JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Efficient Free Energy Calculation of Biomolecules from Diffusion-Biased Molecular Dynamics

Sadanand Singh, Chi-Cheng Chiu,[†] and Juan J. de Pablo*,[†]

Department of Chemical and Biological Engineering, University of Wisconsin—Madison, Madison, Wisconsin 53706, United States

Supporting Information

ABSTRACT: Recently proposed metadynamics techniques offer an effective means for improving sampling in simulations of complex systems, including polymers and biological macromolecules. One of the drawbacks of such methods has been the absence of well-defined or effective convergence criteria. A solution to this problem is considered here in which an optimal ensemble is introduced to minimize the travel time across the entire order parameter range of interest. The usefulness of the proposed approach is illustrated in the context of two systems consisting of biological molecules dissolved in water. The results presented in this work indicate that the proposed method is considerably faster than other existing algorithms for the study of these systems, and that the corresponding free energy that emerges from the simulations converges to the exact result.

INTRODUCTION

Over the past few decades, much research in the area of molecular simulations and statistical physics has been devoted to the calculation of free energies for systems that exhibit rough free energy landscapes. An important concern in this regard has been that of ensuring significant sampling of relevant regions of phase space. In the particular case of "density-of-states" (DOS) based simulation techniques, on-the-fly estimates of the free energy have been used to bias an otherwise random walk, thereby achieving efficient sampling while providing asymptotic estimates of the density of states. Examples of such a strategy include the Wang−Landau method,[1] expanded-ensemble density-of-states (EXEDOS) methods,[2] adaptive-bias method,s[3] and metadynamics.[4]

Two aspects of Monte Carlo (MC) based DOS methods that have generated considerable debate are convergence and accuracy.[5−8] A number of recent studies have served to clarify the general convergence characteristics of Metropolis based Monte Carlo DOS algorithms.[1] However, while such algorithms offer promise for study of complex systems, they are generally system-specific, thereby limiting their widespread use. In contrast, molecular dynamics based methods can be readily incorporated into existing general-purpose approaches, thereby facilitating their adoption for the study of macromolecular systems. To that end, "metadynamics" methods[4] have been proposed in which one or more relevant order parameters are defined to describe a free energy surface. The evolution of the system along those parameters is then controlled through a history-dependent bias potential (V_G), which is administered in the form of "Gaussians" or "hills" (of height W and width δs) that are used to fill-in local minima of the underlying free energy surface at a rate τ_G. The bias potential is gradually constructed in a manner that achieves uniform sampling along the order parameters, thereby leading to an estimate of the corresponding free energy.

The critical issue with metadynamics[4] considered here is that of convergence.[9] In a conventional metadynamics simulation, the simulated free energy fluctuates around the correct value, leading to an average error that scales with the square root of the bias-potential deposition rate.[9] Several ad-hoc solutions to this problem have been proposed, most notably, averaging,[9,10] well-tempered metadynamics,[11] and Wang−Landau metadynamics.[12] In the first approach, averaging,[9,10] after a predefined amount of computation time (referred to as the "equilibration time"), the free energy is averaged throughout the rest of the simulation run. This simple approach guarantees convergence but suffers from a slow convergence rate. In well-tempered metadynamics,[11] the height of the bias potential hills is decreased on the fly, at a predefined rate ΔT. While such an approach guarantees mathematical convergence, the efficiency of the simulation depends significantly on parameters ΔT, W, δs, and τ_G, which are not known a priori. In Wang−Landau metadynamics,,[1,12] the simulation is started with an arbitrarily large hill height, which is then reduced in cycles based on the flatness of a histogram of visits to distinct values of the order parameters. This method suffers from the convergence problems that are inherent to Wang−Landau sampling.[12]

The traditional metadynamics method can be cast in the language of a Wang−Landau density-of-states approach, where one seeks to sample all states along a specified order parameter in a uniform manner.[4] More precisely, a histogram of visits to all states along the order parameter is maintained, and the simulation process is designed to flatten that histogram. As discussed by in refs 6 and 7, such an approach does not necessarily provide an optimal solution to the problem of sampling rugged energy landscapes. Consider a system that exhibits a free energy profile with two stable states 'A' and 'B' separated by a barrier. Generic free energy and diffusivity profiles for such a system are given in Figure 1. Systems undergoing a phase transition, for example, exhibit free energy and diffusivity profiles with the same general features as those shown in Figure 1. It is difficult for the system to cross the energy barrier; the diffusivity of a random walker along such a

Received: May 11, 2012

dx.doi.org/10.1021/ct3003755 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

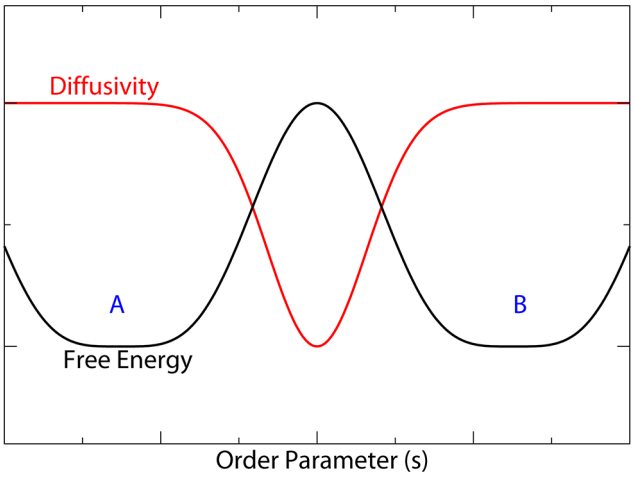

Figure 1. Schematic representation of a generic free energy curve and the corresponding diffusivity profile of a random walker on that curve. If a Wang–Landau approach is followed, the sampling of the barrier would be slow, leading to slow convergence. If the diffusivity profile is taken into account during sampling, by maximizing the number of round trips between states 'A' and 'B', convergence is faster.

free energy profile is therefore lower at the barrier. Following the Wang–Landau approach of uniform sampling of all states would slow sampling of the region of the free energy profile where the diffusivity of the system is small, thereby leading to slow convergence. If, on other hand, one tried to maximize the number of round trips from one extremum of order parameter space to the other, the problem of heterogeneous sampling of low diffusivity states could be avoided.[7]

In a recent publication, we proposed a method that ensures the optimal convergence of metadynamics simulations in general.[13] We refer to such a method as "Flux-Tempered Metadynamics (FTM)." The key idea is to measure a local diffusivity along the order parameter(s) of interest, thereby identifying bottlenecks in the simulations, and then using this information to systematically shift the bias toward these bottlenecks in a feedback loop. In what follows, we briefly revisit some details of the FTM method, and we discuss the improvements in accuracy and convergence that it provides over existing variations of metadynamics. More specifically, in our previous work, we discussed the concept of using local diffusivity information to improve sampling of a simple one-dimensional toy model.[13] In this work, the FTM method is presented for the more relevant case of multiple dimensions, and its general applicability and usefulness are demonstrated in the context of systems consisting of realistic biological macromolecules dissolved in water. It is found that the advantages of FTM are significantly greater in multiple dimensions and for realistic systems than could have been anticipated on the basis of results for the more simplistic model considered earlier.[13]

FLUX TEMPERED METADYNAMICS

In this section, we briefly describe the original FTM method. Note that our discussion of the algorithm is limited to key aspects of the method that we later rely on for its extension and application to multiple dimensions. For a more detailed account of FTM, readers are referred to an earlier publication.[13] The system under consideration is described in terms of a set of atomic coordinates $\mathbf{x}$ and a potential energy function $V(\mathbf{x})$. For a predefined order parameter space $s(\mathbf{x})$, the system dynamics is guided by the free energy profile $F(s)$. Using metadynamics,[4] one can calculate the free energy profile $F(s)$ by adding bias potentials $V_G(s,t)$ in the form of Gaussian functions, at time intervals governed by a rate τ_G. The net force acting on the system is then given by

$$\mathbf{f}_{\text{total}}(t) = -\frac{\partial V(\mathbf{x})}{\partial \mathbf{x}} - \frac{\partial V_G(s(\mathbf{x}), t)}{\partial s}\frac{\partial s}{\partial \mathbf{x}}$$
$$\text{where } V_G(s, t) = W \sum_{t' \le t} \exp\left(-\frac{(s - s(t'))^2}{2\delta s^2}\right) \quad (1)$$

In the limit of long times, $V_G(s,t \to \infty) = -F(s)$. In a Wang–Landau like approach, uniform sampling is sought along the complete range of order parameter space s, which is bounded by $s_- \le s \le s_+$. In the FTM approach, we seek an optimal bias potential V_G that maximizes the number of round trips between s_- and s_+. The details of the method have been presented in a previous publication.[13]

In FTM, the position of the system along the order parameter is recorded by a "walker". To determine the optimal bias potential, two types of walkers are defined: an unlabeled walker and a labeled walker. The unlabeled walker is used to measure the distribution $(n(s))$ of visits to states along the order parameter s. To measure the number of round trips, walkers are assigned a label '+' or '−'. The nature of the label depends on the boundary (s_+ or s_-) that the walker visited last. The two extrema of the order parameter domain act as "reflecting" and "absorbing" boundaries for the labeled walkers: if the label is a '+', a visit to s_+ does not change the label, and that case corresponds to a "reflecting" boundary. In contrast, a visit to s_- changes the label of the walker, and the '+' walker is "absorbed" at the s_- boundary. By construction, the steady-state distribution of labeled walkers satisfies the constraint $n_-(s) + n_+(s) = n(s)$. The quantity $f(s) = (n_+(s))/(n(s))$ corresponds to the fraction of walkers at s having a '+' label. The reflecting and absorbing boundary conditions described above dictate that $f(s_-) = 0$ and $f(s_+) = 1$, respectively.

Using the above definition, one can show that with an optimal V_G that maximizes flux (details in ref 13), the distribution of the system $n(s)$ is given by

$$n^{(\text{optimal})} = \frac{1}{\sqrt{D(s)\lambda}} \quad (2)$$

where λ is a proportionality constant. To obtain such an optimal distribution of the system in the order parameter space s, an iterative metadynamics scheme is defined as

$$V_G(s)^{(\text{new})} = V_G(s)^{(\text{old})} + m(s)$$
$$\text{where } m(s) = -\frac{1}{2}k_B T\left(\ln\left|\frac{df}{ds}\right| - \ln n(s)^{(\text{old})}\right) \quad (3)$$

Such an iterative scheme can be viewed as a "fixed-point" iteration of the form

$$x_{N+1} = h(x_N),\ N = 0, 1, 2, \ldots$$
$$\text{where } h(x) = \sqrt{\frac{1}{j\lambda}x\left|\frac{df}{ds}\right|} \quad (4)$$

which gives rise to a sequence $x_0, x_1, x_2, \ldots$ that converges to a point x. If h is continuous, x is a fixed point of f, i.e., a solution

dx.doi.org/10.1021/ct3003755 | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

of $h(x) = x$. Furthermore, if h is Lipschitz continuous with Lipschitz constant $L < 1$, the fixed point x is unique. Without showing a mathematical derivation of h being Lipschitz continuous with Lipschitz constant $L < 1$, we have verified that irrespective of the initial condition, the iterative scheme indeed converges to a unique value. The unique fixed point is the optimality criterion given by eq 2.

To implement FTM, one can first generate an initial estimate of $V_G^{(old)}$ through any of the existing versions of metadynamics. The first derivatives of $V_G(s)$ with respect to particle coordinates are tabulated. An MD simulation is then run with an additional external force extracted from that table. During this MD run, the histograms $n(s)$ and $n_+(s)$ are collected to calculate f and $(df)/(ds)$. At the end of the MD run, the applied external potential is modified according to eq 3. After each cycle, to ensure convergence, the number of MD steps in the next cycle is doubled, and the relative change in the bias potential is monitored until a predefined accuracy is achieved in the free energy profile. The final free energy is given by

$$F(s) = -V_G(s)^{(\text{optimal})} - k_B T \ln n(s)^{(\text{optimal})} \tag{5}$$

Note that the final free energy or the rate of convergence of the simulation does not depend on the parameters that arise in traditional metadynamics. For a reasonable initial guess of the bias potential $V_G^{(old)}$, the method converges typically in 2 to 3 cycles, which should be contrasted with the large number of iterations required in the Wang–Landau method.[1] Also note that the convergence rate of the FTM method depends on the choice of the initial guess of the bias potential (see Figure S1 of Supporting Information). The iterative FTM scheme described by eq 3 converges to a unique value irrespective of the initial conditions. For complex systems, however, different initial values of $n(s)$ may lead to different convergence rates. For example, if one starts with zero weights for all states along the order parameter, a long simulation time would be required to generate the converged bias (see Supporting Information Figure S1). Nevertheless, as the primary goal of the FTM method is to ensure convergence and high accuracy, it is generally beneficial to start with a good initial bias, typically from a simpler version of metadynamics.

FTM IN MULTIPLE DIMENSIONS

In this section, the FTM method is extended to multiple dimensions. The objective is to arrive at an unbiased random walk in the order parameter space of interest; if multiple order parameters are required to describe the system, the flux (a representation of the random walk) can be decomposed into multiple components. For concreteness, in the following we describe the FTM figure for two order parameters.

In traditional metadynamics with two order parameters s_1 and s_2, the net force acting on a system described by a set of atomic coordinates $\mathbf{x}$ and a potential $V(\mathbf{x})$ is given by

$$\mathbf{f}_{\text{total}}(t) = -\frac{\partial V(\mathbf{x})}{\partial \mathbf{x}} - \sum_{\alpha=1}^{2} \frac{\partial V_G(s_1(\mathbf{x}), s_2(\mathbf{x}), t)}{\partial s_\alpha} \frac{\partial s_\alpha}{\partial \mathbf{x}}$$

$$\text{where } V_G(s_1, s_2, t) = W \sum_{t' \le t} \exp\left(-\frac{(s_\alpha - s_\alpha(t'))^2}{2\delta s_\alpha^2}\right) \tag{6}$$

In the long time limit $V_G(s_1,s_2,t \to \infty) = -F(s_1,s_2)$. Different labeled and unlabeled walkers are defined. Two unlabeled walkers with distributions $n_1(s_1)$ and $n_2(s_2)$ are defined for each order parameter, s_1 and s_2, in addition to a two-dimensional histogram $n(s_1,s_2)$ that measures the frequency of visits to different states along the order parameters. Note that histograms $n_1(s_1)$ and $n_2(s_2)$ can also be directly computed from the two-dimensional histogram $n(s_1,s_2)$

$$n_1(s_1) = \int_{s_2^-}^{s_2^+} n(s_1, s_2)\, ds_2$$

$$n_2(s_2) = \int_{s_1^-}^{s_1^+} n(s_1, s_2)\, ds_1 \tag{7}$$

Labeled walkers are defined in each of the order parameters. Two quantities, $f_1 = (n_1^+)/(n_1)$ and $f_2 = (n_2^+)/(n_2)$, measure the fraction of walkers with a label '+' in the s_1 and s_2 order parameters, respectively. Following a similar definition for current/flux j, an iterative algorithm for the calculation of the α (= 1 or 2) component of the biasing force can be designed as follows:

$$-\frac{\partial V_G^{\text{new}}}{\partial s_\alpha} = -\frac{\partial V_G^{\text{old}}}{\partial s_\alpha} - \frac{dm_\alpha(s_\alpha)}{ds_\alpha}$$

$$\text{where } m_\alpha(s_\alpha) = -\frac{1}{2}\left[\ln\left|\frac{df_\alpha}{ds_\alpha}\right| - \ln n_\alpha(s_\alpha)^{(\text{old})}\right] \tag{8}$$

Note that these equations are scalar in nature, i.e., there is one equation for each direction α. Once converged, the final free energy profile is given by

$$F(s_1, s_2) = -V_G(s_1, s_2)^{(\text{optimal})} - k_B T \ln n(s_1, s_2)^{(\text{optimal})} \tag{9}$$

Following a similar procedure, the proposed FTM approach can be generalized to any number of order parameters. A detailed algorithm of a two-dimensional FTM is provided in the Supporting Information.

SIMULATION DETAILS

System 1: Polyalanine. For the 15-residue polyalanine ($[(ALA)_{15}\text{-}NH_2]$ protein simulations, we used the AMBER99 force field and TIP3P water model.[14] The protein was simulated at a temperature of 298 K and a pressure of 1 bar using a Nosé–Hoover thermostat[15] and a Parinello–Rahman barostat,[16] with a time step of 0.002 ps. Electrostatic interactions were handled using a particle mesh Ewald sum.[17] A cutoff of 1.2 nm was used for nonbonded interactions. The free energy of the system was calculated as a function of the end-to-end distance of the peptide (which is the order parameter in our calculations). The reference free energy profile was generated from umbrella sampling[18] simulations. For umbrella sampling, a total of 40 simulation windows were used to span the end-to-end distance range between 1.8 and 3.4 nm. Each of the windows was then equilibrated for 10 ns. Statistics were collected over the following 50 ns.

For all systems considered here, the following four variations of metadynamics were implemented. Conventional metadynamics[4] was applied with a hill height of $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps. In another variation, hills with $W = 0.1$ kJ/mol were added every $\tau_G = 1$ ps. The final free energy profile was calculated by averaging over the entire simulation time.[10] In the second method, a well-tempered metadynamics (WTM)[11] approach was applied with a hill height of $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps. Again two values of the decay

dx.doi.org/10.1021/ct3003755 | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

parameter ΔT were applied, namely $\Delta T = 60$ and $\Delta T = 5$. An optimal value of the decay parameter ($\Delta T = 45$) was arrived at by comparing the error in the free energy surfaces after simulations of 50 ns for different values of ΔT. Third, Wang–Landau like metadynamics (or Flat Histogram Metadynamics, FHM), as proposed by Min et al.,[12] was implemented with an initial hill height $W = 2$ kJ/mol added every $\tau_G = 1$ ps. Finally, the FTM method was applied from an initial bias obtained from a WTM run with hills $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps for 2 ns, as well as from an FHM run of 1 ns. The initial MD run consisted of 5 ns. For all methods, a hill width δs of 0.1 was chosen, while the order parameter was divided into a grid with a bin size of 0.02. For all methods (except umbrella sampling), we performed 20 ns runs.

System 2: Alanine Dipeptide. The proposed FTM method was also implemented in two dimensions for an alanine dipeptide. The sought after free energy profile is a function of two torsional angles, namely ϕ and ψ. The simulations were conducted in the OBC implicit solvent with the surface term modeled using GBSA to represent water solvent effects with an effective salt concentration of 0.2 M.[19] The protein was simulated at a temperature of 298 K using a Nosé–Hoover thermostat,[15] with a time step of 0.002 ps. Electrostatic interactions were handled using a cutoff of 1.8 nm, and a cutoff of 1.2 nm was used for nonbonded interactions. The reference free energy profile was generated from umbrella sampling[18] simulations along with a 2-D WHAM calculation.[20] For umbrella sampling, a total of 1296 simulation windows were used to span the ϕ and ψ range between $-\pi$ and π. Each of the windows was then equilibrated for 4 ns. Statistics were collected over the following 20 ns.

Consistent with the one-dimensional implementation mentioned above, four variations of metadynamics were implemented to simulate the alanine dipeptide system. Conventional metadynamics[4] was applied with a hill height of $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps. In another variation, hills with $W = 0.1$ kJ/mol were added every $\tau_G = 1$ ps. The final free energy profile was calculated by averaging over the entire simulation time of 50 ns.[10] In the second method, a well-tempered metadynamics (WTM)[11] approach was applied with a hill height of $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps. An optimal value of the decay parameter ($\Delta T = 45$) was determined by comparing the error in the free energy surfaces after simulations of 100 ns with different values of ΔT. Third, Wang–Landau like metadynamics (or Flat Histogram Metadynamics, FHM), as proposed by Min et al.,[12] was implemented with an initial hill height $W = 2$ kJ/mol added every $\tau_G = 1$ ps. Finally, the FTM method was applied from an initial bias obtained from a WTM run with hills $W = 0.1$ kJ/mol added every $\tau_G = 10$ ps for 20 ns. The initial MD run consisted of 10 ns. For all methods, a hill width δs of $\pi/90$ was chosen, while the order parameter was divided into a 180×180 grid. The results in the Supporting Information for Meta, WTM, and FHM methods correspond to 50 ns runs. The order parameters ϕ and ψ being periodic, the boundaries s_- and s_+ can be chosen arbitrarily such that they are equidistant in every direction. We have used a value of $s_- = -\pi/2$ and $s_+ = +\pi/2$.

The error bars in the free energy profiles for different cases were obtained from the standard deviation of results from 10 independent runs.

RESULTS AND DISCUSSION

To demonstrate the usefulness of FTM, we applied it here to calculate the unfolding free energy of a 15-residue polyalanine molecule in explicit water ($[(ALA)_{15}\text{-}NH_2]$, which under normal conditions is known to exhibit an α-helical structure). The results from 20 ns runs are shown in Figure 2. A reference free energy profile was also determined from an umbrella sampling[18] simulation with statistics generated over 2 μs runs (see Simulation Details). In order to compare FTM with other available variants of metadynamics, four specific approaches were also implemented: the traditional metadynamics ("Meta") with averaging,[10] well-tempered metadynamics[11] ("WTM"), Wang–Landau metadynamics[12] ("FHM"), and FTM.

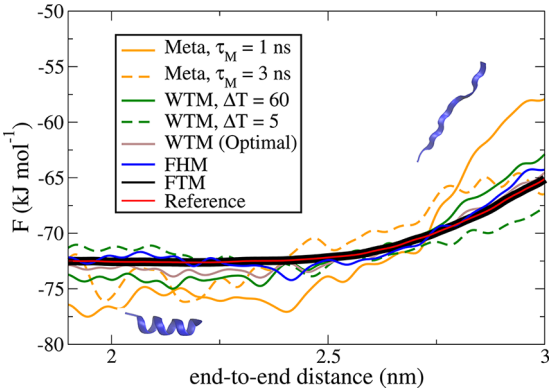

Figure 2. Free energy curve as a function of end-to-end distance for the polyalanine system $[(ALA)_{15}\text{-}NH_2]$ in water. The reference free energy profile generated from umbrella sampling simulations is shown in red. The results of the proposed flux tempered metadynamics (FTM) method are shown in black. The results of conventional metadynamics (Meta) with averaging are shown in orange. Results from well-tempered metadynamics (WTM) are shown in green and brown. Results from flat-histogram metadynamics (FHM) are shown in blue. The error bars for metadynamics and WTM are 0.8 kJ/mol, while those for FHM and FTM are 0.2 kJ/mol. The FTM simulations were performed with two different initial values of the applied bias (see text).

The details of the system and the parameters employed for each of these methods have been provided in the Simulation Details section. Note that the simulation time for FTM includes the time used to calculate the initial bias using conventional metadynamics. The results of the proposed FTM are indistinguishable from the exact or reference free energy profile. The accuracy of the free energy curve obtained from conventional metadynamics (Meta), WTM, or FHM is inferior to that of the proposed FTM profile (see Figure 2). Also note that the results from the WTM method depend on several simulation parameters, including ΔT, W, δs, and τ_G.

The convergence and accuracy of different methods can be compared on a quantitative basis by introducing an error measure according to

$$E(t) = \frac{1}{(s_+ - s_-)} \inf_{C \in \mathbb{R}} \int_{s_-}^{s_+} |F(s, t) - \hat{F}(s) - C(t)|\, ds \tag{10}$$

where $F(s,t)$ is the free energy at time t, and $\hat{F}(s)$ is the reference free energy. As different metadynamics simulations do not yield the same absolute free energy profile, for a fair comparison all calculated profiles need to be shifted to the

dx.doi.org/10.1021/ct3003755 | J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

reference profile. $C(t)$ provides an estimate of the absolute shift in the free energy profile and is calculated using a least-squares fit. Figure 3 shows the error function, $E(t)$, for different

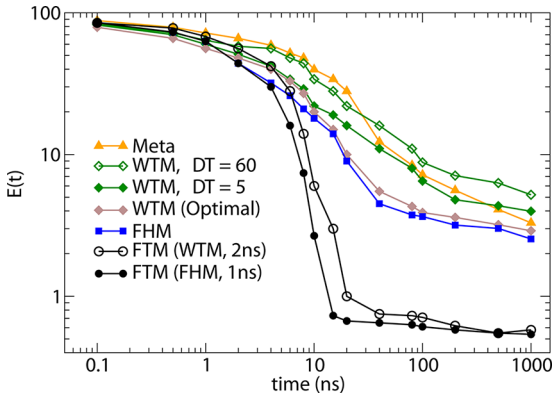

Figure 3. Error in the calculation of free energy (in energy units, kJ/mol), $E(t)$, as defined in eq 10, as a function of simulation time for polyalanine dissolved in water. For each method (see Figure 2 caption for a definition of acronyms), the errors correspond to an average of five independent simulations. The results of the proposed flux tempered metadynamics (FTM) method are shown in black. The FTM simulations were performed with two different initial values of the applied bias (see text). The results of conventional metadynamics (Meta) with averaging are shown in orange. For conventional metadynamics, results were averaged over the entire simulation time. Results from well-tempered metadynamics (WTM) are shown in green and brown. Results from flat-histogram metadynamics (FHM) are shown in blue.

methods for the polyalanine system as a function of simulated time. For conventional metadynamics with averaging, the error depends on the time of averaging and decreases monotonically with simulation time. The dependence of the WTM method on simulation parameters is apparent. With optimal parameters for WTM, the algorithm has similar performance to that of the FHM method. Although FHM converges, FTM converges faster, and furthermore, the error in FTM is significantly smaller $[E(t) \approx 0.6]$ than that of all other methods $[E(t) \geq 4.5]$. On the basis of the extrapolation of results from the WTM/FHM method in Figure 3, we conclude that the proposed FTM method yields results that could only be achievable by WTM/FHM simulations that would have to be much longer. The gains in efficiency of FTM are more pronounced if viewed in terms of actual process/CPU time (see Figure S2 of Supporting Information), as cycles of FTM involve fewer force calculations than other versions of metadynamics considered in this work.

The proposed FTM method can be generalized to multiple dimensions. To examine the performance of FTM in a two-dimensional order parameter space, a similar analysis as that presented above was performed on an alanine dipeptide in implicit water. The free energy profile was determined along two dihedral angles, ϕ and ψ. A description of the method for the case of two-dimensional order parameters and the details of the model are provided in the Simulation Details section. The reference free energy profile was obtained by means of extensive umbrella sampling simulations.

As before, four variants of metadynamics were used to generate the free energy surface of the dipeptide. Only the reference free energy surface and that obtained from the FTM method (after a 50 ns simulation) are shown in Figure 4 (see

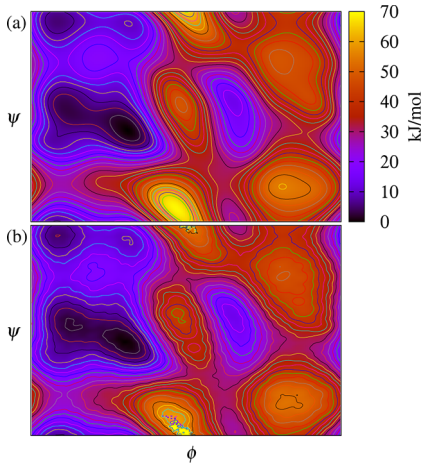

Figure 4. (a) Two-dimensional free energy surface for an alanine dipeptide peptide from an FTM simulation after 50 ns of time. (b) The reference free energy profile was obtained from umbrella sampling simulations, with statistics collected over more than 26 μs of time.

the Supporting Information for additional results). To compare these profiles, we used an error function analogous to that given by eq 10:

$$E(t) = \frac{1}{4\pi^2} \inf_{C \in \mathbb{R}} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} |F - \hat{F} - C| \, ds_1 \, ds_2 \tag{11}$$

where $\hat{F}$ represents the reference free energy profile. Results are shown in Figure 5a for different methods, as a function of

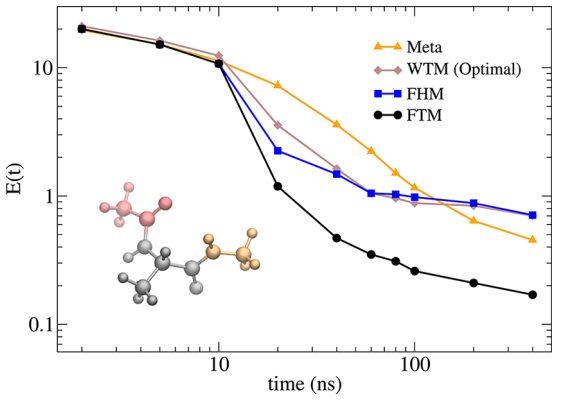

Figure 5. Error in the calculation of the free energy (in energy units, kJ/mol), $E(t)$, as defined in eq 11, as a function of simulation time for the alanine dipeptide system. For each method, the errors presented here correspond to an average of four independent simulations. The traditional metadynamics results (Meta) were averaged over the complete simulation time and are shown in orange. For conventional metadynamics, results were averaged over the entire simulation time. Results from well-tempered metadynamics (WTM) are shown in brown. Results from flat-histogram metadynamics (FHM) are shown in blue.

simulation time. The trends observed for the dipeptide in two dimensions are the same as those found for the polyalanine system in one dimension. As can be seen in the figure, in two dimensions the advantages of the FTM method over all other techniques considered here are equally significant; FTM is significantly faster than other versions of metadynamics. As before, the gains in efficiency of FTM are equally significant if viewed in terms of actual process/CPU time (see Figure S3 of

dx.doi.org/10.1021/ct3003755 | J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

the Supporting Information). Specific results for the free energy surface corresponding to other versions of metadynamics are provided in Figure S4 of the Supporting Information.

In both the unidimensional (polyalanine) and two-dimensional (alanine dipeptide) systems, one stark difference in the convergence of the FTM method and that of other methods considered here is the sudden decrease in the error after the first few cycles. Existing methods exhibit a $1/\sqrt{t}$ convergence, with different prefactors. Their convergence is based directly or indirectly on the uniformity of histograms of visits to different order parameter states, and hence they follow the same convergence scaling with time.[21] In general, the decay of the error is proportional to the square root of the round-trip time[7] (defined as the time taken per round trip across the order parameters space). In the case of the FTM method, which minimizes round-trip time by construction, the scaling of the error with simulation time improves dramatically. Figure 6

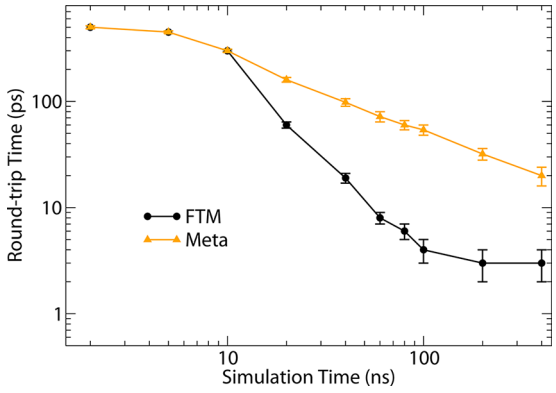

Figure 6. Round-trip time, defined as simulation time required to go from one boundary to another (s_- to s_+) of the ϕ order parameter, as a function of simulated time for the alanine dipeptide system. Results are shown for the FTM and traditional metadynamics methods (Meta).

shows the time taken per round trip across the ϕ order parameter space for the alanine dipeptide for FTM and traditional metadynamics. In contrast with the traditional metadynamics method, for the FTM method, the round-trip time decreases significantly faster, leading to sharper convergence (see Figure 5). An additional feature of the FTM method (as compared to other methods considered here that involve on-the-fly modification of the potential energy by addition of Gaussians) is that it satisfies detailed balance after the calculation of the initial bias potential.

CONCLUSIONS

To summarize, the proposed FTM method relies on metadynamics and density of states calculations in an optimal ensemble that maximizes the number of round trips between the boundaries of the order parameter range of interest. For the chemically detailed systems considered here, namely a large polypeptide dissolved and an alanine dipeptide in water, the method has been shown to be considerably more efficient than existing, widely used state-of-the art sampling algorithms based on metadynamics.

ASSOCIATED CONTENT

Supporting Information

Additional results about the effects of the initial bias on the results, errors in free energy profiles as a function of CPU time, and two-dimensional free energy profiles of the alanine dipeptide system for different methods. This material is available free of charge via the Internet at http://pubs.acs.org/.

AUTHOR INFORMATION

Corresponding Author

*E-mail: depablo@uchicago.edu.

Present Address

†Institute for Molecular Engineering, University of Chicago, Chicago, IL 60637.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work is supported by NIH (1R01DK088184). Additional support for algorithm development for self-assembly from NSF grant DMR-0832760 is also acknowledged.

REFERENCES

(1) Wang, F. G.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(2) Escobedo, F. A.; Martinez-Veracoechea, F. J. *J. Chem. Phys.* **2008**, *129*, 154107.
(3) Dickson, B. M.; Legoll, F.; Lelievre, T.; Stoltz, G.; Fleurat-Lessard, P. *J. Phys. Chem. B* **2010**, *114*, 5823−30.
(4) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.
(5) Morozov, A. N.; Lin, S. H. *Phys. Rev. E* **2007**, *76*, 026701.
(6) Dayal, P.; Trebst, S.; Wessel, S.; Wurtz, D.; Troyer, M.; Sabhapandit, S.; Coppersmith, S. N. *Phys. Rev. Lett.* **2004**, *92*, 097201.
(7) Trebst, S.; Huse, D. A.; Troyer, M. *Phys. Rev. E* **2004**, *70*, 046701.
(8) Rathore, N.; Knotts, T. A.; de Pablo, J. J. *Biophys. J.* **2003**, *85*, 3963−3968.
(9) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
(10) Crespo, Y.; Marinelli, F.; Pietrucci, F.; Laio, A. *Phys. Rev. E* **2010**, *81*, 055701.
(11) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(12) Min, D. H.; Liu, Y. S.; Carbone, I.; Yang, W. *J. Chem. Phys.* **2007**, *126*, 194104.
(13) Singh, S.; Chiu, C.-c.; de Pablo, J. J. *J. Stat. Phys.* **2011**, *145*, 932.
(14) Sorin, E. J.; Pande, V. S. *Biophys. J.* **2005**, *88*, 2472−2493.
(15) Martyna, G. J.; Klein, M. L.; Tuckerman, M. *J. Chem. Phys.* **1992**, *97*, 2635−2643.
(16) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(17) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(18) Kastner, J. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, 1−11.
(19) Onufriev, A.; Bashford, D.; Case, D. A. *Proteins: Struct., Funct., Bioinf.* **2004**, *55*, 383−394.
(20) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(21) Zhou, C. G.; Bhatt, R. N. *Phys. Rev. E* **2005**, *72*, 025701.