PHYSICAL REVIEW E **85**, 066702 (2012)

Free-energy calculation via mean-force dynamics using a logarithmic energy landscape

Tetsuya Morishita,[1,*] Satoru G. Itoh,[2] Hisashi Okumura,[2] and Masuhiro Mikami[1,†]
[1]*Nanosystem Research Institute (NRI), National Institute of Advanced Industrial Science and Technology (AIST), 1-1-1 Umezono, Tsukuba, Ibaraki 305-8568, Japan and CREST-JST, 4-1-8 Honcho, Kawaguchi, Saitama 332-0012, Japan*
[2]*Department of Theoretical and Computational Molecular Science, Institute for Molecular Science, Okazaki, Aichi 444-8585, Japan and Department of Structural Molecular Science, Graduate University for Advanced Studies, Okazaki, Aichi 444-8585, Japan*

(Received 14 March 2012; published 11 June 2012)

A method for free-energy calculation based on mean-force dynamics (fictitious dynamics on a potential of mean force) is presented. The method utilizes a logarithmic form of free energy to enhance crossing barriers on a free-energy landscape, which results in efficient sampling of "rare" events. Invoking a conserved quantity in mean-force dynamics, free energy can be estimated on-the-fly without postprocessing. This means that an estimate of the free-energy profile can be locally made in contrast to the other methods based on mean-force dynamics such as metadynamics. The method is benchmarked against conventional methods and its high efficiency is demonstrated in the free-energy calculation for a glycine dipeptide molecule.

DOI: 10.1103/PhysRevE.85.066702 PACS number(s): 05.10.−a, 02.70.Ns, 82.20.Wt, 87.10.Tf

I. INTRODUCTION

Reliable and efficient techniques for sampling energy landscapes are crucial to obtain thermodynamic properties and free-energy profiles of complex systems using molecular simulations. This is particularly significant in biosystems that usually exhibit rugged energy landscapes [1]. To tackle this problem, methods that sample "rare" events efficiently, often referred to as biased sampling approaches, have been widely used [2]. Biased sampling such as umbrella sampling [3] enables us to survey an energy landscape more than the sampling based on the usual Boltzmann weight. Conformations (or states) that are rarely realized in a normal thermostatted MD run can be easily sampled using biased sampling and thus it has been applied to a variety of complex systems [2]. Biased sampling is typically used for exploring potential-energy landscapes. It however can also be applied to exploring free-energy landscapes using "mean-force dynamics."

Recently, to efficiently obtain free-energy profiles with respect to a set of collective variables $\{X_1,...,X_N\}$ ($\equiv\mathbf{X}$), mean-force dynamics (MFD) has been proposed [4] in which $\mathbf{X}$ are treated as dynamical variables whose time evolution is determined by dynamical equations of motion using "mean force" (MF). MF is the force acting on $\mathbf{X}$ and is a slope of free energy $F(\mathbf{X})$ (or "potential of mean force").

Metadynamics can be seen as a free-energy calculation method based on MFD utilizing biased sampling [5]. In this method, $\mathbf{X}$ are driven by MF and the additional force derived from a time-dependent potential $V(t)$, the latter being the key to implementing the biased sampling of $F(\mathbf{X})$. $V(t)$ works as a biased potential so that $\mathbf{X}$ show a random walk without being trapped in local free-energy minima, compensating its original free-energy landscape. $F(\mathbf{X})$ is then estimated as $-V(t\to\infty)$ once a flat free-energy surface (for $\mathbf{X}$) is obtained.

Adiabatic free-energy dynamics (AFED) [4,6] or temperature-accelerated MD (TAMD) [7] are also free-energy calculation methods based on MFD. Although the normal Boltzmann distribution $e^{-F/k_BT'}$ is designed to be sampled in these methods, the temperature T' for $\mathbf{X}$ is artificially set to be high enough to enhance crossing energy barriers on a free-energy landscape, which is equivalent to the uniform scaling of the free energy.

It has been demonstrated that both the methods are of use in constructing free-energy profiles of complex systems [5,6,8,9]. However careful adjustments of the parameters used in the methods are necessary, particularly in the former case [5]. Furthermore, an estimate of the free-energy profile in a certain region does depend on how accurately the "whole" free-energy landscape is sampled, which might degrade the efficiency in free-energy calculations.

In this paper, we introduce a method for free-energy calculation based on MFD. The method is developed to ideally meet the following requirements: (1) Realizing rare events by simple modification of the free-energy landscape, (2) the free energy can be locally estimated, and (3) the free energy is calculated on-the-fly without postprocessing. To meet (1), we employ a logarithmic form of free energy as the potential energy function in MFD. In the following, we show how the method, *logarithmic mean-force dynamics* (LogMFD), yields free-energy profiles satisfying these three requirements.

II. THEORY

We consider the following equations of motion (EOM) for the dynamical variable X_i,

$$M_i\ddot{X}_i = -\left(\frac{\alpha\gamma}{\alpha F+1}\right)\frac{\partial F}{\partial X_i} - M_i\dot{X}_i\dot{\eta}, \tag{1}$$

$$Q\ddot{\eta} = \left(\sum_i M_i\dot{X}_i^2 - Nk_BT_X\right), \tag{2}$$

where M_i is the fictitious mass for X_i, η is the thermostat variable in the Nosé-Hoover thermostat [10] that controls the temperature for X_i, T_X, and Q is the mass for η. The force acting on X_i comes from the logarithmic form

---

[*] t-morishita@aist.go.jp; http://staff.aist.go.jp/t-morishita/

[†] Present address: Global Leader Carrier Development Network, University of Tsukuba, 1-1-1 Tenno-dai, Tsukuba, Ibaraki, 305-8577, Japan.

of $F(\mathbf{X})$, $\gamma\log(\alpha F+1)$, where α and γ are the positive parameters we set so as to efficiently sample the landscape of $F(\mathbf{X})$ (for simplicity, γ can often be taken to be $1/\alpha$). To avoid $(\alpha F+1)<1$, F is shifted by c, if necessary, so that $F=F^{'}+c>0$ for the whole $\mathbf{X}$ range we focus on. $-\partial F/\partial X_i$ is the mean force estimated by the following relation:

$$\begin{aligned}\frac{\partial F(X_0)}{\partial X_i} &= \frac{1}{Z}\int d\mathbf{q}\left[\frac{\partial\Phi(\mathbf{q})}{\partial X_i}\right]_{X_i=X_0} e^{-\Phi(\mathbf{q})/k_BT}\\ &\simeq \frac{1}{\tau}\int_0^{\tau} dt\left[\frac{\partial\Phi(\mathbf{q})}{\partial X_i}\right]_{X_i=X_0},\end{aligned} \tag{3}$$

where

$$Z=\int d\mathbf{q}e^{-\Phi(\mathbf{q})/k_BT}, \tag{4}$$

and $\Phi(\mathbf{q})$ is the potential energy of the system, being a function of particles' positions $\mathbf{q}$. As discussed in detail in Ref. [4], the right-hand side of Eq. (3) is assumed to be obtained by a short-time average of $\partial\Phi/\partial X_i$ in a thermostatted (canonical) MD simulation. Thus, a short canonical MD run with fixed $\mathbf{X}$ is performed each time $\mathbf{X}$ is updated by Eqs. (1) and (2) (each MFD step may consist of ~5–100 MD steps). $\mathbf{X}$ itself can be either dependent on or independent of $\mathbf{q}$ [11]. In the former case, a harmonic restraining potential is often used to fix $\mathbf{X}$ in the short MD run as $1/2k[\mathbf{X}(\mathbf{q})-\tilde{\mathbf{X}}]^2$, where k is the spring constant. $\tilde{\mathbf{X}}$ are independent of $\mathbf{q}$ and are treated as dynamical variables instead of $\mathbf{X}$.

We note that in metadynamics and AFED (TAMD), $\tilde{\mathbf{X}}$ and $\mathbf{q}$ evolve simultaneously with the same time step using large M_i, which ensures their adiabatic separation as in the Car-Parinello MD [12]. This can also be followed in LogMFD. A multiple time scale algorithm such as r-RESPA [13] is particularly effective in this case. One however should take notice that forces on $\mathbf{q}$ and $\tilde{\mathbf{X}}$ are derived from different energy functions in LogMFD, while they are from the same energy function in metadynamics and AFED (TAMD).

$F(\mathbf{X})$ at t should be known to calculate the force on X_i [see Eq. (1)], while $F(\mathbf{X})$ itself is the property we are to obtain. This seemingly critical problem can be obviated by invoking a conserved quantity, $\hat{H}_{\mathrm{MFD}}$,

$$\begin{aligned}\hat{H}_{\mathrm{MFD}} &= \sum_i^N \frac{1}{2}M_i\dot{X}_i^2+\gamma\log[\alpha F(\mathbf{X})+1]\\ &\quad+\frac{1}{2}Q\dot{\eta}^2+Nk_BT_X\eta,\end{aligned} \tag{5}$$

which necessarily exists when a Nosé-Hoover type thermostat is employed. [$\hat{H}_{\mathrm{MFD}}$ in Eq. (5) applies to the case where a single Nosé-Hoover thermostat is used. $\hat{H}_{\mathrm{MFD}}$ should contain additional terms when multiple thermostats are used. See Ref. [14] for details.] $F(\mathbf{X})$ at t is then obtained from Eq. (5) as

$$\begin{aligned}F(\mathbf{X}) &= \frac{1}{\alpha}\left\{\exp\left[\frac{1}{\gamma}\left(\hat{H}_{\mathrm{MFD}}-\sum_i\frac{1}{2}M_i\dot{X}_i^2\right.\right.\right.\\ &\quad\left.\left.\left.-\frac{1}{2}Q\dot{\eta}^2-Nk_BT_X\eta\right)\right]-1\right\}.\end{aligned} \tag{6}$$

We do not need to set c, but instead $\hat{H}_{\mathrm{MFD}}$ needs to be set at the beginning of a MFD run to ensure $F\ (=F^{'}+c)\geqslant 0$. This means that $\hat{H}_{\mathrm{MFD}}$ should be chosen to satisfy the following,

$$\left(\hat{H}_{\mathrm{MFD}}-\sum_i\frac{1}{2}M_i\dot{X}_i^2-\frac{1}{2}Q\dot{\eta}^2-Nk_BT_X\eta\right)\equiv F_{\log}\geqslant 0. \tag{7}$$

Since the kinetic energy of X_i and the terms associated with the thermostat variables are easily estimated for a preset temperature T_X, one can roughly estimate an appropriate value of $\hat{H}_{\mathrm{MFD}}$. In order to take full advantage of the logarithmic form, it is desirable that $F_{\log}/\gamma$ is close to 0 at $\mathbf{X}=\mathbf{X}^{'}$ where $F(\mathbf{X}^{'})$ is the bottom of the free-energy landscape, i.e., $F(\mathbf{X}^{'})\sim 0$. Unfortunately, we do not know in general the exact position of $\mathbf{X}^{'}$ before running a MFD calculation. However the LogMFD results are not sensitive to the choice of $\hat{H}_{\mathrm{MFD}}$; thus only a rough estimate of $\hat{H}_{\mathrm{MFD}}$ is sufficient.

Equation (6) clearly shows that a free-energy profile is constructed on-the-fly (in principle) as $\mathbf{X}$ evolves in a LogMFD run. If MF is accurately calculated using Eq. (3), a reasonable estimate of $F(\mathbf{X})$ is obtained from Eq. (6). It may not be valid, however, to regard $\hat{H}_{\mathrm{MFD}}$ as a constant [Eq. (6) may lose its validity], unless MF is well estimated. We expect that in practice, an average of more than one $F(\mathbf{X})$ trajectories will give a reasonable estimate of $F(\mathbf{X})$.

III. RESULTS AND DISCUSSION

We now demonstrate how the logarithmic form enhances energy barrier crossing. To this end, we perform a MD simulation for a one-dimensional one-particle system using Eqs. (1) and (2) which can be applied to normal MD as well as MFD. X_i and F are in this case replaced with q and Φ, respectively, where q is the particle position and Φ is the double-well potential, $\Phi(q)=3(q^2-1)^2$. Figure 1(a) shows how the energy barrier is reduced by the logarithmic form. It is clearly seen that as α increases keeping $\gamma=1/\alpha$, the barrier height is substantially lowered. The time evolution of q [Fig. 1(b)] shows that the logarithmic MD trajectory samples a wider range of q and realizes barrier crossing more frequently than the standard canonical MD trajectory. The notable point is that correct canonical distributions can be easily recovered by reweighting [15]. As shown in Fig. 1(c), the position distribution $P(q)$ recovered from the logarithmic MD result using a reweighting technique exhibits two peaks with almost the same height, while the standard canonical MD fails to reproduce the double peaks, which again illustrates the advantage of the logarithmic form.

Next, we show how to construct the free-energy profile using LogMFD. We consider the free-energy profile of a glycine dipeptide molecule in vacuum at 300 K along the dihedral angle ψ, $F(\psi)$, constraining another dihedral angle ϕ at -80 degrees (see Fig. 2). ψ is the collective variable but is dependent on $\mathbf{q}$. We thus introduce a restraining potential and an auxiliary variable $\tilde{X}$ that is treated as the dynamical variable as in metadynamics or TAMD. Specifically, two harmonic potentials were used to keep the dihedral angles at desired values; one is $1/2k[\psi(\mathbf{q})-\tilde{X}]^2$ and another is $1/2k[\phi(\mathbf{q})-\phi_0]^2$, where k is 100 kcal/(mol rad^2) and ϕ_0 is

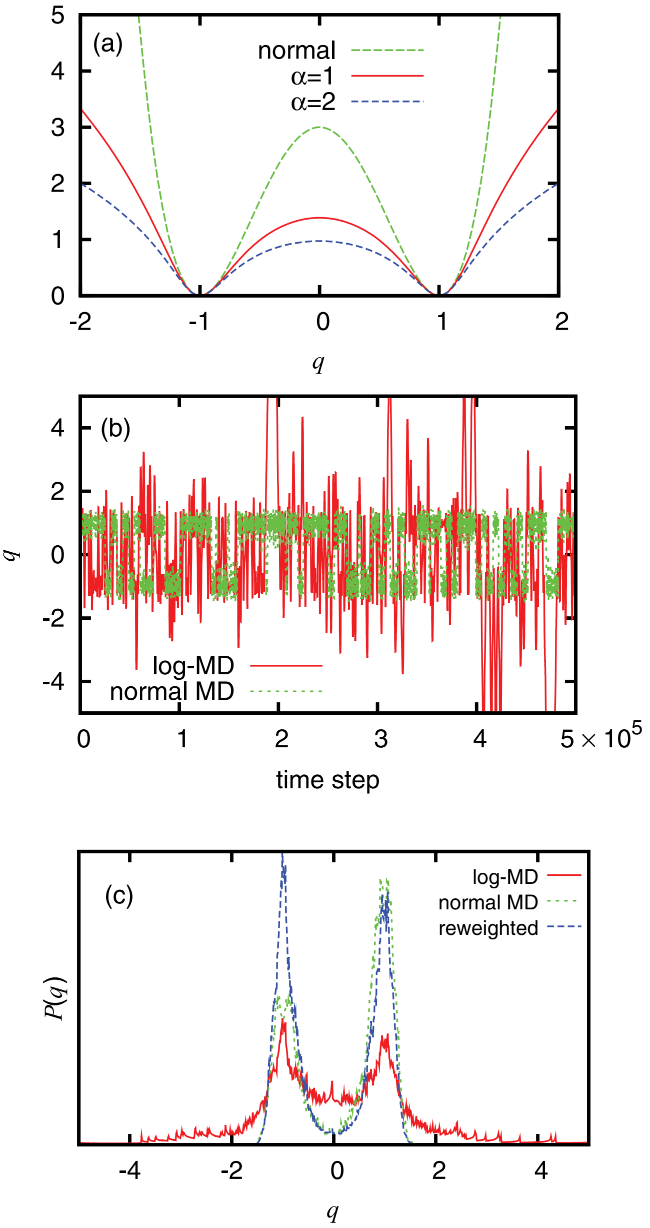

FIG. 1. (Color online) (a) Overall shape of $\Phi = 3(q^2 - 1)^2$ (normal) and its logarithmic form, $\gamma \log(\alpha\Phi + 1)$ ($\alpha = 1$ and 2 with $\gamma = 1/\alpha$). (b) Time evolution of q in the logarithmic MD (with $\alpha = 1.3$ and $\gamma = 0.5$) and the standard MD simulations. The mass and the temperature are set to 1 and the time step is 0.0025. Two Nosé-Hoover thermostats are recursively attached to the particle, which ensures better sampling of the canonical distribution than the linearly chained Nosé-Hoover thermostats [14]. (c) Position distribution from the logarithmic MD and standard MD trajectories. Its canonically reweighted distribution is also shown.

-1.3962634 in radians (-80 degrees). To efficiently sample the whole range of ψ, we used the following parameter set: $\alpha = 3$, $\gamma = 1/\alpha$, $\hat{H}_{\mathrm{MFD}} = 1$, and $M = 10^6$. Although T_X need not be the same as the temperature for $\mathbf{q}$, both the temperatures were maintained at 300 K by attaching the recursive Nosé-Hoover (RNH) thermostat [14] to each particle and $\tilde{X}$, with each containing two thermostats [16]. The glycine dipeptide molecule was described by the CHARMM 22 force field [17] and the EOM for $\mathbf{q}$ were integrated with a time step of 0.5 fs. Using the r-RESPA algorithm, MF on $\tilde{X}$ was updated every 20 MD steps; i.e., MF on $\tilde{X}$ was obtained by averaging $k(\psi - \tilde{X})$ over 20 MD steps (we may perform MFD runs without r-RESPA, in which M and the time step for $\tilde{X}$ can be chosen independently of those for $\mathbf{q}$). The MFD run consisted of 20 000 MFD steps where each MFD step was updated by a 20 MD step calculation; thus 4×10^5 MD steps in total were calculated.

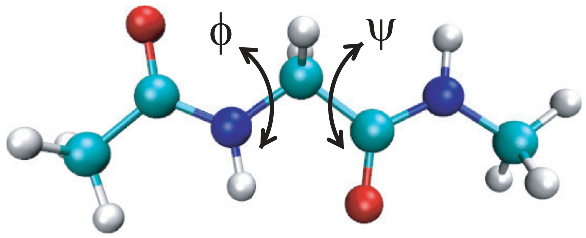

FIG. 2. (Color online) Structure of a glycine dipeptide molecule in vacuum.

We also performed thermodynamic integration (TI) [2] to benchmark LogMFD against TI, which demonstrates the efficiency and accuracy of the LogMFD method. The TI calculation was performed at 100 evenly spaced points in the ψ range of $[-\pi : \pi]$. At each point the MF (slope of F) was evaluated by averaging $k(\psi - \tilde{X})$ over 30 000 MD steps using the RNH thermostat. This number of MD steps is found to be necessary to obtain a fully converged MF. (Note that the convergence rate is found to be worse when the Nosé-Hoover chain thermostat [18] is used instead of the RNH thermostat.) As in the LogMFD calculation, harmonic potentials with the same force constant k were used to fix ϕ to -1.3962634 and ψ to each grid point.

Figure 3 shows the free-energy profile along ψ while keeping $\phi = -1.3962634$. The solid (red) line indicates the result by LogMFD, while filled (green) circles denote the result by TI. It is remarkable that both the results are almost coincident with each other. $\tilde{X}$ in this LogMFD run gradually moved from $\tilde{X} = 1$ toward higher $\tilde{X}$ and overcame the hill around $\psi \sim -2$ after passing the periodic boundary, and then

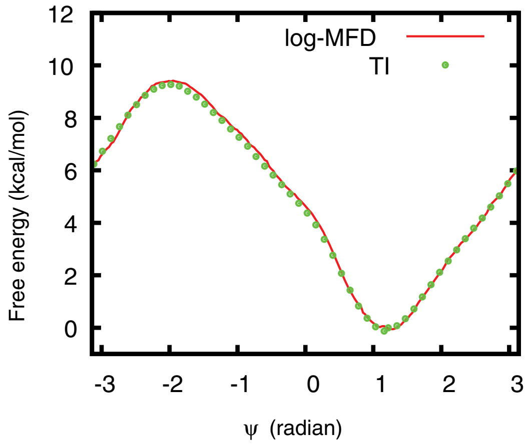

FIG. 3. (Color online) Free-energy profile of a glycine dipeptide molecule in vacuum as a function of the dihedral angle ψ while constraining the dihedral angle ϕ at -1.3962634 (-80 degrees).

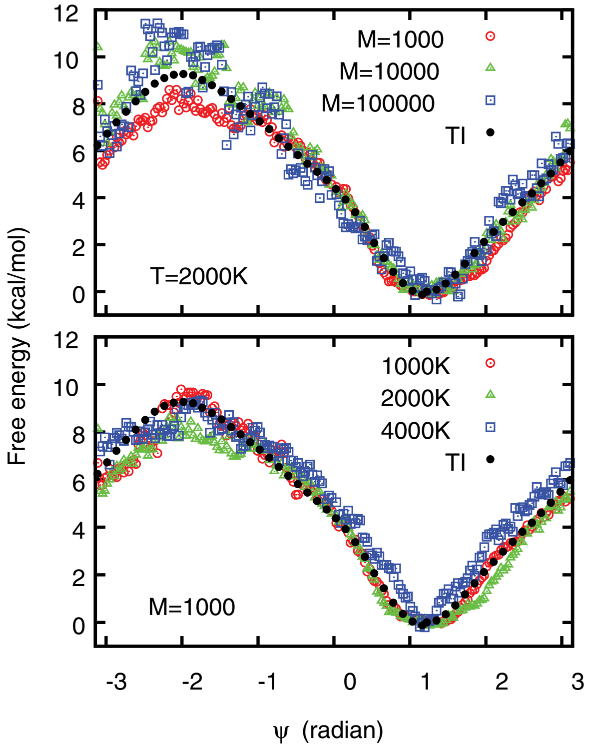

FIG. 4. (Color online) Same as Fig. 3 except that the profiles are obtained by AFED using various parameter sets.

came back to the initial position. Since $\hat{H}_{\mathrm{MFD}}$ was nicely conserved in this run, a single sweep over the whole ψ range was sufficient to construct $F(\psi)$ with high accuracy. A longer MFD trajectory in which $\tilde{X}$ sweeps repeatedly the whole ψ range (including stochastic changes of direction due to the perturbation by the thermostat) was found not to be necessary in the present case. $F(\psi)$ shows the minimum at $\psi \sim 1$ (57.3°) and the maximum at $\psi \sim -2$ (−114.6°). The free-energy barrier is thus about 9.3 kcal/mol, which is much larger than kT of 300 K indicating the effectiveness of the logarithmic form. What should be stressed here is that only ∼13% of the computation cost for the TI calculation was needed for the LogMFD calculation to construct $F(\psi)$ maintaining the same accuracy. This clearly shows the high efficiency of the LogMFD method for free-energy calculation.

In addition to TI, AFED was also performed for the same system and $F(\psi)$ thus obtained is shown in Fig. 4. Similar to the LogMFD calculation, the MF was calculated by averaging $k(\psi - \tilde{X})$ over 20 MD steps. We tried several parameter sets for the temperature and mass of $\tilde{X}$ which are tuned to enhance barrier crossing [4]. Although long MFD runs (6×10^6 MD steps in total) were carried out, none of the parameter sets used in the AFED calculation succeeds in reproducing $F(\psi)$ with the same accuracy as in LogMFD or TI. It is seen that $F(\psi)$ around the maximum ($\psi \sim -2$) is particularly constructed with less accuracy. This indicates that correct sampling of that region is quite difficult only by increasing the T_X (this is not due to insufficient MFD steps as the profile obtained using 4×10^6 MD steps is almost the same as shown in Fig. 4). It is thus strongly indicated that techniques to effectively reduce energy barriers such as a logarithmic form are highly desirable in free-energy calculation based on MFD.

We have attempted several parameter sets for α, M, and T_X in the LogMFD calculations and found that the LogMFD result is not sensitive to the parameter sets as long as the displacement of $\tilde{X}$ at each MFD step is small enough to ensure the adiabatic separation. It should be remarked that while very large $\alpha\,(= 1/\gamma)$ would be of use to enhance crossing energy barriers, it demands extremely precise calculations to reduce numerical errors originated from the logarithmic (exponential) form. In the present system, $\alpha >\sim 5$ is found not to be necessary.

IV. CONCLUSION

We have proposed the logarithmic mean-force dynamics for free-energy calculation. The method enables us to easily sample rare events and to calculate free-energy profiles locally, showing that free-energy profiles can be obtained on-the-fly without any postprocessing. The LogMFD method thus allows us to perform remarkably efficient free-energy calculations, which has been demonstrated by applying the method to a glycine dipeptide molecule making a comparison with conventional TI and AFED methods. It is also worth noting that although TI may not be effective in constructing more than one-dimensional free-energy profiles, LogMFD can easily handle multidimensional free-energy landscapes.

We have examined the effect of the parameter sets used in LogMFD and found that the results are not sensitive to the choice of the parameter sets. We will discuss more details about the protocol for choosing the parameters and multidimensional free-energy construction in LogMFD elsewhere.

ACKNOWLEDGMENTS

T.M. thanks Peter Daivis, Billy Todd, and Ian Snook for helpful discussion and careful reading of the manuscript. This work is partly supported by the Next Generation Super Computing Project, the Nanoscience Program, Japan.

[1] D. J. Wales, *Energy Landscapes* (Cambridge University Press, Cambridge, 2003).
[2] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic Press, San Diego, 2002).
[3] G. M. Torrie and J. P. Valleau, J. Comput. Phys. **23**, 187 (1977).
[4] L. Rosso, P. Mináry, Z. Zhu, and M. E. Tuckerman, J. Chem. Phys. **116**, 4389 (2002).
[5] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. USA **99**, 12562 (2002); A. Laio and F. L. Gervasio, Rep. Prog. Phys. **71**, 126601 (2008).
[6] J. B. Abrams and M. E. Tuckerman, J. Phys. Chem. B **112**, 15742 (2008).
[7] L. Maragliano and E. Vanden-Eijnden, Chem. Phys. Lett. **426**, 168 (2006).
[8] J. B. Abrams, L. Rosso, and M. E. Tuckerman, J. Chem. Phys. **125**, 074115 (2006).

[9] C. F. Abrams and E. Vanden-Eijnden, Proc. Natl. Acad. Sci. USA **107**, 4961 (2010).

[10] S. Nosé, Prog. Theor. Phys. Suppl. **103**, 1 (1991).

[11] For example, supercell parameters can be chosen as **X** that is independent of **q**. See Ref. [5].

[12] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).

[13] M. E. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[14] T. Morishita, Mol. Phys. **108**, 1337 (2010).

[15] A. M. Ferrenberg and R. H. Swendsen, Phys. Rev. Lett. **61**, 2635 (1988); **63**, 1658 (1989).

[16] It is worth noting that any type of thermostat can be used as long as $\hat{H}_{\mathrm{MFD}}$ or its analog is conserved during a MFD run. Even a microcanonical MFD run (no thermostat, $\hat{H}_{\mathrm{MFD}} = \sum M_i \dot{X}_i^2/2 + \gamma \log[\alpha F(\mathbf{X}) + 1]$) should work as long as the MF is accurately estimated.

[17] A. 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).

[18] G. J. Martyna, M. L. Klein, and M. E. Tuckerman, J. Chem. Phys. **97**, 2635 (1992).