JCTC Journal of Chemical Theory and Computation

Article

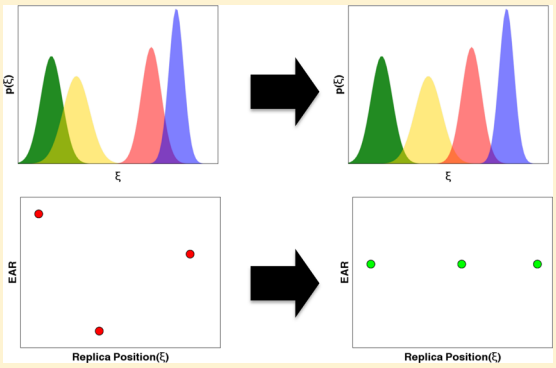
pubs.acs.org/JCTC

Optimization of Umbrella Sampling Replica Exchange Molecular Dynamics by Replica Positioning

Danial Sabri Dashti[†,§] and Adrian E. Roitberg*[,‡,§]

[†]Departments of Physics and [‡]Chemistry and [§]Quantum Theory Project, University of Florida, Gainesville, Florida 32611-8435, United States

ABSTRACT: The positioning of sampling windows in an umbrella sampling simulation has an effect on the rate of convergence and computational efficiency. When such simulation is coupled with a Hamiltonian replica exchange setup, we show that such positioning can be optimized for maximal convergence of the results. We present a method for estimating the exchange acceptance ratio (EAR) between two arbitrary positions on a reaction coordinate in umbrella sampling replica exchange (USRE) molecular dynamics (MD). We designed a scoring function to optimize the position of the set of replicas (windows). By maximizing the scoring function, we make EAR the same for all neighbor replica pairs, increasing the efficiency of the method. We tested our algorithm by sampling a torsion for butane in implicit solvent and by studying a salt bridge in explicit solvent. We found that the optimized set of replicas recovers the correct free energy profile much faster than for equally spaced umbrellas.



INTRODUCTION

In recent decades atomistic molecular dynamics (MD) has improved our understanding about the dynamics, structure, and energetics of molecular systems including large biomolecular systems. However, because of the small size of the time steps and the roughness of potential energy surfaces in large systems, regular MD sampling is mostly limited to small systems. To overcome this limitation, different enhanced sampling methods such as multicanonical, replica exchange MD (REMD), and biased sampling MD algorithms have been developed in recent years.[1−8]

Among those, the umbrella sampling method[9−11] (US) is well-known for improving the sampling of rare events along the reaction coordinate. US attempts to solve the bad sampling problem by applying a biasing potential to the system, to guarantee the effective sampling along the reaction coordinates. This can be achieved either in one or several simulations (windows). While the efficiency of US is clear, it needs careful tuning. For example, the positions of windows and the strength of the bias potentials have to be such that they have enough overlap.

REMD techniques[3,12−15] have become increasingly popular schemes. REMD is not only a promising method for tackling the quasi-ergodicity[16] problem, but it is also able to combine intuitively with other enhanced sampling techniques.[7,12−14,17−19] In REMD methods, *N* noninteracting copies (replicas) of the system with different temperatures (Temperature REMD[3,12]) or different potential energy parameters (Hamiltonian REMD[17−22]) run concurrently and each replica attempts to exchange with its neighbors every few steps. In other words, the system performs a random walk in the replicas' ladder space.

Three types of efforts have been made to optimize REMD with most of those being only applicable to TREMD. First, some techniques try to keep the number of replicas limited. Second, there are methods that try to increase the efficiency of REMD by increasing the probability of accepting an exchange. Third, are methods that increase the efficiency by optimal selection of the positions of replicas on a temperature/Hamiltonian ladder.

In TREMD, the number of needed replicas in a system increases with the number of degrees of freedom.[20,23,24] This scaling confines TREMD to small systems. A clear way to avoid this problem is to decrease the number of degrees of freedom.[25,26] Approaches comprise the use of hybrid explicit/implicit solvent models for the MD and exchange parts respectively,[23] the use of separate baths for solvent and solute,[24] the use of a coarse grained model for a subset of the system[22] etc. Also, some studies have tried to decrease the number of replicas by coupling the system to a presampled reservoir.[27,28]

A few efforts have been made to increase the probability of exchange. Rick[29] used dynamically scaled replicas between conventional replicas at broad temperature ranges. Li et al.[30] devised a new approach called temperature intervals with global energy reassignment (TIGER). In their method, short runs of heating−sampling−quenching sequences are done, and at the end of each sequence the potential energies of all replicas are compared using Metropolis MC sampling and finally reassigned to the temperature levels. Kamberaj et al.[31] added appropriate bias potentials to the system to flatten the potential of mean force and to achieve a random walk along the temperature

Received: May 17, 2013
Published: October 9, 2013

ladder. Also, Ballard et al.[7] used nonequilibrium works simulations to generate the attempted configuration swaps.

For the optimal choice of the ladder of replicas, it has been argued that the highest efficiency is achieved when the EAR between neighbor replicas is about 20%.[32−35] In 2002, Kofke[32,33] showed that when the specific heat of the system is independent of temperature, assigning temperatures in a geometric progression leads to equal EAR for TREMD. But, if the specific heat is a function of temperature, then the choice of temperatures is not system-independent.

Although setting the equi-EAR between the neighbors is the most accepted criterion for optimizing REMD (especially in TREMD) and it has been shown[36−40] that for systems with phase transitions along replica ladders, this is not the optimal arrangement. A few techniques have been proposed for selecting the set of temperatures when the specific heat changes with temperature.[41,42]

The majority of proposed optimization methods for REMD are applicable only to TREMD, with those that are applicable to HREMD methods being computationally demanding. In this paper, we propose a method for optimizing umbrella sampling replica exchange.[17] We use three unique properties of USRE for optimizing the position of replicas.

First, in the case of USRE, there is no phase transition along the reaction coordinate, because the bias potentials restrain the sampled coordinate locally and prevent unsmooth phase changes along the reaction coordinate. The optimal replica set in the absence of phase transitions are such that each replica spends the same amount of time over each position on the reaction coordinate. Similar to TREMD, in the case of USRE this can be realized by a set of positions that yields a uniform exchange acceptance ratio between all neighbor replicas. In such a case EAR has no bottleneck along the replicas' ladder, and this maximizes the roundtrip rate. Second, the exchange probability in USRE is only a function of the difference between ξs at the moment of exchange. This exchange acceptance ratio (EAR) is very easy to compute. Third, in USRE we can safely approximate the reaction coordinate distribution in each window as a Gaussian distribution.[43,44] In this article, we do not discuss or propose any optimal value of EAR between the neighbors, but we provide a way to optimize the efficiency of USRE given either the number of replicas or the desired EAR between all neighbors.

To test our method we applied it to optimize USRE simulations to sampling butane dihedral rotation in implicit solvent and to compute the free energy of formation of a NH_4^+ + OH^- salt bridge in explicit solvent. The proposed positions of replicas (windows) by our method not only resulted in better mixing and convergence with respect to that of equal distance on the reaction coordinate but also the best possible arrangement with given range and number of replicas. We show that any deviation from this arrangement produces a lower rate of mixing.

In the rest of the paper we use replica and windows interchangeably.

THEORY

1. HREMD and USRE. The details of the HREMD and USRE ideas have been explained in detail by others.[3,12,17,45] We briefly review them for completeness.

In HREMD, N noninteracting replicas run simultaneously with N different Hamiltonians but at the same temperature. After a certain number of MD steps, neighboring replicas try to exchange their conformation according to exchange criterion. As in TREMD, the core of the exchange criteria in HREMD is derived by imposing a detailed balanced equation to the generalized ensemble. At the exchange time, the detailed balance equation can be written as

$$p(X)w(X \to X') = p(X')w(X' \to X) \tag{1}$$

here X and X' are respectively the generalized states of the system before and after exchange, $p(X)$ is the probability of the generalized ensemble being at the state of X and $w(X \to X')$ is the probability of the system going from state of X to that of X'.

The attempted exchange would move X to X' as follows:

$$X = \begin{pmatrix} q_1 \dots q_i & q_j \dots q_N \\ H_1 \dots H_i & H_j \dots H_N \end{pmatrix} \to X' = \begin{pmatrix} q_1 \dots q_j & q_i \dots q_N \\ H_1 \dots H_i & H_j \dots H_N \end{pmatrix} \tag{2}$$

where q_k and H_k respectively represent configuration and Hamiltonian of replica k just before the exchange.

Since each replica is independent of the others, the probability of the system being at generalized state of X can be written as $p(X) = \Pi_{i=1}^{N} P_i(q_i, H_i)$. Each $P_i(q_i,H_i)$ is the probability of replica i with conformation q_i and Hamiltonian H_i. By writing the Boltzmann canonical ensemble probability density for P_i values, eq 1 can be rewritten as

$$\frac{w(X \to X')}{w(X' \to X)} = \frac{\exp(-\beta(H_i(q_j) + H_j(q_i)))}{\exp(-\beta(H_i(q_i) + H_j(q_j)))} = \exp(-\beta\Delta) \tag{3}$$

where β is $(1/k_BT)$ and $\Delta = H_i(q_j) + H_j(q_i) - H_i(q_i) - H_j(q_j)$.

Using the Metropolis criterion,[46] the probability of accepting the move $X \to X'$ is computed as follows:

$$p_{acc} = w(X \to X') = \min\{1, \exp(-\beta\Delta)\} \tag{4}$$

USRE[17] is a combination of umbrella sampling and hamiltonian replica exchange MD, i.e., in USRE N noninteracting replicas run simultaneously with N different bias potentials. The bias potentials are harmonic functions with distinct strengths and coordinate references.

The bias of the ith window has a quadratic form of

$$B_i(\xi) = k_i(\xi(q) - \xi_i^0)^2 \tag{5}$$

where k_i, ξ, ξ_i^0, and q are the bias strength, the reaction coordinate, the bias center, and configuration, respectively. We will present results only for the case for which all our replicas have the same force constant but different coordinate centers. The method is easily expandable to dissimilar potentials that also have different bias strengths.

The Hamiltonian for the ith window can be written as

$$H_i = K + U_{UB} + B_i \tag{6}$$

where U_{UB} and K are the unbiased potential energy and the kinetic energy, respectively.

For two exchanging replicas, i and j, kinetic and potential energy terms cancel individually. Finally, with $k_i = k_j = k$, Δ can be simplified to the following:

$$\begin{aligned}\Delta &= [k(\xi_i - \xi_j^0)^2 + k(\xi_j - \xi_i^0)^2] \\ &\quad - [k(\xi_i - \xi_i^0)^2 + k(\xi_j - \xi_j^0)^2] \\ &= 2k(\xi_i^0 - \xi_j^0)(\xi_i - \xi_j) \\ &= 2k\Delta\xi_{ij}^0\Delta\xi_{ij} \end{aligned} \quad (7)$$

Given a set of bias centers for each window, $\Delta\xi_{ij}^0$ is a constant and then Δ becomes a linear function of $\Delta\xi_{ij}$ which is not dependent on the potential energies of exchanging replicas.

2. Calculating the EAR in USRE. By inserting eq 7 into eq 4, we can write the probability of accepting an exchange as

$$p_{\mathrm{ACC}} = \min(1, \exp[-2\beta k(\Delta\xi_{ij}^0)(\Delta\xi_{ij})]) \quad (8)$$

where $\Delta\xi_{ij}$ is $\xi_i - \xi_j$.

Using the probability density function (PDF) of $\Delta\xi_{ij}$ (i.e., $p(\Delta\xi_{ij})$), the EAR_{ij} (or the average probability of acceptance between replicas i and j) can be calculated as:

$$\mathrm{EAR}_{i,j} = \langle p_{\mathrm{ACC}} \rangle_{p(\Delta\xi_{ij})} = \int p_{\mathrm{ACC}}(\Delta\xi_{ij}) p(\Delta\xi_{ij})\, \mathrm{d}\Delta\xi_{ij} \quad (9)$$

We can estimate the $p(\Delta\xi_{ij})$ as a normal distribution (see Appendix A for details)

$$p(\Delta\xi_{ij}) = \frac{1}{\sqrt{2\pi(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}} \exp\left[-\frac{(\Delta\xi_{ij} - (\bar{\xi}_i - \bar{\xi}_j))^2}{2(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}\right] \quad (10)$$

here, $\bar{\xi}_i$ and $\sigma_{\xi_i}^2$ are the mean and the variance of ξ_i distribution. Note that $\bar{\xi}_i \neq \xi_i^0$.

By substituting eqs 8 and 10 in eq 9, the average probability of acceptance becomes

$$\begin{aligned}\mathrm{EAR}_{i,j} = \int_{-\infty}^{\infty} \mathrm{d}\Delta\xi_{ij} \min(1, \exp[-2\beta k(\Delta\xi_{ij}^0)(\Delta\xi_{ij})]) \\ \frac{1}{\sqrt{2\pi(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}} \exp\left[-\frac{(\Delta\xi_{ij} - (\bar{\xi}_i - \bar{\xi}_j))^2}{2(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}\right] \end{aligned} \quad (11)$$

Without loss of generality, we can simplify the min by assuming $\xi_i^0 < \xi_j^0$, and finally

$$\begin{aligned}\mathrm{EAR}_{i,j} &= \frac{1}{2}\mathrm{erfc}\left(-\frac{(\bar{\xi}_i - \bar{\xi}_j)}{\sqrt{2(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}}\right) \\ &\quad + \frac{1}{2}\mathrm{erfc}\left(\frac{((\bar{\xi}_i - \bar{\xi}_j) - 2\beta k(\Delta\xi_{ij}^0)(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2))}{\sqrt{2(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)}}\right) \\ &\quad \exp[-2\beta k(\Delta\xi_{ij}^0)(\bar{\xi}_i - \bar{\xi}_j) + 2\beta^2 k^2(\sigma_{\xi_i}^2 + \sigma_{\xi_j}^2)(\Delta\xi_{ij}^0)^2] \end{aligned} \quad (12)$$

In eq 12, all other parameters except the $\bar{\xi}_{i|j}$ and $\sigma_{\xi_{i|j}}^2$ are known. In Appendix B we present a method for estimating the mean and variance of ξ at any point of the reaction coordinate using very few short presimulated windows.

3. USRE Optimization. As we discussed above, there is no phase transition along the replicas ladder in the case of USRE. If there is no phase transition the best solution for positions of replicas is for the EAR between all neighbor pairs to be equal.

We present here two different approaches for making a set of equi-EARs among neighbor pairs. We discuss fitness functions for each case, that, when maximized, lead to equal EAR among neighbor replicas.

In the first method, the number of replicas and the centers of the first and last replicas are fixed (the range for the sampling coordinate). In order to find the set of equi-EAR replicas, we define a full batch scoring function (FBSF) as follows:

$$\mathrm{FBSF}(\xi_2^0, \ldots, \xi_{N-1}^0) = -\sum_{i=2}^{N} (\mathrm{EAR}_{i-1,i}(\Delta\xi_{i-1i}^0) - \overline{\mathrm{EAR}})^2 \quad (13)$$

and maximize it with respect to the center of biases in all replicas except the last and first ones (i.e., $\xi_2^0, \ldots, \xi_{N-1}^0$). Here N is the number of replicas and $\overline{\mathrm{EAR}} = (\Sigma_{i=2}^{N}\mathrm{EAR}_{i-1,i}(\Delta\xi_{i-1i}^0))/(N-1)$.

In FBSF we set the position of the first and the last replicas. Consequently the total number of degrees of freedom is equal to $N-2$. Since $\mathrm{EAR}_{i-1,i}$ is a strictly decreasing function of $\xi_{i-1}^0 - \xi_i^0$, then it can be concluded that the scoring function (which is an $N-2$-dimensional parabola) has a unique maximum. Maximizing the FBSF leads to equal EAR between adjacent replicas.

In the second method, we choose a desired value for the EAR between replicas, and we fix the center of bias for the first replica (lowest value on reaction coordinate) and the range of the reaction coordinate. In this scheme, starting from the lowest value on the reaction coordinate, we add a new replica at a time and maximize the single batch scoring function (SBSF) with respect to that center of bias in that replica. We continue this scheme until the last replica passes the maximum range of the reaction coordinate. In order to adjust the position ith replica, we try to maximize the following SBSF at every replica accumulation:

$$\mathrm{SBSF}(\xi_i^0) = -(\mathrm{EAR}_{i-1,i}(\Delta\xi_{i-1i}^0) - \mathrm{EAR}_c)^2$$

where EAR_c is the desired value of EAR, which is bounded between 0 and 1. At the end of this process, the EAR between all neighbor pairs equals EAR_c.

We used the Barzilai and Borwein[47,48] optimization method for maximizing both SBSF and FBSF (see Appendix E for more details).

4. USRE Optimization Workflows. On the basis of the choice of optimization, two types of workflows exist. In both methods we presimulate M windows.

In the FBSF method, the number of replicas is set, but the final average EAR is not known a priori (Figure 1).

In the SBSF method the EAR_c is a chosen parameter and the number of replicas in not known in advance (Figure 2).

We have written scripts (for both FBSF and SBSF) that implement the ideas presented here. Both scripts are available at http://www.clas.ufl.edu/users/roitberg/software.html.

5. Simulation Details. In order to validate our method, we performed USRE simulations on two systems. First, we studied the PMF along the butane torsional angle in the range from −65° to 67° using 12 replicas. We performed four sets of simulations for different arrangements of replicas on the reaction coordinate. We applied the OBC Generalized Born[49] implicit solvent model (igb = 5) with a cutoff of 12 Å for nonbonded interactions. A spring constant of k = 30 kcal mol^{-1} radian^{-2} has been applied to all replicas. The length of each simulation was about 50 ns in order to have enough sampling for estimating the average roundtrip time (ART) between the highest and the lowest replicas in each set.

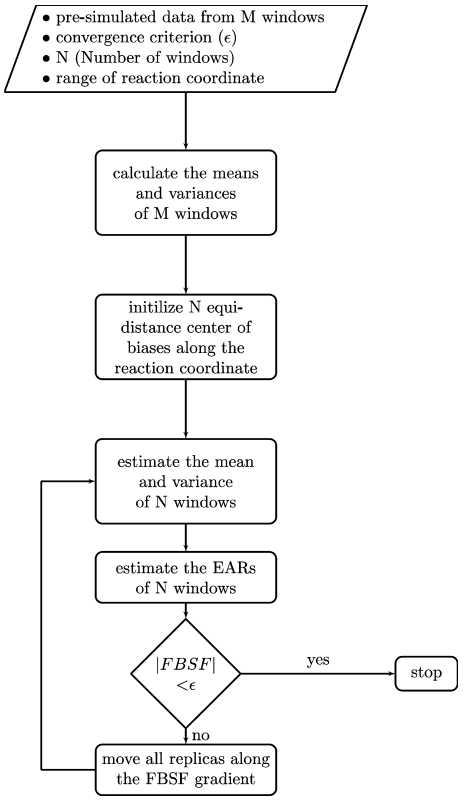


Figure 1. FBSF workflow.

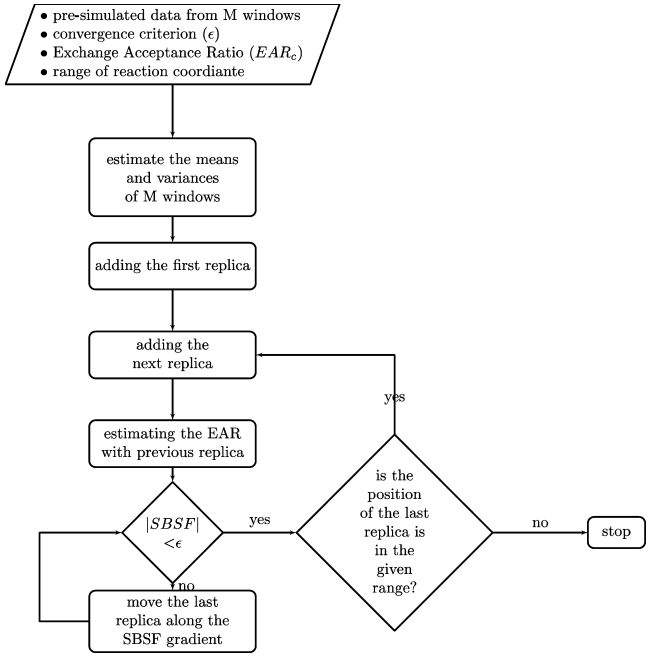


Figure 2. SBSF workflow.

Second, we performed two sets of simulations using 24 replicas on a NH_4^+ + OH^- salt bridge immersed in a box of TIP3P[50] solvent model of 21.4 Å × 21.4 Å × 21.4 Å. We calculated the long-range electrostatic interactions with the particle mesh Ewald method, and used an 8 Å cutoff for the short-range nonbonded interactions. A spring constant of k = 30 kcal mol^{-1} Å^{-2} has been applied to all replicas. The lengths of all simulations were 40 ps; however, we used a 5 ns USRE simulation as a reference in calculation of the Kullback−Leibler divergence.[51]

In all calculations in this paper, we used the AMBER 12 molecular simulation suite[52] and the general AMBER force field (GAFF).[53] The SHAKE[54] algorithm was employed to constrain the distance between hydrogens and heavy atoms, which permitted the use of 2 fs MD steps. Replica exchanges were attempted every 50 MD steps[55] and Langevin dynamics with friction coefficient of 2.0 ps^{-1} was employed to sustain the systems at 300 K. In order to produce independent simulations,[56] we used distinct random seeds (ig = −1 in AMBER) for the Langevin thermostats in all simulations. We saved ξ every 10 MD steps (i.e., 20 fs) and we used Grossfield's[57] implementation of Weighted Histogram Analysis[10] Method (WHAM) for calculating the Potential of Mean Forces versus reaction coordinate in all simulations.

RESULTS AND DISCUSSION

To study the effect of replica optimization in USRE, we computed and compared the ART for different positions of the replicas, including our optimized one. Furthermore, we measured the convergence speed for optimized and non-optimized USRE arrangement of NH_4^+ + OH^- salt bridge in explicit solvent. In both simulations, we chose the total number of replicas to be equal to that of presimulated replicas (i.e., M = N).

1. Potential of Mean Force along the Butane Dihedral Angle. We performed four sets of USRE calculations on butane dihedral in implicit solvent.

In the first one, the centers of biases were equally distributed (every 12°) on the reaction coordinate. In the second set of calculations we moved two neighboring replicas in the middle of the reaction coordinate and we kept all other replicas in place. For the third set, we optimized the replica positions by maximizing the FBSF. For the optimization, we used the first 50 ps of the first set as presimulated data. Finally, in the fourth set of simulations we perturbed the position of one replica in the third (optimized) set. The positions of replicas and corresponding EAR in each set has been shown in Table 1.

For each set, the ART between extremum replicas was computed. To gather sufficient statistics, each simulation was repeated 10 times. Since there were 12 replicas in each simulation, the total number of measured ARTs in each set was 120. We present below the probability density of the ARTs (Figure 3).

Figure 3 can be analyzed based on the central limit theorem (CLT). The CLT describes the characteristics of the population of the means of an infinite number of random samples of size N, drawn from an infinite parent population. According to the CLT, the distribution of the mean converges to a Gaussian distribution, and the mean of the means (i.e., center of the Gaussian) is equal to the mean of the parent distribution.

On the basis of the CLT, the centers of the distributions are equal to the mean of the parent pools. According to Figure 3, the optimized set (set 3) has the lowest ART among all the sets and as the replica positions deviate more from the optimized arrangement (i.e., sets 4, 1, and 2, respectively) the ART increases. This shows the equi-EAR set corresponds to the minimum ART and the mixing is the best in this arrangement. Since the number of exchange attempts is equal in all sets, we expect the sets with higher ARTs to have the smaller sample sizes (i.e., the number of roundtrips in each simulation), which

dx.doi.org/10.1021/ct400366h | J. Chem. Theory Comput. 2013, 9, 4692−4699

Table 1. Position and EAR for Four Different Settings of USRE Calculations on the Butane Dihedral Simulation in Implicit Water[a]

set 1		set 2		set 3		set 4	
position	EAR	position	EAR	position	EAR	position	EAR
−65.00	0.18	−65.00	0.19	−65.00	0.14	−65.00	0.14
−53.00	0.18	−53.00	0.18	−51.72	0.14	−51.72	0.14
−41.00	0.16	−41.00	0.16	−38.58	0.14	−38.58	0.14
−29.00	0.13	−29.00	0.13	−26.01	0.14	−26.01	0.14
−17.00	0.10	−17.00	0.13	−14.44	0.15	−14.44	0.15
−5.00	0.07	−6.00	0.03	−4.16	0.14	−4.16	0.10
7.00	0.10	9.00	0.18	5.73	0.15	7.00	0.21
19.00	0.14	19.00	0.14	16.10	0.14	16.10	0.13
31.00	0.16	31.00	0.17	27.98	0.14	27.98	0.14
43.00	0.18	43.00	0.18	40.60	0.14	40.60	0.14
55.00	0.19	55.00	0.19	53.71	0.14	53.71	0.14
67.00	0.00	67.00	0.00	67.00	0.00	67.00	0.00

[a]Position of each replica corresponds to the center of bias for that replica. Also, the EAR of each position corresponds to the EAR between that replica position and its higher neighbor on the reaction coordinate.

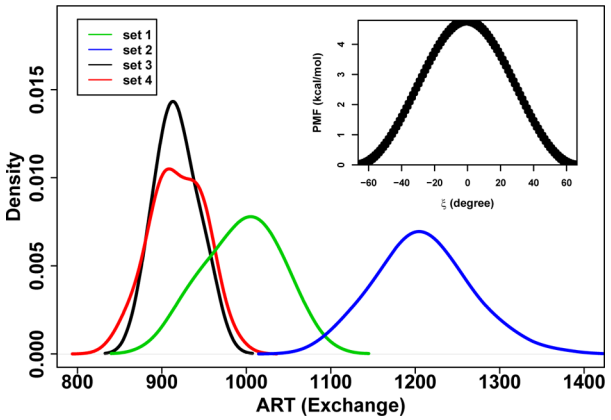


Figure 3. ART probability density. The inset shows the PMF vs reaction coordinate (dihedral angle).

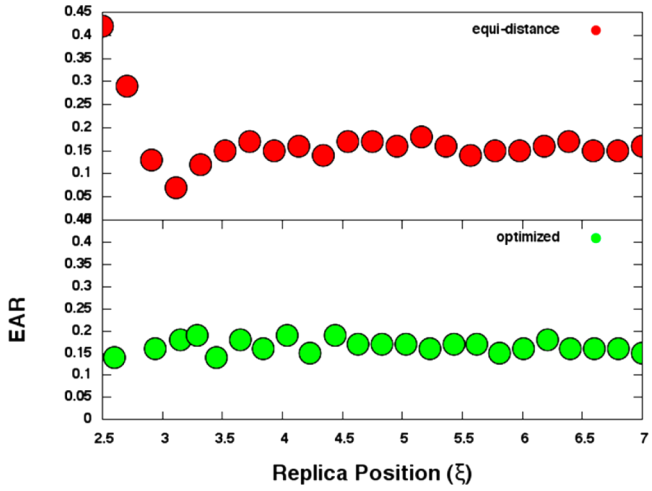


Figure 4. Window positions vs EARs in the equi-distance (non-optimized) (red circles) and optimized (set 2) (green circles) simulations. The EAR of each position corresponds to EAR between that replica position and its higher neighbor on the reaction coordinate.

according to the CLT leads to wider ART distributions. Figure 3 also shows the naive arrangement of equal distance between the centers of biases (set 1) is not the optimized arrangement of replicas.

2. Potential of Mean Force for NH_4^+ + OH^- Salt Bridge in Explicit Solvent. In order to show the effect of USRE optimization on the rate of the convergence for free energy calculations, we applied it to a more computationally demanding system, i.e., NH_4^+ + OH^- salt bridge in explicit solvent. We computed the salt bridge PMF vs N−O distance (inset of Figure 5) by performing two sets of simulations on the salt bridge and for each set we repeated the calculations 10 times. In set 1, we used the naive setting of equal distance between the adjacent bias centers along the reaction coordinate and in the set 2 we optimized the positions of replicas. For the optimization leading to set 2, we used dumped values of the first 20 ps of one of the simulations in set 1 to maximize the FBSF (Figure 4).

Figure 5 shows the mean and root mean square error of Kullback−Leibler (KL) divergence for the computed PMF over 10 independent simulations for each of the optimized and nonoptimized sets. The zero value for Kullback−Leibler divergence means the distribution has converged to the reference distribution (Appendix F). As Figure 5 suggests, on average the optimized set converged to the reference values in less than 20 ps while the nonoptimized set did not converge even after 40 ps. Furthermore, the RMSEs of KL show that each individual simulation in the optimized set converged to the same final answer after 20 ps, while in the nonoptimized set the value of RMSE is still not zero even after 40 ps.

CONCLUSIONS

In the present work, we used statistical mechanics techniques to estimate the Exchange Acceptance Ratios between any two windows on the reaction coordinate for USRE.

Using the proposed schemes we set the positions of replicas such that all replicas have the same EAR with their immediate neighbors. Since in USRE there is no phase transition along the reaction coordinate then this setup maximizes the number of roundtrips between the lowest and highest replicas and consequently the efficiency of USRE. We applied our optimization method to butane in implicit solvent and NH_4^+ + OH^- salt bridge in explicit solvent. From the results, it was evident that this optimization substantially increased the mixing and convergence rate for both systems. We expect the

dx.doi.org/10.1021/ct400366h | J. Chem. Theory Comput. 2013, 9, 4692−4699

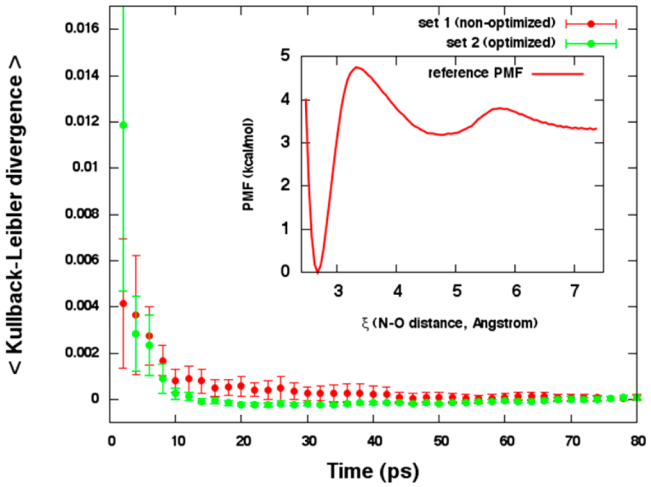


Figure 5. Mean and RMSE of Kullback−Leibler divergence over 10 simulations for each set. Set 1 (red) is the naive equi-distance arrangement and set 2 (green) is our optimized arrangement. We used the PMF of 5 ns USRE simulation as a reference for both simulation sets (the inset). The inset shows PMF vs reaction coordinate (N−O distance).

optimization of USRE to significantly increase the speed of convergence in larger systems.

APPENDIX A: CALCULATING THE PDF OF $\Delta\xi$

In US (and subsequently USRE), approximating the probability density function (PDF) of ξ by a normal distribution is a valid assumption.[58] So e.g., for replica i we can write:

$$p_i(\xi) = \frac{1}{\sqrt{2\pi\sigma_{\xi_i}^2}}\exp\left[-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_{\xi_i}^2}\right] \tag{14}$$

Since the windows are independent, the joint PDF of ξ_i and ξ_j (i.e., $p(\xi_i, \xi_j)$) is equal to multiplication of their PDFs, then we can compute the PDF of $\Delta\xi_{ij} = \xi_i - \xi_j$ by integrating the joint PDF of ξ_i and ξ_j, i.e.,

$$\begin{aligned}
p(\Delta\xi_{ij}) &= \int_{-\infty}^{\infty} \mathrm{d}\xi_i p(\xi_i, \xi_j) \\
&= \int_{-\infty}^{\infty} \mathrm{d}\xi_i p_i(\xi_i) p_j(\xi_j) \\
&= \int_{-\infty}^{\infty} \mathrm{d}\xi_i p_i(\xi_i) p_j(-\Delta\xi_{ij} + \xi_i) \\
&= \frac{1}{2\pi\sigma_{\xi_i}\sigma_{\xi_j}} \int_{-\infty}^{\infty} \mathrm{d}\xi_i \exp\left\{-\frac{1}{2}\left[\frac{(\xi_i-\overline{\xi}_i)^2}{\sigma_{\xi_i}^2} + \frac{(-\Delta\xi_{ij}+\xi_i-\overline{\xi}_j)^2}{\sigma_{\xi_j}^2}\right]\right\} \\
&= \frac{1}{\sqrt{2\pi(\sigma_{\xi_i}^2+\sigma_{\xi_j}^2)}}\exp\left[-\frac{(\Delta\xi_{ij}-(\overline{\xi}_i-\overline{\xi}_j))^2}{2(\sigma_{\xi_i}^2+\sigma_{\xi_j}^2)}\right]
\end{aligned} \tag{15}$$

which itself is a normal distribution.

APPENDIX B: ESTIMATING MEANS AND VARIANCES OF THE WINDOWS USING NNWA AND REWEIGHTING

In order to find the optimal positions of replicas on the reaction coordinate we need to estimate the mean and variance at the new positions of the replicas. The most primitive way to estimate the mean and variance of ξ (i.e., $\overline{\xi}$ and σ^2) for each window is to run a short US simulation at that window. Since optimizing the positions of replicas requires moving the positions (i.e., center of biases) of replicas at each step of optimization, it is not very efficient to use this method for estimating the mean and variance of ξ. Instead, we develop a method below, which approximates $\overline{\xi}$ and σ^2 for any arbitrary replica on the reaction coordinate using the $\overline{\xi}$ and σ^2 of a few, very short time presimulated windows.

Suppose we know the center of bias, variance and mean for N presimulated windows, i.e.,

$$\begin{cases} \xi_1^0 \ldots \xi_N^0 \\ \overline{\xi}_1 \ldots \overline{\xi}_N \\ \sigma_1^2 \ldots \sigma_N^2 \end{cases}$$

if replica i is a presimulated replica, then theoretically we can estimate the first two noncentral moments of the ξ distribution (i.e., $\overline{\xi}$ and $\overline{\xi^2}$) on any new window m on the reaction coordinate using the reweighting technique (Appendix C) as

$$\overline{\xi_{m;i}} = \frac{\int_{-\infty}^{\infty} \xi \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_i^2}\right]\exp[\beta k((\xi-\xi_i^0)^2-(\xi-\xi_m^0)^2)]\,\mathrm{d}\xi}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_i^2}\right]\exp[\beta k((\xi-\xi_i^0)^2-(\xi-\xi_m^0)^2)]\,\mathrm{d}\xi} \tag{16}$$

and

$$\overline{\xi_{m;i}^2} = \frac{\int_{-\infty}^{\infty} \xi^2 \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_i^2}\right]\exp[\beta k((\xi-\xi_i^0)^2-(\xi-\xi_m^0)^2)]\,\mathrm{d}\xi}{\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma_i^2}}\exp\left[-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_i^2}\right]\exp[\beta k((\xi-\xi_i^0)^2-(\xi-\xi_m^0)^2)]\,\mathrm{d}\xi} \tag{17}$$

We can solve the above integrations by writing the power of exponential in eq 17 in a single square form:

$$\begin{aligned}
&-\frac{(\xi-\overline{\xi}_i)^2}{2\sigma_i^2} + \beta k((\xi-\xi_i^0)^2-(\xi-\xi_m^0)^2) \\
&= [-2\overline{\xi}_i\beta k(\xi_i^0-\xi_m^0) + 2\sigma_i^2(\beta k(\xi_i^0-\xi_m^0))^2] \\
&\quad - \frac{(\xi-(\overline{\xi}_i-2\sigma_i^2\beta k(\xi_i^0-\xi_m^0)))^2}{2\sigma_i^2} + \beta k(\xi_i^{0^2}-\xi_m^{0^2})
\end{aligned} \tag{18}$$

and using

$$\begin{cases} \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi C}}\xi\exp\left[-\frac{(\xi-B)^2}{C}\right]\mathrm{d}x = B \\ \int_{-\infty}^{\infty} \frac{1}{\sqrt{\pi C}}\xi^2\exp\left[-\frac{(\xi-B)^2}{C}\right]\mathrm{d}\xi = \frac{C}{2} + B^2 \end{cases} \tag{19}$$

Equations 16 and 17 simplify to

dx.doi.org/10.1021/ct400366h | J. Chem. Theory Comput. 2013, 9, 4692−4699

$$\begin{cases} \overline{(\xi)_{m;i}} = \overline{\xi_i} - 2{\sigma_i}^2\beta k(\xi_i^0 - \xi_m^0) \\ \overline{(\xi^2)_{m;i}} = (\overline{\xi_i} - 2{\sigma_i}^2\beta k(\xi_i^0 - \xi_m^0))^2 + {\sigma_i}^2 \end{cases}$$

If replicas i and $i + 1$ are the nearest presimulated neighbors of our imaginary replica m, i.e., $\xi_i^0 < \xi_m^0 < \xi_{i+1}^0$, then using the NNWA (Appendix D) we can estimate the mean and variance of ξ on windows m as following:

$$\begin{cases} \overline{\xi_m} = w_i(\overline{\xi_i} - 2{\sigma_i}^2\beta k(\xi_i^0 - \xi_m^0)) \\ \quad + w_{i+1}(\overline{\xi_{i+1}} - 2{\sigma_{i+1}}^2\beta k(\xi_{i+1}^0 - \xi_m^0)) \\ \overline{{\xi_m}^2} = w_i[(\overline{\xi_i} - 2{\sigma_i}^2\beta k(\xi_i^0 - \xi_m^0))^2 + {\sigma_i}^2] \\ \quad + w_{i+1}[(\overline{\xi_{i+1}} - 2{\sigma_i}^2\beta k(\xi_{i+1}^0 - \xi_m^0))^2 \\ \quad + {\sigma_{i+1}}^2] \\ {\sigma_m}^2 = \overline{{\xi_m}^2} - \overline{\xi_m}^2 \end{cases} \tag{20}$$

where we choose the power parameter equal to 1.2 and number of involving nearest neighbors equals to 2, so eq 26 in Appendix D implies the following:

$$w_i = \begin{cases} \dfrac{(|\xi_m^0 - \xi_i^0|)^{-0.6}}{(|\xi_m^0 - \xi_i^0|)^{-0.6} + (|\xi_m^0 - \xi_{i+1}^0|)^{-0.6}}; \\ \quad \xi_m^0 \neq \xi_i^0 \text{ and } \xi_m^0 \neq \xi_{i+1}^0 \\ 1;\ \xi_m^0 = \xi_i^0 \\ 0;\ \xi_m^0 = \xi_{i+1}^0 \end{cases} \tag{21}$$

Using the above setup, we are able to estimate the mean and variance of any arbitrary window on the reaction coordinate. Subsequently it is possible to approximate the EAR between any two replicas using eq 12.

APPENDIX C: REWEIGHTING FITTED GAUSSIAN DISTRIBUTION ON WINDOWS

Here we describe the reweighting method that we used for estimation of the mean and the variance of arbitrary windows on the reaction coordinate (Appendix B).

Consider a continuous function $f(\xi)$. We can calculate the average of $f(\xi)$ over a distribution of $p(\xi)$ as follows $\overline{(f(\xi))}_p = (\int f(\xi)p(\xi)\,d\xi)/(\int p(\xi)\,d\xi)$.

If we want to calculate the average of $f(\xi)$ over another distribution $q(\xi)$, where the domain of $q(\xi)$ is a subset of $p(\xi)$, then we may write:

$$\overline{(f(\xi))}_{q;p} = \frac{\int f(\xi)q(\xi)\,dx}{\int q(\xi)\,d\xi} = \frac{\int f(\xi)p(\xi)\frac{q(\xi)}{p(\xi)}\,d\xi}{\int p(\xi)\frac{q(\xi)}{p(\xi)}\,d\xi} = \frac{\overline{\left(\frac{q(\xi)}{p(\xi)}f(\xi)\right)}_p}{\overline{\left(\frac{q(\xi)}{p(\xi)}\right)}_p} \tag{22}$$

where $\overline{(f(\xi))}_{q;p}$ is the average of $f(\xi)$ over distribution of q using the distribution of p.

In the case of umbrella sampling and by assuming Gaussian distributions for the collective variables for every window (i.e., using eq 14), we can estimate the average of $f(\xi)$ over window i:

$$\overline{(f(\xi))}_{p_i} = \frac{\int_{-\infty}^{\infty} f(\xi)\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left[-\frac{(\xi-\overline{\xi_i})^2}{2{\sigma_i}^2}\right]d\xi}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left[-\frac{(\xi-\overline{\xi_i})^2}{2{\sigma_i}^2}\right]d\xi} \tag{23}$$

Using eqs 5 and 6, we can compute the ratio of probability distribution of replicas i and m as

$$\frac{p_m(\xi)}{p_i(\xi)} = \frac{\exp[-\beta(U_{UB} + B_m)]}{\exp[-\beta(U_{UB} + B_i)]} = \exp[\beta k((\xi - \xi_i^0)^2 - (\xi - \xi_m^0)^2)] \tag{24}$$

Finally using eqs 22 and 24, we can exploit

$$\overline{(f(\xi))_{m;i}} = \overline{(f(\xi))}_{p_m;p_i} = \frac{\int_{-\infty}^{\infty} f(\xi)\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left[-\frac{(\xi-\overline{\xi_i})^2}{2{\sigma_i}^2}\right]\exp[\beta k((\xi - \xi_i^0)^2 - (\xi - \xi_m^0)^2)]\,d\xi}{\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi{\sigma_i}^2}}\exp\left[-\frac{(\xi-\overline{\xi_i})^2}{2{\sigma_i}^2}\right]\exp[\beta k((\xi - \xi_i^0)^2 - (\xi - \xi_m^0)^2)]\,d\xi} \tag{25}$$

which is the average of $f(\xi)$ over the window m using the window i. Using this equation we calculate the values of $\overline{\xi}$ and $\overline{\xi^2}$ for any arbitrary windows, using the presimulated windows (Appendix B).

APPENDIX D: NEAREST NEIGHBOR WEIGHTED AVERAGING (NNWA)

We used NNWA to make a better estimation of the mean and variance of ξ for any arbitrary window via presimulated windows. NNWA belongs to the family of locally weighted learning methods.[59] Consider a function $f(\xi)$ with a few known points (i.e., $\{\xi_i\}$) and assume that the value of f at any arbitrary ξ position can be estimated separately from each of those known points. We can then, using k known nearest neighbors of ξ to approximate the value of f as

$$f(\xi) = \sum_{i=1}^{k}\frac{w_i(\xi)f_{\xi_i}(\xi)}{\sum_{i=1}^{k} w_i(\xi)} \tag{26}$$

where $f_{\xi_i}(\xi)$ is the estimation of $f(\xi)$ using ξ_i and $w_i(\xi) = (1/d(\xi, \xi_i))$. d is so-called metric operator, which in its simplest form is $d(\xi, \xi_i) = [(\xi-\xi_i)^2]^{p/2}$. p is a positive real number known as a power parameter. Depending on the problem, k can be changed from the first nearest neighbors to the total number of known points. In this paper, we set $p = 1.2$ and $k = 2$, which helped the optimizer to smoothly converge.

APPENDIX E: BARZILAI AND BORWEIN OPTIMIZATION (BB METHOD)

The steepest descent or gradient descent method is among the most important first-order optimization algorithms. It can be formulized as follows:

$$x_{k+1} = x_k + \lambda_k g_k$$

where $g_k = -\nabla f(x_k)$ and λ_k is the step size at step k.

There are many methods for estimating the optimal step size.[48,60] BB method[47] is one of the most efficient routines among those, in which the step size is determined by minimizing either $\|\Delta x - \lambda\Delta g\|^2$ or its correspondent $\|\Delta g - \lambda^{-1}\Delta x\|^2$ with respect to λ, where $\Delta x = x_k - x_{k+1}$ and $\Delta g = g_k - g_{k+1}$. On the basis of those, the optimal step size can be derived as

dx.doi.org/10.1021/ct400366h | J. Chem. Theory Comput. 2013, 9, 4692−4699

$$\lambda_k = \frac{\Delta x^{\mathrm{T}} \Delta g}{\Delta g^{\mathrm{T}} \Delta g} \tag{27}$$

APPENDIX F: KULLBACK–LEIBLER DIVERGENCE

The Kullback–Leibler divergence[51,61] is defined as

$$D_{\mathrm{KL}}(P\|R) = \sum_{i=1}^{N} P(i)\log(P(i)/R(i)) \tag{28}$$

where P and R are the probability densities of a discrete random variable and N is the number of bins. It can be interpreted as the average of logarithmic difference between the probability distributions of P and R over the distrbition of R. $D_{\mathrm{KL}}(P\|R)$ is always non-negative, and a small value of it indicates that P and R are highly overlapped. It can be used as a metric for convergence and measure how P is different from R, where R is the reference distribution.

AUTHOR INFORMATION

Corresponding Author

*E-mail: roitberg@ufl.edu.

Notes

The authors declare no competing financial interest.

REFERENCES

(1) Earl, D. J.; Deem, M. W. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3910–3916.
(2) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 20227–20232.
(3) Mitsutake, A.; Sugita, Y.; Okamoto, Y. *Biopolymers.* **2001**, *60*, 96–123.
(4) Faller, R.; Yan, Q.; de Pablo, J. J. *J. Chem. Phys.* **2002**, *116*, 5419.
(5) Fenwick, M. K.; Escobedo, F. a. *J. Chem. Phys.* **2003**, *119*, 11998.
(6) Chodera, J. D.; Shirts, M. R. *J. Chem. Phys.* **2011**, *135*, 194110.
(7) Ballard, A. J.; Jarzynski, C. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 12224–12229.
(8) Kästner, J. *WIREs. Comput. Mol. Sci.* **2011**, *1*, 932–942.
(9) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.
(10) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
(11) Bartels, C. *Chem. Phys. Lett.* **2000**, *331*, 446–454.
(12) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141–145.
(13) Meng, Y.; Dashti, D. S.; Roitberg, A. E. *J. Chem. Theory Comput.* **2011**, *7*, 2721–2727.
(14) Sabri Dashti, D.; Meng, Y.; Roitberg, A. E. *J. Phys. Chem. B* **2012**, *116*, 8805–8811.
(15) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
(16) Neumann, J. V. *Proc. Natl. Acad. Sci. U.S.A.* **1932**, *18*, 70–82.
(17) Sugita, Y.; Kitao, A.; Okamoto, Y. *J. Chem. Phys.* **2000**, *113*, 6042.
(18) Itoh, S. G.; Damjanović, A.; Brooks, B. R. *Proteins* **2011**, *79*, 3420–3436.
(19) Woods, C. J.; Essex, J. W.; King, M. A. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.
(20) Fukunishi, H.; Watanabe, O.; Takada, S. *J. Chem. Phys.* **2002**, *116*, 9058.
(21) Liu, P.; Kim, B.; Friesner, R. A.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 13749–13754.
(22) Lyman, E.; Ytreberg, F.; Zuckerman, D. *Phys. Rev. Lett.* **2006**, *96*, 028105.
(23) Okur, A.; Wickstrom, L.; Layten, M.; Geney, R.; Song, K.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2006**, *2*, 420–433.
(24) Cheng, X.; Cui, G.; Hornak, V.; Simmerling, C. *J. Phys. Chem. B* **2005**, *109*, 8220–8230.
(25) Zhou, R.; Berne, B. J. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12777–12782.
(26) Zhou, R. *Proteins* **2003**, *53*, 148–161.
(27) Roitberg, A. E.; Okur, A.; Simmerling, C. *J. Phys. Chem. B* **2007**, *111*, 2415–2418.
(28) Okur, A.; Roe, D. R.; Cui, G.; Hornak, V.; Simmerling, C. *J. Chem. Theory Comput.* **2007**, *3*, 557–568.
(29) Rick, S. W. *J. Chem. Phys.* **2007**, *126*, 054102.
(30) Li, X.; O'Brien, C. P.; Collier, G.; Vellore, N. A.; Wang, F.; Latour, R. A.; Bruce, D. A.; Stuart, S. J. *J. Chem. Phys.* **2007**, *127*, 164116.
(31) Kamberaj, H.; van der Vaart, A. *J. Chem. Phys.* **2009**, *130*, 074906.
(32) Kone, A.; Kofke, D. A. *J. Chem. Phys.* **2005**, *122*, 206101.
(33) Kofke, D. A. *J. Chem. Phys.* **2002**, *117*, 6911.
(34) Kofke, D. A. *J. Chem. Phys.* **2004**, *121*, 1167.
(35) Predescu, C.; Predescu, M.; Ciobanu, C. V *J. Chem. Phys.* **2004**, *120*, 4119–4128.
(36) Sanbonmatsu, K.; Garcia, A. *Proteins: Struct., Funct., Bioinf.* **2002**, *46*, 225–234.
(37) Rathore, N.; Chopra, M.; de Pablo, J. J. *J. Chem. Phys.* **2005**, *122*, 024111.
(38) Katzgraber, H. G.; Trebst, S.; Huse, D. a; Troyer, M. *J. Stat. Mech. Theor. Exp.* **2006**, *2006*, 03018.
(39) Hritz, J.; Oostenbrink, C. *J. Chem. Phys.* **2007**, *127*, 204104.
(40) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.
(41) Shenfeld, D.; Xu, H.; Eastwood, M.; Dror, R.; Shaw, D. *Phys. Rev. E* **2009**, *80*, 046705.
(42) Crooks, G. *Phys. Rev. Lett.* **2007**, *99*, 100602.
(43) Kästner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104.
(44) Kästner, J. *J. Chem. Phys.* **2012**, *136*, 234102.
(45) Murata, K.; Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **2004**, *385*, 1–7.
(46) Metropolis, N.; Rosenbluth, A. W.; Rosenbluth, M. N.; Teller, A. H.; Teller, E. *J. Chem. Phys.* **1953**, *21*, 1087.
(47) Barzilai, J.; Borwein, J. M. *J. Inst. Math. Its App.* **1988**, *8*, 141–148.
(48) Yuan, Y. *AMS/IP Stud. Adv. Math.* **2008**, *42*, 785–796.
(49) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712–3720.
(50) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(51) Kullback, S.; Leibler, R. A. *Ann. Math. Statist.* **1951**, *22*, 79–86.
(52) Case, D. A.; Darden, T. A.; Cheatham, T. E. I.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Kollman, P. A. *AMBER*, version 12; University of California: San Francisco, CA, 2012.
(53) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. a; Case, D. a *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(54) Ryckaert, J. P.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327–341.
(55) Sindhikara, D. J.; Emerson, D. J.; Roitberg, A. E. *J. Chem. Theory Comput.* **2010**, *6*, 2804–2808.
(56) Sindhikara, D. J.; Kim, S.; Voter, A. F.; Roitberg, A. E. *J. Chem. Theory Comput.* **2009**, *5*, 1624–1631.
(57) Grossfield, A. WHAM: an implementation of the weighted histogram analysis method, version 2.0.3, University of Rochester: Rochester, NY, 2010.
(58) Kästner, J. *WIREs. Comput. Mol. Sci.* **2011**, *1*, 932–942.
(59) Atkeson, C.; Moore, A.; Schaal, S. *Artif. Intell. Rev.* **1997**, 11–73.
(60) Burachik, R.; Graña Drummond, L. M.; Iusem, A. N.; Svaiter, B. F. *Optimization* **1995**, *32*, 137–146.
(61) McClendon, C. L.; Hua, L.; Barreiro, A.; Jacobson, M. P. *J. Chem. Theory Comput.* **2012**, *8*, 2115–2126.