due to screening of the solvent.

The high efficiency of focused confinement enables treatment of large systems. This is illustrated by its application to the loop closing of hydrophobin I (Fig. 1) in explicit water. When limiting the conformationally active region to the loop (residues 59 to 68), large shifts in the position of the rest of the protein were observed in the vacuum simulation of the open state needed for ΔF_{3a} (Fig. 6). This was particularly problematic for regions near the loop. These shifts were not observed in the vacuum closed state simulation (nor in the explicit water simulations). Since the non-loop region sampled different positions in the vacuum closed and open state simulations for this selection of the conformationally active and inactive regions, the assumption that the non-loop region is conformationally inactive was invalidated. Consequently, poor distribution overlaps and large statistical errors in ΔF_{3a} were observed for this setup.

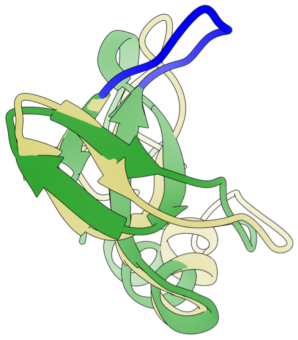

Figure 6: Positional shifts in the vacuum simulation of the open state of hydrophobin I when limiting the conformationally active region to the loop. In green the open state reference structure with the loop in blue; in khaki a representative snapshot of the vacuum simulation.

In order to prevent these shifts, the conformationally active region was redefined as the loop and all C^{α} atoms of the protein. With this definition the conformationally inactive region indeed sampled the same space, and small statistical errors for ΔF_{3a} were obtained. Table 3 summarizes the focused confinement results while Fig. 7a shows that the HO states

were reached at $\nu_{\max}$. Fig. 7b shows the chimeric trajectories: no clashes between the conformationally active and inactive regions were observed. The calculations show that the open state is marginally more stable by about 1 kcal/mol. This free energy difference was obtained with 258.2 million MD equilibration and production steps for the open and 302.8 million for the closed state; the FEP simulations added a total of 72 million production steps.

The hydrophobin I simulations suggest that the workflow should start with the vacuum simulations for ΔF_{3a} in order to assure that the conformationally active and inactive regions are properly defined. A proper selection can be readily assessed based on the statistical errors of ΔF_{3a}. Given the relatively small expense of the vacuum simulations, multiple selections could be tried before starting the confinement simulations. In order to minimize the expense of the confinement simulations, one would then pick the selection with low statistical error that has the least number of conformationally active atoms. While this approach was not pursued here, it is likely that good overlap for hydrophobin I could have been reached with fewer C^{α} atoms in the conformationally active region, which would have reduced the overall calculation cost.

Table 3: Focused confinement results for the open to closed transition of hydrophobin I in explicit water. The conformationally active region was chosen as the loop and all C^{α} atoms. PB+SASA and FEP indicate the different ways in which the desolvation free energies were calculated; for reference the values of $\Delta\Delta F_{\text{desolv}} = \Delta F_{2b} - \Delta F_{1b}$ are shown as well. All energies in kcal/mol.

	PB+SASA	FEP
ΔF_{rr}	0.1	0.1
ΔF_{1a}	1367.3 ± 0.6	1367.3 ± 0.6
ΔF_{2a}	1358.4 ± 0.6	1358.4 ± 0.6
ΔF_{1b}	763.4 ± 0.4	576.1 ± 0.1
ΔF_{2b}	739.5 ± 0.4	551.8 ± 0.1
$\Delta\Delta F_{\text{desolv}}$	-24.0 ± 0.6	-24.3 ± 0.1
ΔF_{3a}	-32.2 ± 0.0	-32.2 ± 0.0
ΔF	0.8 ± 1.0	1.1 ± 0.9

While hydrophobin I is a large system, its desolvation free energy could be readily calculated by FEP. Fast convergence was aided by the fact that a portion of the molecule (the

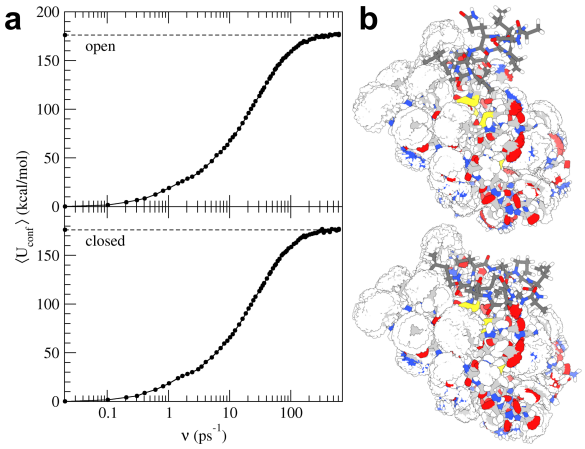

Figure 7: Focused confinement of hydrophobin I in explicit water. The conformationally active region was chosen as the loop and all C^{α} atoms. a) HO state convergence. The HO value of $\frac{1}{2}N_{\mathrm{DOF}}k_{\mathrm{B}}T$ is indicated by the dashed lines. Statistical errors are less than 0.2 kcal/mol and shown by vertical bars. b) Chimeric trajectories. Loop shown in dark grey.

conformationally active region) was held fixed in the FEP simulations. Consistent with the peptide calculations, $\Delta\Delta F_{\mathrm{desolv}}$ of the PB+SASA and FEP methods closely agreed (Table 3); therefore the conformational free energy ΔF also agreed between the two desolvation methods. Given that PB+SASA calculations are much cheaper than FEP simulations, these observations suggest that PB+SASA with a small grid spacing might be an attractive alternative to FEP for the calculation of peptide and protein desolvation free energies in CM and focused confinement.

5 Conclusion

The focused confinement method is shown to be an accurate and efficient method for the calculation of conformational free energy differences in implicit and explicit solvent. This newly introduced reaction coordinate-free technique uses restrained simulations to transform the conformationally active part of the molecule to independent harmonic oscillators, while the rest of the system is left unrestrained and free to move. While this restraining leads to end points that are mixed harmonic-anharmonic states for which the partition functions cannot be easily evaluated, the free energy difference between these end points can be readily

calculated by constructing chimeric trajectories from vacuum simulations. This is due to the fact that the conformational space of the conformationally active region effectively collapsed into a single point, while the conformationally inactive region samples the same space in both end points.

Whether the conformationally active and inactive regions are properly selected can be readily established from structural overlays and energy distributions of the vacuum end point simulations; since these calculations are performed in vacuum, they are cheap in terms of computer time (much cheaper than each of the restraining simulations), and, if needed, several selections can be tried at low cost.

The free energy cost of restraining part of the molecule is necessarily less than restraining the entire molecule, and consequently fewer frequencies are needed to accurately bridge the free energy gap in focused confinement. Therefore, significant amounts of computer time are saved compared to the traditional confinement method. These savings stem from the reduction in the number of restraint frequencies and are in proportion to the fraction of conformationally inactive atoms. The efficiency of the focused confinement method enables its application to large systems.

Acknowledgement

This work was supported by NSF MCB-1919096. Computer time was provided by USF Research Computing, sponsored in part by NSF MRI CHE-1531590.

References

(1) Spiriti, J.; Kamberaj, H.; van der Vaart, A. Development and application of enhanced sampling techniques to simulate the long-time scale dynamics of biomolecular systems. *Int. J. Quantum Chem.* **2012**, *112*, 33–43.

(2) Miao, Y.; McCammon, J. A. Unconstrained enhanced sampling for free energy calculations of biomolecules: a review. *Mol. Sim.* **2016**, *42*, 1046–1055.

(3) Valsson, O.; Tiwary, P.; Parrinello, M. In *Annual Review of Physical Chemistry, Vol 67*; Johnson, MA and Martinez, TJ„ Ed.; Annual Review of Physical Chemistry; 2016; Vol. 67; pp 159–184.

(4) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta Gen. Subj.* **2015**, *1850*, 872–877.

(5) Doshi, U.; Hamelberg, D. Towards fast, rigorous and efficient conformational sampling of biomolecules: Advances in accelerated molecular dynamics. *Biochim. Biophys. Acta Gen. Subj.* **2015**, *1850*, 878–888.

(6) Kaestner, J. Umbrella sampling. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2011**, *1*, 932–942.

(7) Zuckerman, D. M. In *Annual Review of Biophysics, Vol 40*; Rees, DC and Dill, KA and Williamson, JR„ Ed.; Annual Review of Biophysics; 2011; Vol. 40; pp 41–62.

(8) van der Vaart, A. Simulation of conformational transitions. *Theor. Chem. Acc.* **2006**, *116*, 183–193.

(9) Earl, D.; Deem, M. Parallel tempering: Theory, applications, and new perspectives. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.

(10) Wu, D.; Fajer, M. I.; Cao, L.; Cheng, X.; Yang, W. In *Computational Approaches for Studying Enzyme Mechanism, Pt A*; Voth, GA„ Ed.; Methods in Enzymology; 2016; Vol. 577; pp 57–74.

(11) Mitsutake, A.; Sugita, Y.; Okamoto, Y. Generalized-ensemble algorithms for molecular simulations of biopolymers. *Biopolymers* **2001**, *60*, 96–123.

(12) Shirts, M. R.; Chodera, J. D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.* **2008**, *129*, 124105.

(13) Stoessel, J. P.; Nowak, P. Absolute free-energies in biomolecular systems. *Macromolecules* **1990**, *23*, 1961–1965.

(14) Tyka, M. D.; Clarke, A. R.; Sessions, R. B. An efficient, path-independent method for free-energy calculations. *J. Phys. Chem. B* **2006**, *110*, 17212–17220.

(15) Tyka, M. D.; Sessions, R. B.; Clarke, A. R. Absolute free-energy calculations of liquids using a harmonic reference state. *J. Phys. Chem. B* **2007**, *111*, 9571–9580.

(16) Cecchini, M.; Krivov, S. V.; Spichty, M.; Karplus, M. Calculation of free-energy differences by confinement simulations. Application to peptide conformers. *J. Phys. Chem. B* **2009**, *113*, 9728–9740.

(17) Ovchinnikov, V.; Cecchini, M.; Karplus, M. A simplified confinement method for calculating absolute free energies and free energy and entropy differences. *J. Phys. Chem. B* **2013**, *117*, 750–762.

(18) Capelli, R.; Villemot, F.; Moroni, E.; Tiana, G.; van der Vaart, A.; Colombo, G. Assessment of mutational effects on peptide stability through confinement simulations. *J. Phys. Chem. Lett.* **2016**, *7*, 126–130.

(19) Villemot, F.; Capelli, R.; Colombo, G.; van der Vaart, A. Balancing accuracy and cost of confinement simulations by interpolation and extrapolation of confinement energies. *J. Chem. Theory Comp.* **2016**, *12*, 2779–2789.

(20) Villemot, F.; Peguero-Tejada, A.; van der Vaart, A. Calculation of conformational free energies by confinement simulations in explicit water with implicit desolvation. *Mol. Sim.* **2018**, *44*, 1082–1089.

(21) Esque, J.; Cecchini, M. Accurate calculation of conformational free energy differences in explicit water: The confinement-solvation free energy approach. *J. Phys. Chem. B* **2015**, *119*, 5194–5207.

(22) Roy, A.; Perez, A.; Dill, K. A.; MacCallum, J. L. Computing the relative stabilities and the per-residue components in protein conformational changes. *Structure* **2014**, *22*, 168–175.

(23) Boonstra, S.; Onck, P. R.; van der Giessen, E. Computation of hemagglutinin free energy difference by the confinement method. *J. Phys. Chem. B* **2017**, *121*, 11292–11303.

(24) Hakanpää, J.; Szilvay, G. R.; Kaljunen, H.; Maksimainen, M.; Linder, M.; Rouvinen, J. Two crystal structures of Trichoderma reesei hydrophobin HFBI - The structure of a protein amphiphile with and without detergent interaction. *Protein Sci.* **2006**, *15*, 2129–2140.

(25) Weeks, J.; Chandler, D.; Andersen, H. Role of Repulsive Forces in Determining Equilibrium Structure of Simple Liquids. *J. Chem. Phys.* **1971**, *54*, 5237+.

(26) Deng, Y.; Roux, B. Hydration of amino acid side chains: Nonpolar and electrostatic contributions calculated from staged molecular dynamics free energy simulations with explicit water molecules. *J. Phys. Chem. B* **2004**, *108*, 16567–16576.

(27) Bennett, C. Efficient estimation of free energy differences from Monte Carlo data. *J. Comp. Phys.* **1976**, *22*, 245–268.

(28) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D., Jr. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone phi, psi and Side-Chain chi(1) and chi(2) Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.

(29) Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell, A. D., Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages between Hexopyranoses. *J. Chem. Theory Comput.* **2009**, *5*, 2353–2370.

(30) Raman, E. P.; Guvench, O.; MacKerell, A. D., Jr. CHARMM Additive All-Atom Force Field for Glycosidic Linkages in Carbohydrates Involving Furanoses. *J. Phys. Chem. B* **2010**, *114*, 12981–12994.

(31) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(32) Essmann, U.; Perera, L.; Berkowitz, M.; Darden, T.; Lee, H.; Pedersen, L. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(33) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D., Jr.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular Simulation Program. *J. Comp. Chem.* **2009**, *30*, 1545–1614.

(34) Lee, M.; Feig, M.; Salsbury, F.; Brooks, C. New analytic approximation to the standard molecular volume definition and its application to generalized born calculations. *J. Comp. Chem.* **2003**, *24*, 1348–1356.

(35) MacKerell, A.; Bashford, D.; Bellott, M.; Dunbrack, R.; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W.;

Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.

(36) Mackerell, A.; Feig, M.; Brooks, C. Extending the treatment of backbone energetics in protein force fields: Limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comp. Chem.* **2004**, *25*, 1400–1415.

(37) Im, W.; Lee, M.; Brooks, C. Generalized born model with a simple smoothing function. *J. Comp. Chem.* **2003**, *24*, 1691–1702.

(38) Green, D. F. Optimized parameters for continuum solvation calculations with carbohydrates. *J. Phys. Chem. B* **2008**, *112*, 5238–5249.

(39) Chen, J.; Im, W.; Brooks, C. Balancing solvation and intramolecular interactions: Toward a consistent generalized born force field. *J. Am. Chem. Soc.* **2006**, *128*, 3728–3736.

(40) Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.

(41) Torrie, G.; Valleau, J. Non-physical sampling distributions in Monte-Carlo free-energy estimation - umbrella sampling. *J. Comp. Phys.* **1977**, *23*, 187–199.

(42) Cremer, D.; Pople, J. General definition of ring puckering coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.

(43) Anandakrishnan, R.; Aguilar, B.; Onufriev, A. V. H++3.0: automating pK prediction and the preparation of biomolecular structures for atomistic molecular modeling and simulations. *Nucleic Acids Res.* **2012**, *40*, W537–W541.

Graphical TOC Entry