Thermodynamics of a conformational change using a random walk in energy-reaction coordinate space: Application to methane dimer hydrophobic interactions

A. N. Morozov and S. H. Lin

Citation: The Journal of Chemical Physics **130**, 074903 (2009); doi: 10.1063/1.3077658
View online: http://dx.doi.org/10.1063/1.3077658
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/130/7?ver=pdfcov
Published by the AIP Publishing

Thermodynamics of a conformational change using a random walk in energy-reaction coordinate space: Application to methane dimer hydrophobic interactions

A. N. Morozov[1,2,3,a] and S. H. Lin[1,2]
[1]*National Chiao Tung University, 1001 Ta Hsuen Road, Hsinchu, Taiwan Republic of China*
[2]*Institute of Atomic and Molecular Sciences, Academia Sinica, P.O. Box 23-166, Taipei, Taiwan 10617, Republic of China*
[3]*Institute of Surface Chemistry, National Academy of Sciences of Ukraine, Generala Naumova str. 17, Kyiv 03164, Ukraine*

(Received 31 October 2008; accepted 12 January 2009; published online 19 February 2009)

A random walk sampling algorithm allows the extraction of the density of states distribution in energy-reaction coordinate space. As a result, the temperature dependences of thermodynamic quantities such as relative energy, entropy, and heat capacity can be calculated using first-principles statistical mechanics. The strategies for optimal convergence of the algorithm and control of its accuracy are proposed. We show that the saturation of the error [Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003); E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007)] is due to the use of histogram flatness as a criterion of convergence. An application of the algorithm to methane dimer hydrophobic interactions is presented. We obtained a quantitatively accurate energy-entropy decomposition of the methane dimer cavity potential. The presented results confirm the previous results, and they provide new information regarding the thermodynamics of hydrophobic interactions. We show that the finite-difference approximation, which is widely used in molecular dynamic simulations for the energy-entropy decomposition of a free energy potential, can lead to a significant error. © *2009 American Institute of Physics*. [DOI: 10.1063/1.3077658]

I. INTRODUCTION

A free energy potential $F(r)$ describes the relative thermodynamics for a conformational change of a system with a large number of degrees of freedom as a function of a reaction coordinate r (this usually represents a limited number of coordinates related to the conformational change of interest); this is performed by averaging over the ensemble of configurations in a state r,[1]

$$F(r) = -k_B T \ln W(r) = -k_B T \ln \int \delta(r - r'(\mathbf{x}))w(\mathbf{x})d\mathbf{x}, \tag{1}$$

where $W(r)$ is the probability density of observing the macrostate r at a temperature T, $\mathbf{x}$ is a point in the configurational space, $w(\mathbf{x})$ is the probability density in the $\mathbf{x}$ space, k_B is Boltzmann's constant, and δ is the Dirac delta. For methane dimer hydrophobic association-dissociation in a water solution, for example, the distance between a pair of the methane particles serves as the reaction coordinate, while averaging is performed over the solvent degrees of freedom. In complex systems, the mapping of a potential function $F(r)$ is hindered by the low statistical weights of the transitional states and/or by a rugged energy landscape. To alleviate these issues, different implementations of a nonphysical sampling approach are used in conjunction with molecular dynamic (MD) or Monte Carlo (MC) simulations.[2–7] Upon calculating the free energy potential, the relative stability of the different conformations can be found. However, to better understand the mechanism of the conformational transition, it is important to also know the related changes in thermodynamic quantities such as energy, entropy, and heat capacity. Due to large fluctuations in the energy and entropy quantities (in proportion to the total number of particles), direct estimations of their relative changes by classical MD/MC methods are extremely inefficient.[8] Few indirect methods have been developed to calculate the thermodynamics of a conformational change using MD simulations. Brooks[9] proposed a method based on the evaluation of derivatives of the partition function. Smith *et al.*[10] used the thermodynamic integration approach and the finite-difference approximation[8] (FDA) to estimate the entropy component of a free energy potential. Among these methods, the FDA shows better performance.[8] Mainly the FDA method is used for the energy-entropy decomposition of a free energy potential or related potential of mean force (PMF).[11–18] Despite the use of long MD simulations, calculations of energy and entropy changes are still subject to large statistical errors. Consequently, such calculations are relatively rare, and their results are of a qualitative level.

Here, we show that reliable theoretical calculations of relative thermodynamic quantities can be achieved using the algorithm of a random walk in energy-reaction coordinate (r,E) space. The output of this algorithm is the joint density of states (JDOS) $g(r,E)$. With the $g(r,E)$ function in hand, the temperature dependences of relative energy, entropy, and

[a]Electronic mail: morozov@gate.sinica.edu.tw.

heat capacity can be calculated from a first-principles approach. Other useful thermodynamic information, e.g., the FDA error, can be estimated as well. The algorithm of a random walk to extract the JDOS is better known for calculations of thermal averages.[19–26] Apparently the JDOS algorithm for (r,E) variables is a promising tool for calculations of the thermodynamics of conformational changes as well.[25]

The JDOS algorithm is a straightforward extension of the original Wang–Landau (WL) sampling algorithm.[27,28] It should be noted that MD or classical Metropolis MC (Ref. 29) simulations produce an "importance sampling" that is in the vicinity of the maximum of the statistical distribution; meanwhile, the WL algorithm generates, upon convergence, "a random walk in energy space" by accepting stochastic configurations with a probability that is proportional to the reciprocal of the density of states. Such an organized stochastic walk moves randomly on the complex energy landscape, where the classical MD/MC methods used may be trapped in the concurrent potential wells.

The implementation of the WL/JDOS method in its original formulation leads to the saturation of the error (nonconvergence) of the algorithm.[21,30] Here we show that the nonconvergence is due to the use of an arbitrary chosen value of histogram flatness as the criterion of convergence. Based on the results of Morozov and Lin,[31] we propose in this work an efficient criterion for the convergence of the WL/JDOS algorithm. The accuracy of the WL algorithm depends on the value of the modification factor.[31,32] It also depends on the density of states function and the method used to pick up a trial state in the stochastic process.[31] Here we propose a strategy to improve accuracy by tuning the stochastic process. The JDOS algorithm provides better sampling than the original WL algorithm by maintaining the correlations of the joint variables;[24] however, it is also more complex and costly to calculate. The improvements and modifications proposed to overcome the calculation difficulties often sacrifice universality and elegance of the original WL method (see a comprehensive comparison made by Poulain *et al.*[24]). Here we propose strategies for the optimal performance of the WL/JDOS algorithm while preserving a universal character and simplicity of its implementation.

In this work, the MC algorithm of a random walk in (r,E) space is applied for calculations of hydrophobic interactions of the methane dimer in a water solution. The hydrophobic effect[33] plays an important role in the conformational transitions of solvated biomolecules[34,35] and, along with hydrogen bonding, is believed to be a determining factor in protein folding.[36–38] For the simple model of the hydrophobic effect, the methane dimer in a water solution was intensively studied using a variety of force fields and MD/MC methods.[8,10,12,14,15,39–42]

The rest of the paper is organized as follows. First, in Sec. II we briefly review the statistical mechanics background for the thermodynamics of a conformational change. Then, we describe the algorithm of a random walk in (r,E) space and the proposed strategies for convergence, accuracy, and performance of the algorithm. Section IV contains the details of the presented simulations. In Sec. V, we compare our result for the methane dimer PMF to those of the previous MD simulations and present the quantitatively accurate results for the relative energy and entropy quantities. The estimation of the FDA error is presented in this section as well.

II. STATISTICAL MECHANICS FORMULATIONS

According to Eq. (1), the probability density $W(r)$ is the projection of the probability density $w(\mathbf{x})$ on the r space. The projection of $w(\mathbf{x})$ on the (r,E) space gives the probability density $W(r,E)$. Since in the canonical ensemble $w(\mathbf{x}) \equiv w(E(\mathbf{x}))$, one finds that

$$W(r,E) = \int \delta(r - r'(\mathbf{x}))\delta(E - E'(\mathbf{x}))w(\mathbf{x})d\mathbf{x} = w(E)g(r,E), \tag{2}$$

where $w(E)=Z^{-1}\exp(-E/k_BT)$ and Z is the partition function. Equation (1) for a free energy potential can be rewritten using the probability density $W(r,E)$ as follows:

$$F(r) = -k_BT \ln \int W(r,E)dE = -k_BT \ln \frac{Z_r}{Z}, \tag{3}$$

where $Z_r=\int g(r,E)\exp(-E/k_BT)dE$ is the contribution of the state r to the partition function. The relative energy quantity $E(r)$ reads

$$E(r) = -T^2 \frac{\partial}{\partial T}\left(\frac{F(r)}{T}\right)_V = \langle E\rangle_r - \langle E\rangle. \tag{4}$$

In Eq. (4), $\langle\cdots\rangle_r$ designates the average over the ensemble of configurations in the state r; for example, the average energy $\langle E\rangle_r$ is given by the relationship

$$\langle E\rangle_r = Z_r^{-1}\int Eg(r,E)\exp(-E/k_BT)dE. \tag{5}$$

It is convenient to express the relative entropy quantity $S(r)$ using $\langle E\rangle_r$,

$$S(r) = -\left(\frac{\partial F(r)}{\partial T}\right)_V = T^{-1}(\langle E\rangle_r - \langle E\rangle - F(r)). \tag{6}$$

Similarly, the relative heat capacity quantity $c_V(r)$ is

$$c_V(r) = T\left(\frac{\partial S(r)}{\partial T}\right)_V = \frac{\langle E^2\rangle_r - \langle E\rangle_r^2}{k_BT^2} - c_V. \tag{7}$$

The temperature dependence of $c_V(r)$ is related to the error of the FDA. The FDA requires three MD simulations at temperatures T and $T\pm\Delta T$ to obtain the approximate entropy quantity $S_{\mathrm{FDA}}(r)$ from the relationship

$$S_{\mathrm{FDA}}(r) = -\frac{F(r)|_{T+\Delta T} - F(r)|_{T-\Delta T}}{2\Delta T}. \tag{8}$$

The series expansion of the right hand side of Eq. (8) to the third degree over ΔT gives an estimation of the $S_{\mathrm{FDA}}(r)$ error, i.e.,

$$S_{\mathrm{FDA}}(r) \cong S(r) - \frac{1}{6}\frac{\partial^3 F(r)}{\partial T^3}\Delta T^2$$

$$= S(r) + \frac{1}{6}\frac{\partial}{\partial T}\left(\frac{c_V(r)}{T}\right)_V \Delta T^2. \quad (9)$$

Using Eq. (7) for $c_V(r)$, one can obtain from Eq. (9) the relationship

$$TS_{\mathrm{FDA}}(r) \cong TS(r) + \frac{1}{6}\left[\frac{\langle \Delta E^3\rangle_r/k_BT - 3\langle \Delta E^2\rangle_r}{k_BT}\right]\left(\frac{\Delta T}{T}\right)^2$$

$$+ \frac{T}{6}\frac{\partial}{\partial T}\left(\frac{c_V}{T}\right)_V \Delta T^2, \quad (10)$$

where ΔE is $\Delta E = E - \langle E\rangle_r$. From this, the deviation $T(S_{\mathrm{FDA}} - S)$ can be estimated, providing that the function $g(r,E)$ is known.

III. A RANDOM WALK IN (r,E) SPACE

Originally, the WL algorithm[27,28] was proposed for discrete systems. We further assume that for continuous systems a binning scheme with discretization width Δr and ΔE is used to fulfill a piecewise constant approximation of the density of states function (later we shall discuss the choice of Δr, ΔE). A random walk in (r,E) space is organized as follows. The transition probability for the stochastic process $\mathbf{x}_1 \to \mathbf{x}_2 \cdots \to \mathbf{x}_{l-1} \to \mathbf{x}_l \to \cdots$ in the configurational space $\mathbf{x}$ is proportional to the reciprocal of the current approximation of the density of states,

$$w(\mathbf{x}_{l-1} \to \mathbf{x}') = \min\left[\frac{\exp S_i(r(\mathbf{x}_{l-1}),E(\mathbf{x}_{l-1}))}{\exp S_i(r(\mathbf{x}'),E(\mathbf{x}'))}, 1\right], \quad (11)$$

where $\mathbf{x}_{l-1}$ and $\mathbf{x}'$ are the old and trial configurations, respectively, the subscript i denotes the ith iteration of the algorithm, and $S_i(r,E) = \ln g_i(r,E)$ is the dimensionless entropy. For the new configuration $\mathbf{x}_l$, the dimensionless entropy $S_i(r,E)$ and the histogram $H_i(r,E)$ are modified according to the relationships

$$S_i(r(\mathbf{x}_l),E(\mathbf{x}_l)) \to S_i(r(\mathbf{x}_l),E(\mathbf{x}_l)) + \ln f_i,$$

$$H_i(r(\mathbf{x}_l),E(\mathbf{x}_l)) \to H_i(r(\mathbf{x}_l),E(\mathbf{x}_l)) + 1, \quad (12)$$

where f_i denotes the modification factor. For the first iteration, the initial modification factor $f_1 > 1$ and initial approximation $S_1^{(0)}(r,E)$ must be chosen. The beginning of the ith iteration is described by the relationships

$$S_i^{(0)}(r,E) = S_{i-1}(r,E), \quad f_i = \sqrt[a]{f_{i-1}}, \quad H_i(r,E) = 0. \quad (13)$$

For example, in the original Wang and Landau[27,28] works, the values $f_1 = e = 2.7182\ldots$, $S_1^{(0)} = 0$, and $a = 2$ were chosen. Finally, when $f_i \to 1$ and the stochastic chain is long enough, the approximation $S_i(r,E)$ converges to within an additive constant of the exact dimensionless entropy $S(r,E)$. It should be noted that the JDOS algorithm does not satisfy detailed balance and cannot be approximated by the Marcovian chain.

A. Convergence

For long enough MC chains, using the exact density of states in Eq. (11) produces a flat histogram, i.e., $H(r,E) \cong$ const. Therefore, histogram "flatness" can be used as a criterion of convergence.[27,28] Flatness which can be defined as $x = \max(|H - \langle H\rangle|/\langle H\rangle)$, where H runs over all histogram entries and $\langle H\rangle$ denotes the average histogram, is a system dependent criterion.[28] Moreover, as mentioned in Sec. I, the choice of some constant value of x will lead to the saturation of the algorithm error[21,30] when $f_i \to 1$. Another approach is to control the number of visits per histogram entry. In our previous work,[31] this type of convergence condition was obtained for a one-dimensional random walk in E space [see Eq. (24) in Ref. 31] by extrapolation of the rigorous result for a two-state system onto a system with multilevel energy spectra. The proposed criterion in Ref. 31 depends on the system properties and on the implementation of the MC process as well. However, using the same approach as in Ref. 31, it is easy to show (see Appendix A) that with an increase in the error given by a factor of $\sqrt{2}$, the criterion of convergence can be derived in a system and MC independent manner. That is, each entry of the histogram must satisfy the following inequality:

$$H_i \geq \frac{\ln a}{2 \ln f_i}. \quad (14)$$

Convergence to the two-dimensional surface $g_i(r,E)$ requires satisfaction of the criterion in expression (14) for at least two independent directions in (r,E) space. Providing that the distribution of trial states has no preferred direction in (r,E) space, the convergence of a random walk can be controlled by the criterion

$$H_i \geq \frac{\ln a}{\ln f_i}. \quad (15)$$

With increasing i the number of MC steps due to the criterion in Eq. (24) of Ref. 31 grows as $\sim i/\ln f_i$, while inequalities (14) and (15) result in $\sim 1/\ln f_i$ MC steps. The use of inequality (15) requires the exclusion of the histogram entries that, due to physical reasons, are not accessible for a random walk. Troster and Dellago[43] proposed the self-adaptive method to determine the accessible entries of the histogram. For the solvated methane dimer (see Sec. IV for the simulation details), we calculated the histogram flatness at convergence of the ith iteration controlled by the criterion in expression (15). The value of histogram flatness x, averaged over 16 MC trajectories, is shown in Fig. 1. The result shows that the use of histogram flatness as the criterion of convergence is inefficient both at the early and late stages of the iterative process. At the early stages, it leads to excessive calculations. Meanwhile at $f_i \to 1$, the arbitrary chosen constant x finally leads to the above mentioned error saturation.

Since inequality (15) must be satisfied for each accessible entry of the histogram, the choice of a too small discretization will result in slow convergence. On the other hand, large discretization can lead to miscalculations of thermal averages.[24] For chosen set of discretization parameters, it is necessary to check that a decrease in their values does

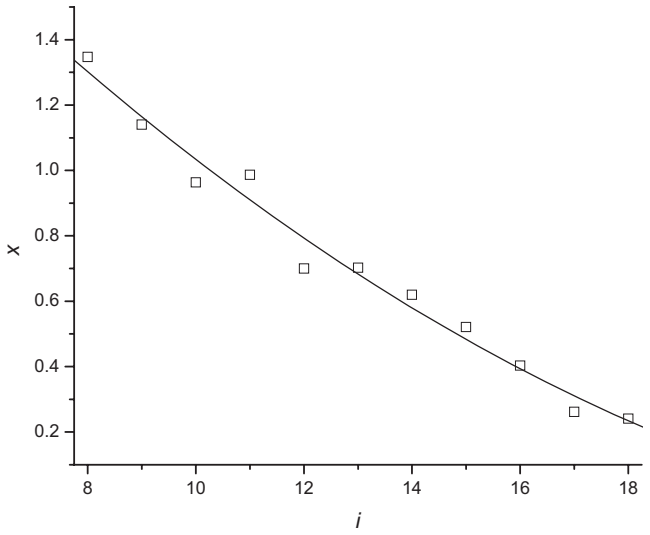

FIG. 1. Histogram flatness x for the solvated methane dimer system (averaged over 16 MC trajectories) calculated at convergence of the ith iteration controlled by the criterion in expression (15).

not change the thermodynamic functions of interest. To save calculation time, such a check can be done on small-size system.

B. Accuracy

For a physical quantity A we shall estimate the accuracy of thermal averaging $\Delta_i^{(A)}$,

$$\langle A\rangle_i = \langle A\rangle + \Delta_i^{(A)}, \tag{16}$$

where $\langle A\rangle_i$ and $\langle A\rangle$ are the results of thermodynamic averaging using, respectively, the density of states due to the ith iteration and the exact density of states. It is convenient to define the error of the density of states as follows:

$$g_{i,\mathbf{m}} = g_{\mathbf{m}} c_i \exp(\Delta_{i,\mathbf{m}}), \tag{17}$$

where subscript $\mathbf{m}$ describes the two-dimensional (r,E) lattice due to discretization of (r,E) space, $g_{\mathbf{m}}$ denotes the exact density of states, and c_i is the multiplicative constant. The value of $\Delta_{i,\mathbf{m}}$ can be estimated by calculating the standard deviation $\sigma(\Delta_{i,\mathbf{m}})$ for the distribution of $\Delta_{i,\mathbf{m}}$ over all possible MC trajectories.[31,32] Taking into account that $\sigma(\Delta_{i,\mathbf{m}})$ in the limit $\ln f_i \to 0$ scales[31,32] as $\sqrt{\ln f_i}$, one can represent $\Delta_i^{(A)}$ as follows:

$$\Delta_i^{(A)} = \Delta_A \sqrt{\ln f_i}, \tag{18}$$

where, in a zeroth-order approximation over $\ln f_i$, the factor Δ_A depends only on a particular MC trajectory. The factor Δ_A can be estimated as $\sqrt{\overline{\Delta_A^2}}$ (a line over the symbol denotes averaging over MC trajectories). Assuming that K statistically independent trajectories are available, Δ_A can be estimated using the formula

$$\Delta_A \cong \pm \sqrt{\frac{\sum_{k1=1}^{K}\sum_{k2>k1}^{K}(\langle A\rangle_{i,k1} - \langle A\rangle_{i,k2})^2}{K(K-1)\ln f_i}}, \tag{19}$$

where $\langle A\rangle_{i,k}$ denotes averaging using the density of states at the ith iteration of the kth trajectory. The cross terms $\Delta_{i,k1}^{(A)}\Delta_{i,k2}^{(A)}$ are neglected in the right hand side of Eq. (19). The reason for this is that $\overline{\Delta_{i,\mathbf{m}}}$ scales as $\ln f_i$ [see Eq. (18) in Ref. 31]. Therefore, $\overline{\Delta_i^{(A)}}$ also scales as $\ln f_i$. Hence, for statistically independent trajectories, the estimation $\overline{\Delta_{i,k1}^{(A)}\Delta_{i,k2}^{(A)}} = (\overline{\Delta_i^{(A)}})^2 \sim (\ln f_i)^2$ is valid. Providing that $K \gg 1$, the contribution of the cross terms on the right hand side of Eq. (19) tends to zero in the limit $\ln f_i \to 0$. Next, we take into account that averaging over K statistically independent MC trajectories reduces[32] the standard deviation of $\Delta_{i,\mathbf{m}}$,

$$\sigma_{i,K} \cong \sigma_i/\sqrt{K}. \tag{20}$$

Therefore, the error of thermal averaging is also reduced by the same factor, i.e.,

$$\Delta_{i,K}^{(A)} = \Delta_A \sqrt{K^{-1}\ln f_i}. \tag{21}$$

Since the calculation of statistically independent MC trajectories also improves performance of the WL/JDOS algorithm (see Sec. III C), Eqs. (19) and (21) provide a practical method for estimating the error of thermal averaging for a physical quantity A.

The error of a random walk sampling depends not only on the modification factor but also on the density of states and the implementation of the MC algorithm.[31] Let us introduce the relative error for a pair of histogram entries $\mathbf{m}$ and $\mathbf{k}$,

$$\frac{g_{i,\mathbf{m}}}{g_{i,\mathbf{k}}} = \frac{g_{\mathbf{m}}}{g_{\mathbf{k}}}\exp(\Delta_{i,\mathbf{mk}}) = \left(1 + \frac{\Delta g_{\mathbf{mk}}}{g_{\mathbf{k}}}\right)\exp(\Delta_{i,\mathrm{mk}}), \tag{22}$$

where $\Delta_{i,\mathbf{mk}} = \Delta_{i,\mathbf{m}} - \Delta_{i,\mathbf{k}}$, $\Delta g_{\mathbf{mk}} = g_{\mathbf{m}} - g_{\mathbf{k}}$. If with high probability $\mathbf{m}$ and $\mathbf{k}$ entries occur as the nearest neighbors on the MC trajectory, we will refer to $\Delta_{i,\mathbf{mk}}$ as the local error. For the case $\ln f_i \to 0$, the estimation of the local error is given by[31]

$$\sigma(\Delta_{i,\mathbf{mk}}) \cong k\sqrt{\left(1 + \frac{|\Delta g_{\mathbf{mk}}|}{g_{\mathbf{k}}}\right)\ln f_i}, \tag{23}$$

where k is the factor of the order of magnitude unity [it should be noted that for the distant entries the scaling relationship $\sigma(\Delta_{i,\mathbf{mk}}) \sim \sqrt{\ln f_i}$ is different from Eq. (23) due to correlations of the MC process]. For a particular MC trajectory, the distribution of $\Delta g_{\mathbf{mk}}$ in Eq. (23) depends on the implementation of a method to pick up a trial state. It is reasonable to assume that for a given f_i, the local error of the WL/JDOS algorithm is determined by the average of $\langle|\Delta g_{\mathbf{mk}}|\rangle$ per one MC step over the MC trajectory. In a linear approximation,

$$\frac{\langle|\Delta g_{\mathbf{mk}}|\rangle}{g_{\mathbf{k}}} \cong \left\langle \left| \frac{\partial S}{\partial r}(r_{m_1} - r_{k_1}) + \frac{\partial S}{\partial E}(E_{m_2} - E_{k_2}) \right| \right\rangle \le \left|\frac{\partial S}{\partial r}\right| \langle|(r_{m_1} - r_{k_1})|\rangle + \left|\frac{\partial S}{\partial E}\right| \langle|(E_{m_2} - E_{k_1})|\rangle, \tag{24}$$

where $\langle|(r_{m_1} - r_{k_1})|\rangle$ and $\langle|(E_{m_2} - E_{k_1})|\rangle$ are the averages of the absolute change of r and E per one MC step, respectively. We propose that in an efficient implementation of the MC process, these averages are equal to the corresponding width of discretization, i.e.,

$$\langle|(r_{m_1} - r_{k_1})|\rangle = \Delta r, \langle|(E_{m_2} - E_{k_1})|\rangle = \Delta E. \quad (25)$$

It follows from Eqs. (23) and (24), for a given modification factor f_i, that this strategy minimizes the WL/JDOS local error by adjusting the parameters of the MC process. On the other hand, because convergence depends on the number of $\mathbf{m}\rightarrow\mathbf{k}$ and $\mathbf{k}\rightarrow\mathbf{m}$ transitions,[31] the choice of $\langle|(r_{m_1}-r_{k_1})|\rangle < \Delta r$ or $\langle|(E_{m_2}-E_{k_1})|\rangle < \Delta E$ will result in poor convergence of the algorithm. It should be noted that according to Eq. (24), the error of the WL/JDOS algorithm increases with increases in the gradient of the dimensionless entropy S.

C. Performance

We shall describe the strategies to improve the performance of the algorithm assuming that K processor units are available. First, let us consider a slowdown of the expansion of the volume of random walking on the initial stage, $i=1$, of the simulation. The initial slowdown is an important issue if the number of the histogram entries is large. Zhou *et al.*[25] described the mechanism of the initial slowdown and proposed a modification of the algorithm to overcome this issue. This modification is characterized by the complicated and system dependent input.[24] On the other hand, the parallel implementation of the algorithm[19] for K overlapping windows is an effective way to improve performance on the initial stage without complicating the input of the algorithm and without sacrificing its universal character.

Second, we shall describe the strategy to improve performance of the WL/JDOS algorithm for the case $\ln f_i \rightarrow 0$. At this stage, the parallel implementation of the algorithm bears the additional cost of calculations due to the boundary regions of the overlapping windows. Also, the parallel implementation can cause an increase in the error that, to our knowledge, has not yet been investigated. Instead it is more efficient to calculate K statistically independent MC trajectories. Comparing Eq. (18) and Eq. (21), one can see that the error given by Eq. (21) can be achieved in the standard one-trajectory WL/JDOS algorithm at the iteration with the modification factor $\sqrt[K]{f_i}$. Using expression (15), one can easily estimate that for $a=2$, averaging over K independent MC trajectories is almost two times cheaper compared to the equivalent one-trajectory calculation starting with f_i and going up to $\sqrt[K]{f_i}$. Consequently, as it follows from Eqs. (17) and (21), in the case of K statistically independent trajectories, expression (13) of the WL/JDOS algorithm is modified to

$$S_i^{(0)}(r,E) = K^{-1}\sum_{k=1}^{K} S_{i-1,k}(r,E), \quad f_i = \sqrt[K]{f_{i-1}}, H_i(r,E) = 0, \quad (26)$$

where index k designates the kth MC trajectory. It should be noted that if, because of a computer precision limit or high cost of computer calculations, the modification factor cannot be reduced further, the accuracy of the WL/JDOS algorithm can still be improved by accumulating the number of the statistically independent trajectories.

IV. SIMULATION DETAILS

Water-water interactions were approximated by the TIP3P potential.[44] For methane-methane interactions, the optimized Lennard-Jones potential[45]

$$V(r) = 4V_{\text{meth}}\left[\left(\frac{R_{\text{meth}}}{r}\right)^{12} - \left(\frac{R_{\text{meth}}}{r}\right)^{6}\right] \quad (27)$$

was used with $V_{\text{meth}}=0.294$ kcal/mol and $R_{\text{meth}}=3.73$ Å. The Lennard-Jones parameters for the water-methane interactions were obtained using the standard Lorentz–Berthelot combination rules.[46] MC simulations were performed using the periodic boundary conditions and shifting of the potentials at a cutoff radius $R_{\text{cut}}=9$ Å. The long range effect of Coulomb interactions was taken into account using the reaction field method[47] with the dielectric constant of surrounding continuum ε_{rf} equal to the dielectric constant of TIP3P water at $T=300$ K,[48] i.e., $\varepsilon_W=92$. Although the choice $\varepsilon_{\text{rf}}=\varepsilon_W$ is in accordance with the original motivation of the reaction field approximation, any constant value of ε_{rf} in the range $\varepsilon_W \leq \varepsilon_{\text{rf}} \leq \infty$ is acceptable for the boundary conditions.[47] First, we placed 216 water molecules in a cubic box with a box side of about 18.7 Å, which corresponds to the equilibrium density of the TIP3P water model at $T=300$ K and a pressure of 1 atm.[16] Using the gradient descent method, the configuration with energy $E_{\text{mol}} \cong -9.5$ kcal/mol was obtained (E_{mol} designates energy per molecule). Next, two water molecules were replaced by two methane particles. This configuration was used as the starting point for the MC simulations.

The simulation was started in the parallel implementation[19] with $f_1=e=2.7182\ldots$, $S_1^{(0)}=0$, and $a=2$. The depth of the methane-methane potential well was chosen as the energy discretization width, i.e., $\Delta E=V_{\text{meth}}$. Such discretization was successfully used by Yan *et al.*[20] in simulations of a Lennard-Jones fluid. The discretization width of the reaction coordinate was set to $\Delta r=0.015R_{\text{meth}}$. We confirmed that within an accuracy of 0.01 kcal/mol, decreases in the discretization parameters to $\Delta E=V_{\text{meth}}/2$ and $\Delta r=0.015R_{\text{meth}}/2$ do not change the free energy potential of the methane dimer solvated by 114 water molecules. The parameters of the random movements were adjusted to satisfy the relationships in expression (25). On average, for each random movement of a water molecule, one movement of a methane particle occurred. Accordingly, the convergence was controlled by inequality (15).

The initially chosen energy range -11.5 kcal/mol $\leq E_{\text{mol}} \leq -7$ kcal/mol was divided into four overlapping windows. In each window, the reaction coordinate range was $0.875R_{\text{meth}} \leq r \leq R_{\text{cut}}$. The parallel calculation was stopped at the iteration $\ln f=2^{-13}$, and the canonical distribution function $\rho(E,T)$ was calculated for temperatures of 273 and 373 K. At this point, the energy range was reduced to -10.45 kcal/mol $\leq E_{\text{mol}} \leq -8.1$ kcal/mol by applying the conditions $\rho/\rho_{\max} \geq 10^{-8}$, $E<\langle E\rangle$, $T=273$ K for low energies and $\rho/\rho_{\max} \geq 10^{-8}$, $E>\langle E\rangle$, $T=373$ K for high energies [$\rho_{\max}$ designates the maximum of the canonical distribution function, $\rho_{\max}=\rho(\langle E\rangle,T)$]. Next, 16 statistically independent trajectories with $\ln f=2^{-14}$ were started using the result of the

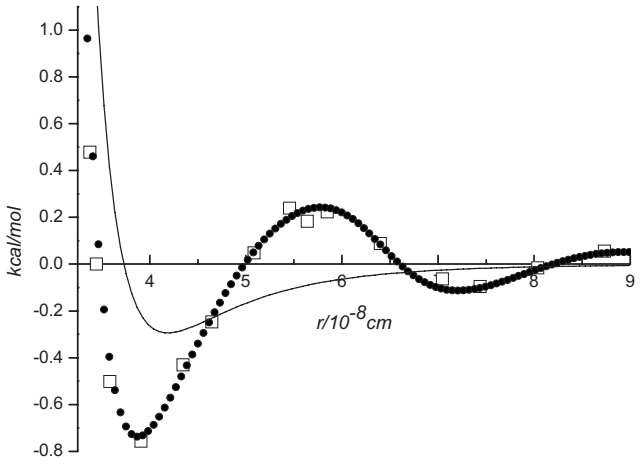

FIG. 2. The PMF of the methane dimer at $T=300$ K: (●) due to a random walk in (r,E) space, the absolute error is ± 0.004 kcal/mol; (□) due to MD simulations by Ghosh *et al.* (Ref. 15). The solid line represents the methane dimer potential energy.

parallel calculation as the initial approximation. After convergence of all 16 trajectories, the next iteration with $\ln f = 2^{-18}$ was started according to expression (26). A total of 64 trajectories ($K=64$) were accumulated with $\ln f = 2^{-18}$ to obtain the relative energy and entropy quantities with high resolution. The calculations were performed using eight Intel Core2 Duo E8500 processors. Approximately 5×10^8 MC sweeps per particle were needed to converge a trajectory with $\ln f = 2^{-18}$. One sweep takes approximately 2.5×10^{-3} s.

V. RESULTS AND DISCUSSION

To exclude the volume effect from a free energy potential of a solvated dimer, it is convenient to use a PMF,

$$G(r) = -k_B T \ln \frac{W(r)}{W_{\mathrm{id}}(r)}, \tag{28}$$

where $G(r)$ denotes the PMF and $W_{\mathrm{id}}(r)$ is the probability density for two ideal particles located at a distance r. It can be easily found that the energy part of the PMF is equal to

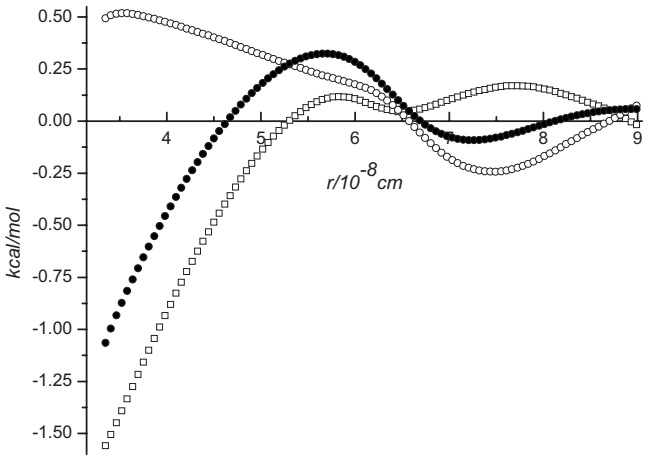

FIG. 3. (●) The cavity potential of the methane dimer at $T=300$ K decomposed into (○) the energy (± 0.04 kcal/mol) and (□) the entropy (± 0.04 kcal/mol) parts.

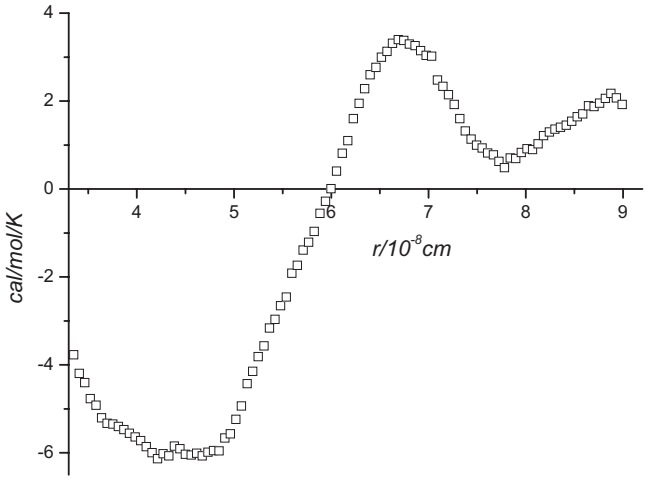

FIG. 4. The relative heat capacity quantity (± 2 kcal/mol) at $T=300$ K.

that of the free energy potential, while the exclusion of the volume effect changes the entropy. The entropy part of $G(r)$ reads

$$S(r) = -\left(\frac{\partial G(r)}{\partial T}\right)_V = T^{-1}(\langle E\rangle_r - \langle E\rangle - G(r)). \tag{29}$$

The $G(r)$ at $T=300$ K for the methane dimer, computed using Eqs. (3) and (28), is shown in Fig. 2. For comparison, the result of long MD simulations by Ghosh *et al.*[15] is also shown in Fig. 2. Using Eq. (19) with $K=64$, we estimated that the inequality $|\Delta_G| \le 14$ kcal/mol is valid for the range $0.875R_{\mathrm{meth}} \le r \le R_{\mathrm{cut}}$. Therefore, substitution of $\ln f = 2^{-18}$ and $K=64$ into Eq. (21) gives the estimation of the PMF absolute error to be $\Delta G(r) \approx \pm 0.004$ kcal/mol. Similarly, for the energy quantity $E(r)$, the estimation of the absolute error yields $\Delta E(r) \approx \pm 0.04$ kcal/mol. Since the PMF error is one order of magnitude lower than $\Delta E(r)$, the error of the entropy quantity is the same as for $\Delta E(r)$, i.e., $\Delta(TS(r)) \approx \pm 0.04$ kcal/mol.

The methane dimer potential energy is shown in Fig. 2 by a solid line. Subtraction of the potential energy from the

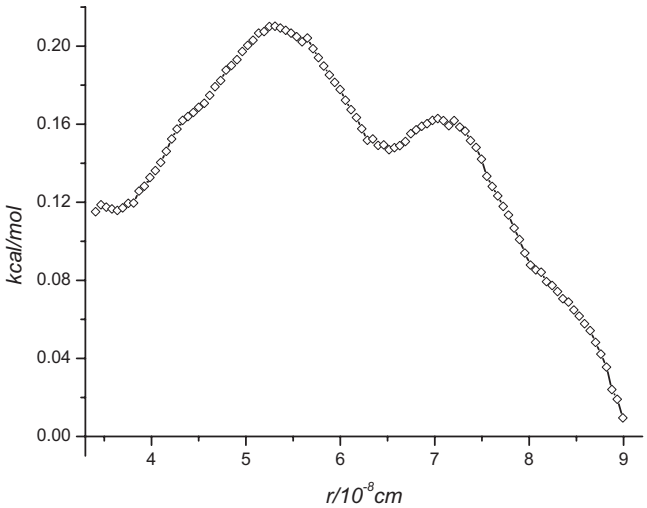

FIG. 5. The FDA error (relative to its value at $r=9$ Å) due to the heat capacity temperature effect for $\Delta T=25$ K and $T=300$ K. The absolute accuracy is ± 0.05 kcal/mol.

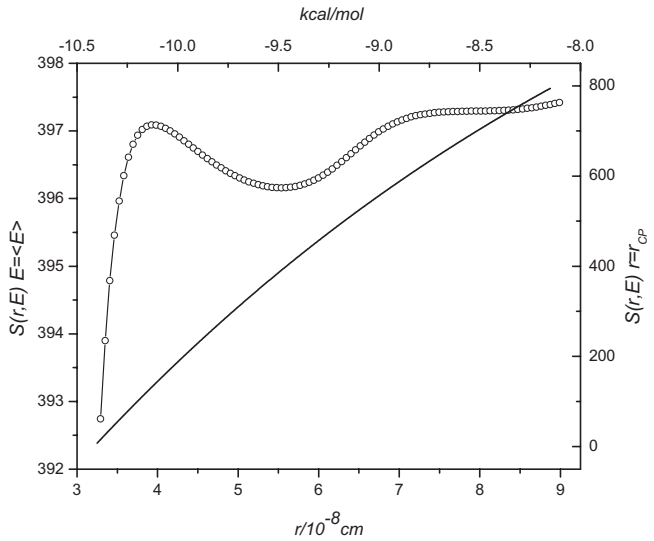

FIG. 6. The one-dimensional cross sections of $S(r,E)$: (○) at constant $E=\langle E\rangle$ (axis: x—at bottom; y—at left); solid line: at constant $r=3.9$ Å (axis: x is E per molecule—at top; y—at right).

PMF gives the solvent contribution to the PMF, which is known as the cavity potential. In the frame of the chosen water and methane models, the presented calculations allow the quantitatively accurate energy-entropy decomposition of the cavity potential. The energy, $E_{\mathrm{cav}}(r)=E(r)-V(r)$, and entropy, $-TS(r)$, parts of the cavity potential are shown in Fig. 3. In line with previous studies, our results show that the entropic stabilization[8] of the contact pair (CP) minimum at around 3.9 Å is opposed by the energy part of the cavity potential. On the contrary, the energetic stabilization[15] of the solvent (SS) separated minimum at around 7.3 Å is opposed by the maximum of the entropy part of the cavity potential. The high accuracy of the presented calculations allows for the detection of features of the methane cavity potential that, to our knowledge, have not been reported previously. First, the desolvation barrier at around 5.8 Å is contributed to by the related maximum of the entropic contribution. Second, the unfavorable energy contribution to the cavity potential decreases almost linearly between 4 and 6 Å.

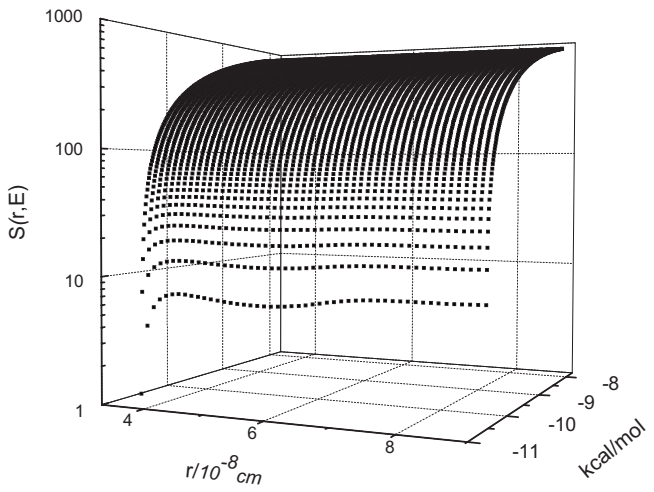

FIG. 7. The two-dimensional surface $S(r,E)$. The thermally averaged relative error at $T=300$ K is 0.002.

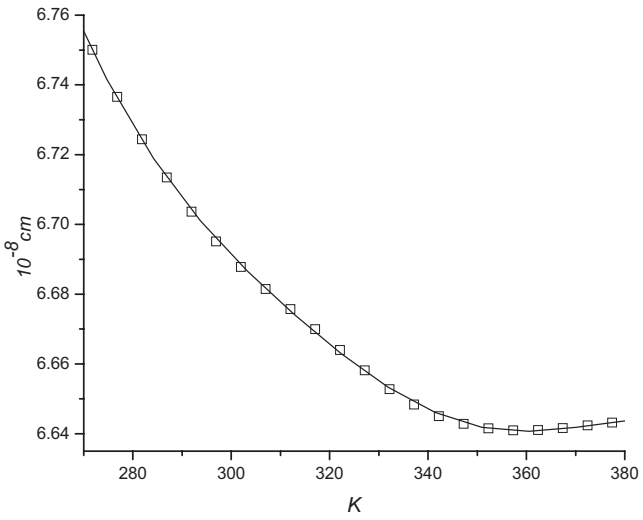

FIG. 8. The temperature dependence of the average separation (±0.02 Å) of the methane particles.

An accurate calculation of the relative heat capacity quantity is more computationally expensive than that of $E(r)$. Our result for $c_V(r)$ is shown in Fig. 4. The estimation of the absolute error of $c_V(r)$ gives $\Delta c_V(r)\approx\pm 2$ cal/mol K. Thus, the presented result is of a qualitative level. In agreement with previous studies,[49–51] we found that $c_V(r)$ shows a peak at the desolvation barrier in the region around 6 Å, followed by the deep minimum around 4.8 Å with a negative heat capacity contribution for the CP state.

Figure 5 shows the deviation $T(S_{\mathrm{FDA}}-S)$ relative to its value at $r=9$ Å for $\Delta T=25$ K and $T=300$ K. The maximum deviation reaches about 0.22 ± 0.05 kcal/mol at the desolvation barrier. Though the systematic error of the FDA is usually obscured by the statistical error of the MD simulations, our result proves that the heat capacity temperature effect cannot be neglected. Since the deviation $T(S_{\mathrm{FDA}}-S)$ is proportional to the second power of $\Delta T/T$, the use of the FDA at room temperatures requires $\Delta T\leq 20$ K to obtain an accurate energy-entropy decomposition to within 0.1 kcal/mol. Thus, the assumed range in Ref. 8 of $\Delta T\sim 30-50$ K is not satisfactory for quantitatively accurate calculations. It should be noted that MD simulations of the solvated xenon dimer[16] indicate that the heat capacity effect of the TIP4P (Ref. 44) and TIP5P (Ref. 52) water models is stronger in comparison with that of the TIP3P model. Consequently the FDA error may be larger for these water models.

The one-dimensional cross sections of the dimensionless entropy surface at constant $E_{\mathrm{mol}}=-9.52$ kcal/mol ($E=\langle E\rangle$) and constant $r=3.9$ Å(CP distance) are shown in Fig. 6, while the two-dimensional surface $S(r,E)$ is shown in Fig. 7. One can see that sampling using a random walk in (r,E) space reveals with high accuracy (the estimation of the thermally averaged relative error gives $\Delta S/S\cong\pm 0.002$ at $T=300$ K) the energy-reaction coordinate correlations of the density of states. Even in the relatively simple case of the solvated methane dimer, these correlations are characterized by the complex landscape and by the relative differences to many orders of magnitude. In addition to the efficient sampling, another advantage is that the output of the WL/JDOS

algorithm allows thermal averaging of the physical values of interest at different temperatures. For example, the temperature dependence of the average distance between the methane particles is shown in Fig. 8.

VI. CONCLUSION

With the proposed strategies in the present paper, the algorithm of a random walk in (r,E) space [the JDOS algorithm for (r,E) variables] should prove useful for calculations of the thermodynamics of a conformational change in complex molecular systems, e.g., for calculations of solvated biomolecules. While the present paper is limited to the thermodynamics of the canonical ensemble, generalization of the algorithm to other ensembles is straightforward.

APPENDIX A: CRITERION OF CONVERGENCE

In our previous work,[31] we showed [see Eq. (17) in Ref. 31] that in the case of a one-dimensional random walk in E space, the local error of the density of states, $\Delta_{i,\mathbf{mk}}$ [see Eq. (22) for the definition], is the sum of the term due to the initial deviation Δ_{init} and the stochastic term Δ_{stoch}. From this,

$$\overline{\Delta_{i,\mathbf{mk}}^2} = \Delta_{\text{init}}^2 + 2\Delta_{\text{init}}\overline{\Delta_{\text{stoch}}} + \overline{\Delta_{\text{stoch}}^2}, \tag{A1}$$

where a line over the symbols denotes averaging over the MC trajectories. We have shown [see Eqs. (18) and (19) in Ref. 31] that in the limit $\ln f_i \to 0$,

$$\overline{\Delta_{\text{stoch}}} \sim \ln f_i, \quad \sigma(\Delta_{\text{stoch}}) \sim \sqrt{\ln f_i}. \tag{A2}$$

From the relationships in expression (A2), it follows that

$$\overline{\Delta_{i,\mathbf{mk}}^2} \cong \Delta_{\text{init}}^2 + 2\Delta_{\text{init}}\overline{\Delta_{\text{stoch}}} + \sigma(\Delta_{\text{stoch}})^2. \tag{A3}$$

In Ref. 31 we showed that with increasing length of MC chain, $\Delta_{\text{init}} \to 0$. The criterion of convergence in Eq. (24) of Ref. 31 was obtained by applying the condition

$$\Delta_{\text{init}} \ll \sigma(\Delta_{\text{stoch}}). \tag{A4}$$

As a result, using Eqs. (A2)–(A4), one can easily estimate the local error, i.e., $\sqrt{\langle \Delta_{i,\mathbf{mk}} \rangle^2} \cong \sigma(\Delta_{\text{stoch}})$. The condition of convergence can be then modified to

$$\Delta_{\text{init}} = \sigma(\Delta_{\text{stoch}}). \tag{A5}$$

In this case, as it follows from Eqs. (A2)–(A4), an increase in the local error is given by a factor of $\sqrt{2}$, i.e., $\sqrt{\langle \Delta_{i,\mathbf{mk}} \rangle^2} \cong \sqrt{2}\sigma(\Delta_{\text{stoch}})$. From Eqs. (20)–(24) in Ref. 31, one can easily find that the condition in Eq. (A5) gives for the histogram entries

$$H_i \geq \frac{\ln a}{2 \ln f_i}. \tag{A6}$$

This criterion of convergence does not depend on the specific properties of the simulated system or on the implementation of the MC process.

[1] S. H. Northrup, M. R. Pear, C. Y. Lee, J. A. McCammon, and M. Karplus, Proc. Natl. Acad. Sci. U.S.A. **79**, 4035 (1982).
[2] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[3] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988).
[4] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **63**, 1195 (1989).
[5] J. Lee, Phys. Rev. Lett. **71**, 211 (1993).
[6] S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman, J. Comput. Chem. **13**, 1011 (1992).
[7] R. W. W. Hooft, B. P. van Eijck, and J. Kroon, J. Chem. Phys. **97**, 6690 (1992).
[8] D. E. Smith and A. D. J. Haymet, J. Chem. Phys. **98**, 6445 (1993).
[9] C. L. Brooks III, J. Phys. Chem. **90**, 6680 (1986).
[10] D. E. Smith, L. Zhang, and A. D. J. Haymet, J. Am. Chem. Soc. **114**, 5875 (1992).
[11] D. J. Tobias and C. L. Brooks III, J. Chem. Phys. **92**, 2582 (1990).
[12] S. W. Rick and B. J. Berne, J. Phys. Chem. B **101**, 10488 (1997).
[13] E. Gallicchio, M. M. Kubo, and R. M. Levy, J. Phys. Chem. B **104**, 6271 (2000).
[14] N. Matubayasi and M. Nakahara, J. Phys. Chem. B **104**, 10352 (2000).
[15] T. Ghosh, A. E. Garcia, and S. Garde, J. Chem. Phys. **116**, 2480 (2002).
[16] D. Paschek, J. Chem. Phys. **120**, 6674 (2004).
[17] N. Choudhury and B. M. Pettitt, J. Phys. Chem. B **110**, 8459 (2006).
[18] M. V. Athawale, S. Sarupria, and S. Garde, J. Phys. Chem. B **112**, 5661 (2008).
[19] M. S. Shell, P. G. Debenedetti, and A. Z. Panagiotopoulos, Phys. Rev. E **66**, 056703 (2002).
[20] Q. Yan, R. Faller, and J. J. de Pablo, J. Chem. Phys. **116**, 8745 (2002).
[21] Q. Yan and J. J. de Pablo, Phys. Rev. Lett. **90**, 035701 (2003).
[22] D. P. Landau, S. H. Tsai, and M. Exler, Am. J. Phys. **72**, 1294 (2004).
[23] E. A. Mastny and J. J. de Pablo, J. Chem. Phys. **122**, 124109 (2005).
[24] P. Poulain, F. Calvo, R. Antoine, M. Broyer, and P. Dugourd, Phys. Rev. E **73**, 056704 (2006).
[25] C. G. Zhou, T. C. Schulthess, S. Torbrugge, and D. P. Landau, Phys. Rev. Lett. **96**, 120201 (2006).
[26] S. H. Tsai, F. G. Wang, and D. P. Landau, Braz. J. Phys. **38**, 6 (2008).
[27] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[28] F. Wang and D. P. Landau, Phys. Rev. E **64**, 056101 (2001).
[29] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, J. Chem. Phys. **21**, 1087 (1953).
[30] R. E. Belardinelli and V. D. Pereyra, J. Chem. Phys. **127**, 184105 (2007).
[31] A. N. Morozov and S. H. Lin, Phys. Rev. E **76**, 026701 (2007).
[32] C. G. Zhou and R. N. Bhatt, Phys. Rev. E **72**, 025701(R) (2005).
[33] B. Widom, P. Bhimalapuram, and K. Koga, Phys. Chem. Chem. Phys. **5**, 3085 (2003).
[34] C. L. Brooks III, M. Karplus, and B. M. Pettit, *Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics* (Wiley, New York, 1987).
[35] A. N. Morozov and S. H. Lin, J. Phys. Chem. B **110**, 20555 (2006).
[36] K. A. Dill, Biochemistry **29**, 7133 (1990).
[37] H. J. Dyson, P. E. Wright, and H. A. Scheraga, Proc. Natl. Acad. Sci. U.S.A. **103**, 13057 (2006).
[38] A. N. Morozov, Y. J. Shiu, C. T. Liang, M. Y. Tsai, and S. H. Lin, J. Biol. Phys. **33**, 255 (2007).
[39] C. Pangali, M. Rao, and B. J. Berne, J. Chem. Phys. **71**, 2982 (1979).
[40] S. Ludemann, H. Schreiber, R. Abseher, and O. Steinhauser, J. Chem. Phys. **104**, 286 (1996).
[41] W. S. Young and C. L. Brooks III, J. Chem. Phys. **106**, 9265 (1997).
[42] E. Sobolewski, M. Makowski, C. Czaplewski, A. Liwo, S. Oldziej, and H. A. Scheraga, J. Phys. Chem. B **111**, 10765 (2007).
[43] A. Troster and C. Dellago, Phys. Rev. E **71**, 066705 (2005).
[44] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).
[45] W. L. Jorgensen, J. D. Madura, and C. J. Swenson, J. Am. Chem. Soc.

106, 6638 (1984).
[46] M. P. Allen and D. G. Tidesley, *Computer Simulation of Liquids* (Oxford University Press, New York, 1987), Chap. 1.4.2, p. 21.
[47] M. Neumann, J. Chem. Phys. **82**, 5663 (1985).
[48] G. Lamoureux, J. A. D. MacKerell, and B. Roux, J. Chem. Phys. **119**, 5185 (2003).
[49] S. Shimizu and H. S. Chan, J. Am. Chem. Soc. **123**, 2083 (2001).
[50] S. W. Rick, J. Phys. Chem. B **107**, 9853 (2003).
[51] D. Paschek, J. Chem. Phys. **120**, 10605 (2004).
[52] M. W. Mahoney and W. L. Jorgensen, J. Chem. Phys. **112**, 8910 (2000).