Introducing sampling entropy in repository based adaptive umbrella sampling

Han Zheng and Yingkai Zhang^{a)}

Department of Chemistry, New York University, New York, New York 10003, USA

(Received 26 June 2009; accepted 4 November 2009; published online 2 December 2009)

Determining free energy surfaces along chosen reaction coordinates is a common and important task in simulating complex systems. Due to the complexity of energy landscapes and the existence of high barriers, one widely pursued objective to develop efficient simulation methods is to achieve uniform sampling among thermodynamic states of interest. In this work, we have demonstrated sampling entropy (SE) as an excellent indicator for uniform sampling as well as for the convergence of free energy simulations. By introducing SE and the concentration theorem into the biasing-potential-updating scheme, we have further improved the adaptivity, robustness, and applicability of our recently developed repository based adaptive umbrella sampling (RBAUS) approach [H. Zheng and Y. Zhang, *J. Chem. Phys.* **128**, 204106 (2008)]. Besides simulations of one dimensional free energy profiles for various systems, the generality and efficiency of this new RBAUS-SE approach have been further demonstrated by determining two dimensional free energy surfaces for the alanine dipeptide in gas phase as well as in water. © 2009 American Institute of Physics. [doi:[10.1063/1.3267549](https://doi.org/10.1063/1.3267549)]

I. INTRODUCTION

Determination of free energy surfaces with computer simulations is of critical importance in quantitatively elucidating many chemical and biological processes, including conformational changes, molecular recognition, and chemical reactions. Besides challenges regarding accuracy of the potential energy surface and its computational efficiency, another key obstacle is to achieve adequate sampling of the configuration space of interest. Due to the complexity of energy landscapes and the existence of high barriers, transitions among distinct thermodynamic states of significance are rare events, which constitute sampling bottlenecks in a straightforward application of molecular dynamics (MD) or Monte Carlo simulations. A widely employed strategy to overcome this difficulty is the introduction of an external biasing potential into the Hamiltonian which forces the system to visit the high barrier regions.^{1–10}

An ideal choice of biasing potential would be the negative of the free energy surface so that uniform sampling along chosen reaction coordinates can be achieved. Unfortunately, such information is exactly what we try to obtain from simulations and is not known in the first place. In past years, many adaptive sampling approaches^{3,8,11–27} have been proposed to overcome this inherent challenge. Among them, very recently we have introduced the repository based adaptive umbrella sampling (RBAUS) method,²⁷ in which a sampling repository is periodically updated based on the latest simulation data, and the accumulated knowledge and sampling history are then employed to determine whether and how to update the biasing umbrella potential to achieve more uniform sampling for subsequent simulations. In comparison

with other adaptive sampling methods, such as adaptive umbrella sampling,^{3,11–14} the Wang–Landau approach,^{15,16} the adaptive biasing force,^{17,18,28} nonequilibrium metadynamics,^{8,19,20} and self-healing umbrella sampling,²⁴ a unique and attractive feature of the RBAUS method is that the frequency for updating the biasing potential is not predetermined but depends on the sampling history and is adaptively determined on the fly, which smoothly bridges nonequilibrium and equilibrium simulations. Such an adaptive updating is achieved by employing the following general principle: If the subsequent simulations still explore a previously oversampled region more than a previously undervisited one, the biasing potential needs to be updated, otherwise, no updating. Thus we can see that a key element of the RBAUS approach is the determination of sampling uniformity. In addition, a well-defined measure of sampling uniformity can also be used to check the convergence of determined free energy surfaces.

Here, we have introduced the sampling entropy (SE) as a valid indicator for uniform sampling as well as for the convergence of free energy simulations. By introducing SE into the biasing-potential-updating scheme, the adaptivity, robustness, and applicability of the RBAUS approach are significantly improved, including applications to the determination of two-dimensional (2D) free energy surfaces.

II. THEORY AND METHOD

This section is organized as follows. First we give a brief review of the RBAUS method, then we present the rationale for employing SE and the entropy concentration theorem (ECT) in adaptive sampling, and finally the resulting RBAUS-SE approach is described.

^{a)}Electronic mail: yingkai.zhang@nyu.edu.

A. Review of RBAUS

In umbrella sampling^{2,24,27} with a biased Hamiltonian H_b ,

$$H_b = H_0 + V_b(\eta(\mathbf{R})), \quad (1)$$

and the free energy surface along a set of predefined reaction coordinates $\eta(\mathbf{R})$ can be computed as

$$A(\eta) = -k_B T \ln \frac{NS(\eta) e^{V_b(\eta)/k_B T}}{\int d\eta NS(\eta) e^{V_b(\eta)/k_B T}} + c, \quad (2)$$

where $NS(\eta)$ as the number of configurations with given reaction coordinates η in the canonical ensemble generated with the biased Hamiltonian H_b , and c can be any constant. For a given biased ensemble, $NS(\eta)$ is the raw sampling number of η and $K(\eta) = NS(\eta) e^{V_b(\eta)/k_B T}$ can be considered as the corresponding unbiased knowledge of η (Ref. 27) from this biased simulation.

In numerical simulations, it is a common practice to divide reaction coordinates η into Z bins. For the RBAUS approach, the sampling repository stores and updates mainly two kinds of information: the accumulated unbiased knowledge $K[i]$ and the accumulated raw sampling number $NS[i]$ for each bin $[i]$. $K[i]$ is employed to determine the biasing potential V_b with the following formula:

$$V_b[i] = k_B T \ln \frac{K[i] * Z}{\sum_{i=1}^Z K[i]}. \quad (3)$$

Meanwhile, $NS[i]$ is used to compute the “histmax-histmin ratio” (HHR), the ratio of its largest value of $NS[i]$ to its lowest value, which is then employed to determine whether to update the biasing potential or not. However, the HHR is not a global uniform sampling indicator and is not defined before all bins are visited. In our original RBAUS approach, several empirical parameters needed to be carefully chosen to result in an efficient heuristic scheme to determine one-dimensional (1D) free energy profiles. Although tests on several very different systems with barriers ranging from 3 to 30 kcal/mol have been successful, the scheme becomes inefficient when applied to determine 2D free energy surfaces. The problem mainly comes from the inadequacy of HHR as a uniform sampling indicator when the dimension of reaction coordinates and the number of bins increase. Meanwhile, the correlation between HHR and the convergence in determining 1D free energy profiles for those different test systems is quite weak and no general criteria to stop the RBAUS simulations can be suggested. Thus, we are motivated to find a better uniform sampling indicator than HHR so that the robustness, adaptivity, and efficiency of the RBAUS approach can be further significantly improved.

B. Sampling entropy, concentration theorem, and uniform sampling

Given Z bins along chosen reaction coordinates, the SE (Refs. 29 and 30) is defined as

$$SE = - \sum_{i=1}^Z \frac{NS[i]}{NS_{\text{tot}}} \ln \frac{NS[i]}{NS_{\text{tot}}}, \quad (4)$$

where $NS[i]$ is the number of times that the i th bin has been visited and $NS_{\text{tot}} = \sum_{i=1}^Z NS[i]$. We can see that if only one bin is visited, the SE has the minimum value of zero; the SE increases with more bins sampled and it has the maximum value of $\ln Z$ if all bins have been equally sampled. Thus the normalized sampling entropy (NSE), $NSE = SE / \ln Z$, which ranges from 0 to 1, would serve as a general uniform sampling indicator independent of the number of bins. It is dependent on sampling data of all bins and is very well defined even when bins are not sampled. Thus, two main deficiencies of HHR are avoided when using NSE as a uniform sampling indicator.

For complex systems, even with flat free energy surfaces along reaction coordinates of interest, to achieve uniform sampling with MD simulations requires the simulation time to be sufficiently long. At the initial period of simulation, the value of NSE would be expected to be much less than 1 regardless of its underlying free energy landscape. This limitation can be overcome by employing Jayne’s concentration theorem, which states that²⁹ given a number of bins Z with equal probability and N independent sampling points, there is a probability P that the resulting $SE(Z, N)$ should be in the range of

$$\ln Z - \frac{\chi_{Z-1}^2(1-P)}{2N} \leq SE(Z, N) \leq \ln Z, \quad (5)$$

where $\chi_{Z-1}^2(1-P)$ denotes the critical chi squared for $Z-1$ degrees of freedom at the $100(1-P)$ percent significance level. Here the ECT factor is defined as

$$ECT(Z, N) = \ln Z - \frac{\chi_{Z-1}^2(1-P)}{2N} \quad (6)$$

and provides continuing guidance regarding the expectation value of the SE with the increasing independent sampling number N . This can provide a theoretical basis to decide whether to update the biasing potential in adaptive sampling. Typically, the conventional 5% significance level can be employed, such as $\chi_{49}^2(0.05) = 66.34$ and $\chi_{5183}^2(0.05) = 5351.60$. The quantiles of chi-square distribution can be easily obtained by the function “qchisq” in R language.³¹

C. Sampling entropy based RBAUS approach

By employing SE and the concentration theorem in the biasing-potential-updating scheme, the recently developed RBAUS approach has been further improved. The resulting RBAUS-SE simulation scheme is presented in detail in the following steps.

- (0) Initial repository setup. Given Z bins along chosen reaction coordinates, set initial values of $K[i]$ and $NS[i]$ for all bins.
- (1) Biasing-potential update. With the accumulated unbiased knowledge $K[i]$, the free energy profile $A[i]$ is calculated by the following equation:

FIG. 1. The flow chart for the RBAUS approach.

$$A[i] = -k_B T \ln \frac{K[i] * Z}{\sum_{i=1}^Z K[i]}. \quad (7)$$

The biased potential V_b is then obtained based on the negative of the calculated free energy profile. It should be noted that with Eq. (7), $A[i]$ would be zero when $K[i]$ is equal to the average of the accumulated knowledge among all bins—a desired property for constructing the biased potential.

- (2) Biased simulations. Perform MD simulations with the biased Hamiltonian $H_0 + V_b$ and collect the raw histogram $h[i]$, which is the number of sampled configurations in each bin.
- (3) Repository update,

$$NS[i] = NS[i] + h[i], \quad (8)$$

$$K[i] = K[i] + h[i]e^{V_b[i]/k_B T}. \quad (9)$$

The updated $K[i]$ would be employed to calculate the current estimation of the free energy profile $A[i]$ through Eq. (7) if desired. At the same time, from the accumulated raw sampling number $NS[i]$, the SE and the ECT factor $ECT(Z, N)$ are calculated with Eqs. (4) and (6), respectively. It should be noted that even if free energy surface is flat, sampling among all bins would be far from random in MD simulations. Thus the independent sampling number N in Eq. (6) should be proportional to the total sampling number NS_{tot} but needs to be divided by a factor C due to correlations among snapshots in the MD trajectory. In principle, the value of C can be determined by calculating the time correlation function, which would be system dependent. In practice, we found any value between 50 and 200 for C

FIG. 2. Schematic of the alanine dipeptide showing the ϕ and ψ angles.

all yields reasonable performance. Thus we have used C to be 100 throughout this work.

- (4) Make decision on whether to update the biasing potential. If the calculated NSE is larger than 0.99 or the SE is larger than the ECT factor $ECT(Z, N)$, the biasing-potential update is not needed, go to step 2. Otherwise, update is needed, go to step 5.
- (5) Repository consolidation. Here the accumulated unbiased knowledge in the sampling repository is consolidated by setting $K[i] = K[i] * [e^{SE}/e^{ECT(Z, N)}]$ when NSE is less than 0.98. Go to step 1.

The above scheme is summarized in a flow chart shown in Fig. 1, which keeps the same structure as the original RBAUS approach²⁷ but employs SE in steps 4 and 5. It should be noted that 1.0 is the theoretical NSE value for ideal uniform sampling and the $ECT(Z, N)$ provides guidance regarding the expectation value of the SE with the independent sampling number N . Thus in step 4, when NSE is larger than 0.99 or SE is larger than $ECT(Z, N)$, the current biasing potential can be considered to be good enough and thus no update is needed. In step 5, the employment of $e^{SE}/e^{ECT(Z, N)}$ as the scaling factor rather than a fixed value of 0.618 in the original RBAUS scheme makes the repository consolidation step more adaptive based on the current knowledge about uniform sampling. Overall, the RBAUS-SE approach is more robust and general than the original one, as is further demonstrated by test results below. It should be also noted that embarrassingly parallel simulations can be as easily implemented in the RBAUS-SE method as in the original one.

III. IMPLEMENTATION AND COMPUTATIONAL DETAIL

We have implemented the above presented RBAUS-SE simulation scheme in the TINKER4.2 suite of molecular simu-

FIG. 3. The RMSD, NSE, and NECT factor curves for the double well model system with a barrier of 24 kcal/mol. The circle symbols represent the biasing-potential-updating points.

FIG. 4. The RMSD, NSE, and NECT factor curves for the dissociation of a sodium chloride ion pair in aqueous solution. The circle symbols represent the biasing-potential-updating points.

lation program.³² The 1D biasing potential is described with the shape-preserving smoothing spline by using Renka's TSPACK (one-dimension) package,³³ and the 2D biasing potential is represented by employing the surface interpolation

package SRFPACK (Ref. 34) in conjunction with TRIPACK.³⁵ First we tested the implemented scheme to determine the 1D potential of mean force (PMF) for two systems that we have examined previously.²⁷ One is a simple system that contains two particles interacting with a double well potential having a barrier of 24 kcal/mol, and the other is the Na⁺ and Cl⁻ ion pair association in water. For both systems, simulations with four replicas have been carried out, and the repository was updated every 1000 data points collected. To check the convergence of simulations, we calculated the root-mean-square deviation (RMSD) of the simulated PMF with respect to the reference one computed from umbrella sampling,

$$\text{RMSD} = \sqrt{\frac{\sum_{i=1}^Z [\text{PMF}_{\text{sim}}[i] - \text{PMF}_{\text{ref}}[i]]^2}{Z}}. \quad (10)$$

It should be noted that in these two cases, the distance of two particles r is chosen as the reaction coordinate. Thus the PMF(r) is related to the free energy profile $[A(r)]$ as $\text{PMF}(r) = A(r) + 2k_B T \ln r$. To further demonstrating the applicability and performance of the RBAUS-SE approach, we

FIG. 5. Free energy surfaces of the alanine dipeptide system as a function of ϕ and ψ torsional angles in gas phase with 1 kcal/mol contour interval calculated by RBAUS-SE with 24 replica. Upper left: the free energy surface obtained after 10 ps simulation; upper right: the free energy surface obtained after 50 ps simulation; lower left: the free energy surface obtained after 100 ps simulation; lower right: the free energy surface obtained after 500 ps simulation.

FIG. 6. Free energy surfaces of the alanine dipeptide system as a function of ϕ and ψ torsional angles in aqueous phase with 1 kcal/mol contour interval calculated by RBAUS-SE with 24 replica. Upper left: the free energy surface obtained after 10 ps simulation; upper right: the free energy surface obtained after 50 ps simulation; lower left: the free energy surface obtained after 100 ps simulation; lower right: the free energy surface obtained after 500 ps simulation.

FIG. 7. The reference free energy surfaces of the alanine dipeptide system as a function of ϕ and ψ torsional angles in gas phase (left graph) and in aqueous phase (right graph) by biasing equilibrium umbrella sampling for 1 ns with 24 replica.

FIG. 8. The RMSD, NSE, and NECT factor curves for the alanine dipeptide in gas phase. The circle symbols represent the biasing-potential-updating points.

have used it to determine 2D free energy surfaces of the alanine peptide in gas phase and in aqueous solution as a function of the backbone dihedral angle ϕ and ψ , as illustrated in Fig. 2. This has become a classical example to test enhanced sampling approaches.^{20,24,27,36–40}

In our simulations, the alanine dipeptide is described by the CHARMM27 force field.⁴¹ For the aqueous phase, the simulation of one alanine dipeptide molecule and 245 TIP3P water⁴² molecules was performed in a cubic box of 19.7429 Å with periodic boundary conditions. The cutoff distance of 9 Å was used for both van der Waals and electrostatic interactions. For the RBAUS-SE simulations, 24 replicas were employed with time steps of 1 fs and temperature of 300 K. The bin width along the reaction coordinates (ϕ and ψ) was set to be 5° and the repository was updated every 1200 collection data points. The reference free energy surfaces for both gas and aqueous phases were calculated from 24 replica equilibrium MD simulations with a fixed biasing potential. Simulations of 6 ns were run for each replica. The biased potential employed for the equilibrium simulations is obtained from the converged RBAUS-SE simulations. In all our test cases, we have assumed no prior knowledge of the system and set all initial $K[i]$ to be 1. The frequencies to

FIG. 9. The RMSD, NSE, and NECT factor curves for the alanine dipeptide in aqueous phase. The circle symbols represent the biasing-potential-updating points.

FIG. 10. The RMSD curves for the alanine dipeptide in gas phase with both the original RBAUS-HHR approach and the current RBAUS-SE. The faster updating refers that the repository is updated every 1200 data points, which is similar to the repository updating frequency in previous studies (Ref. 27). The slower updating means that repository is updated every 12 000 data points.

collect data were set to be 1, 5, and 2 fs, respectively, for the double well system (each particle has a mass of 1 amu), NaCl, and alanine peptide.

IV. RESULTS AND DISCUSSION

We have applied the RBAUS-SE method to the double well potential model with the barrier of 24 kcal/mol and the dissociation of a NaCl molecule in water. The results are shown in Figs. 3 and 4. Simulation efficiency is found to be very similar to the original RBAUS scheme²⁷ in determining such 1D PMFs, and one attractive advantage of the RBAUS-SE approach is that we can monitor the simulation progress by measuring the NSE. As expected, the error in the simulated PMF decreases with the increase in the NSE. When the NSE first reaches 0.99, the error in the simulated PMF is about 0.1 kcal/mol for the double well system and 0.2 kcal/mol for NaCl in aqueous solution, both of which are below $0.5k_B T$ (about 0.3 kcal/mol when $T=300$ K). Meanwhile, the plotted normalized ECT (NECT) and NSE curves also illustrate how such information is employed to deter-

FIG. 11. The RMSD curves for the alanine dipeptide in aqueous phase with both the original RBAUS-HHR approach and the current RBAUS-SE. The faster updating refers that the repository is updated every 1200 data points, which is similar to the repository updating frequency in previous studies (Ref. 27). The slower updating means that repository is updated every 12 000 data points.

FIG. 12. The correlation between the RMSD in simulated free energies and the NSE for all test systems based on RBAUS-SE simulations

mine whether to update the biasing potential or not. We can see that for both systems, initially these are nonequilibrium simulations and there is no biasing potential updating after about 65 ps for the double well system and after about 250 ps for the NaCl system (Fig. 3 and 4).

To further demonstrate the robustness, efficiency, and generality of the RBAUS-SE method, we employed it to determine 2D free energy surfaces (FES) of the alanine dipeptide in vacuum and water. Figures 5 and 6 illustrate the calculated FES contour maps at different simulation times (10, 50, 100, and 500 ps) for the alanine dipeptide in vacuum and aqueous solution, respectively. In each case, it only takes about 50 ps simulation for each replica to obtain some general features of the 2D free energy surface, and the FES at 500 ps is in excellent agreement with the corresponding reference FES in Fig. 7. The RMSD curves of calculated FES as well as NSE and NECT curves are shown in Figs. 8 and 9, which show that with the progress of simulations, the NSE increases and the biasing-potential update stops at about 0.25 ns for both gas phase and aqueous phase. Those figures also indicate that when the NSE first reaches 0.99, the error in calculated FES is a little less than 0.3 kcal/mol for the alanine dipeptide in vacuum and about 0.2 kcal/mol for the alanine dipeptide in aqueous solution. These results indicate that the RBAUS-SE approach is an efficient approach to determine converged 2D free energy surfaces.

FIG. 13. The correlation between the RMSD in simulated free energies and the NSE for all test systems using equilibrium MD calculations with the inverse of the reference free energy surface as the biasing potential.

As discussed previously, one main motivation to develop this RBAUS-SE approach is the difficulty of applying the original HHR based RBAUS method²⁷ to determine 2D free energy surfaces. Figures 10 and 11 show the performance of both approaches in determining 2D FES for the alanine dipeptide in gas and aqueous phase respectively. For the original RBAUS-HHR approach, it would not be able to yield meaningful results with the repository updating frequency of every 1200 data points, which is similar to the value employed in previous simulations to determine 1D PMF.²⁷ Although the results can be improved by updating every 12 000 data points, the convergence is quite slow and it still has the difficulty for the alanine dipeptide in gas phase. On the other hand, for the RBAUS-SE approach, the error in the calculated FES becomes less than 0.3 kcal/mol after 0.25 ns for all four cases, as shown in Figs. 10 and 11. These results clearly indicate that the new RBAUS-SE approach is a much more efficient and robust approach in comparison with the original RBAUS-HHR approach.²⁷

Figures 12 and 13 show the correlation between the error in simulated free energy surfaces and the NSE in both RBAUS-SE simulations and equilibrium MD simulations without free energy barriers along reaction coordinates of interest for all four test systems. These results clearly demonstrate that the NSE is an excellent indicator for the convergence of free energy simulations as well as for the uniform sampling. Based on these results, a criteria to stop RBAUS-SE simulations can be suggested, such as NSE = 0.993. For all of our four test systems, the error is no more than 0.2 kcal/mol when NSE = 0.993.

V. CONCLUSION

In this work, by employing NSE and the concentration theorem, we have further improved the adaptivity, robustness, and applicability of the RBAUS approach. The tests on alanine peptide in gas phase as well as in water indicate that the RBAUS-SE approach is an efficient approach to determine converged 2D free energy surfaces. Our results clearly demonstrate that the NSE is an excellent indicator for the convergence of free energy simulations as well as for the uniform sampling.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation (Grant No. CHE-CAREER-0448156) and the National Institutes of Health (Grant No. R01-GM079223). We appreciate the computational resources and support provided by NYU-ITS.

¹C. Chipot and A. Pohorille, *Free Energy Calculations: Theory and Applications in Chemistry and Biology* (Springer, New York, 2007).

²G. N. Patey and J. P. Valleau, *J. Chem. Phys.* **63**, 2334 (1975).

³M. Mezei, *J. Comput. Phys.* **68**, 237 (1987).

⁴B. Roux, *Comput. Phys. Commun.* **91**, 275 (1995).

⁵H. Grubmüller, *Phys. Rev. E* **52**, 2893 (1995).

⁶A. F. Voter, *Phys. Rev. Lett.* **78**, 3908 (1997).

⁷C. Jarzynski, *Phys. Rev. Lett.* **78**, 2690 (1997).

⁸A. Laio and M. Parrinello, *Proc. Natl. Acad. Sci. U.S.A.* **99**, 12562 (2002).

⁹J. A. Rahman and J. C. Tully, *J. Chem. Phys.* **116**, 8750 (2002).

- ¹⁰D. Hamelberg, J. Mongan, and J. A. Mccammon, *J. Chem. Phys.* **120**, 11919 (2004).
- ¹¹C. Bartels and M. Karplus, *J. Comput. Chem.* **18**, 1450 (1997).
- ¹²C. Bartels, M. Schaefer, and M. Karplus, *J. Chem. Phys.* **111**, 8048 (1999).
- ¹³R. Rajamani, K. J. Naidoo, and J. L. Gao, *J. Comput. Chem.* **24**, 1775 (2003).
- ¹⁴J. Wang, Y. Gu, and H. Y. Liu, *J. Chem. Phys.* **125**, 094907 (2006).
- ¹⁵F. Calvo, *Mol. Phys.* **100**, 3421 (2002).
- ¹⁶F. G. Wang and D. P. Landau, *Phys. Rev. E* **64**, 056101 (2001).
- ¹⁷E. Darve and A. Pohorille, *J. Chem. Phys.* **115**, 9169 (2001).
- ¹⁸J. Hénin and C. Chipot, *J. Chem. Phys.* **121**, 2904 (2004).
- ¹⁹G. Bussi, A. Laio, and M. Parrinello, *Phys. Rev. Lett.* **96**, 090601 (2006).
- ²⁰B. Ensing, M. De Vivo, Z. W. Liu, P. Moore, and M. L. Klein, *Acc. Chem. Res.* **39**, 73 (2006).
- ²¹H. Li, D. Min, Y. Liu, and W. Yang, *J. Chem. Phys.* **127**, 094101 (2007).
- ²²A. Laio and F. L. Gervasio, *Rep. Prog. Phys.* **71**, 126601 (2008).
- ²³T. Lelièvre, M. Rousset, and G. Stoltz, *J. Chem. Phys.* **126**, 134111 (2007).
- ²⁴S. Marsili, A. Barducci, R. Chelli, P. Procacci, and V. Schettino, *J. Phys. Chem. B* **110**, 14011 (2006).
- ²⁵A. Barducci, G. Bussi, and M. Parrinello, *Phys. Rev. Lett.* **100**, 020603 (2008).
- ²⁶L. Q. Zheng and W. Yang, *J. Chem. Phys.* **129**, 014105 (2008).
- ²⁷H. Zheng and Y. Zhang, *J. Chem. Phys.* **128**, 204106 (2008).
- ²⁸E. Darve, M. A. Wilson, and A. Pohorille, *Mol. Simul.* **28**, 113 (2002).
- ²⁹E. T. Jaynes, *Proc. IEEE* **70**, 939 (1982).
- ³⁰S. Park and V. S. Pande, *Phys. Rev. E* **76**, 016703 (2007).
- ³¹R. Development Core Team, *R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing*, <http://www.R-project.org> (2007).
- ³²J. W. Ponder, TINKER, Software Tools for Molecular Design, version 4.2, the most updated version for the TINKER program can be obtained from J. W. Ponder's worldwide website at <http://dasher.wustl.edu/tinker>, June 2004.
- ³³R. J. Renka, *ACM Trans. Math. Softw.* **19**, 81 (1993).
- ³⁴R. J. Renka, *ACM Trans. Math. Softw.* **22**, 9 (1996).
- ³⁵R. J. Renka, *ACM Trans. Math. Softw.* **22**, 1 (1996).
- ³⁶D. J. Tobias and C. L. Brooks, *J. Phys. Chem.* **96**, 3864 (1992).
- ³⁷P. E. Smith, *J. Chem. Phys.* **111**, 5568 (1999).
- ³⁸L. Rosso, J. B. Abrams, and M. E. Tuckerman, *J. Phys. Chem. B* **109**, 4162 (2005).
- ³⁹C. A. F. De Oliveira, D. Hamelberg, and J. A. Mccammon, *J. Chem. Phys.* **127**, 175105 (2007).
- ⁴⁰J. Apostolakis, P. Ferrara, and A. Caflisch, *J. Chem. Phys.* **110**, 2099 (1999).
- ⁴¹A. D. MacKerell, Jr., D. Bashford, M. Bellott, R. L. Dunbrack, Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin, and M. Karplus, *J. Phys. Chem. B* **102**, 3586 (1998).
- ⁴²W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, *J. Chem. Phys.* **79**, 926 (1983).