JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

# Robust Estimation of Diffusion-Optimized Ensembles for Enhanced Sampling

Pengfei Tian,[†] Sigurdur Æ. Jónsson,[‡] Jesper Ferkinghoff-Borg,[§] Sergei V. Krivov,[∥] Kresten Lindorff-Larsen,[⊥] Anders Irbäck,[‡] and Wouter Boomsma*[,⊥]

[†]Niels Bohr Institute, University of Copenhagen, Blegdamsvej 17, 2100 Copenhagen, Denmark

[‡]Computational Biology and Biological Physics, Department of Astronomy and Theoretical Physics, Lund University, Sölvegatan 14A, SE-223 62 Lund, Sweden

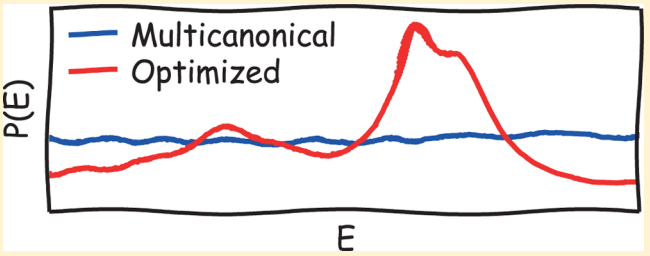
[§]Department of Systems Biology, DTU, Søltofts Plads, 221 DK-2800, Kgs. Lyngby, Denmark

[∥]Astbury Center for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

[⊥]Structural Biology and NMR Laboratory, Department of Biology, University of Copenhagen, Ole Maaløes Vej 5 DK-2200 Copenhagen N, Denmark

Supporting Information

**ABSTRACT:** The multicanonical, or flat-histogram, method is a common technique to improve the sampling efficiency of molecular simulations. The idea is that free-energy barriers in a simulation can be removed by simulating from a distribution where all values of a reaction coordinate are equally likely, and subsequently reweight the obtained statistics to recover the Boltzmann distribution at the temperature of interest. While this method has been successful in practice, the choice of a flat distribution is not necessarily optimal. Recently, it was proposed that additional performance gains could be obtained by taking the position-dependent diffusion coefficient into account, thus placing greater emphasis on regions diffusing slowly. Although some promising examples of applications of this approach exist, the practical usefulness of the method has been hindered by the difficulty in obtaining sufficiently accurate estimates of the diffusion coefficient. Here, we present a simple, yet robust solution to this problem. Compared to current state-of-the-art procedures, the new estimation method requires an order of magnitude fewer data to obtain reliable estimates, thus broadening the potential scope in which this technique can be applied in practice.




## 1. INTRODUCTION

The free-energy landscape sampled in a molecular simulation is typically rugged with a multitude of local minima, and obtaining converged statistics in such simulations remains an important challenge. Over the last decades, a number of enhanced sampling strategies have been proposed to overcome this problem. Early examples include methods such as simulated annealing,[1] replica-exchange (parallel tempering),[2,3] and simulated tempering,[4,5] which in various ways soften the landscape by allowing temporary excursions to higher temperatures. Another important class of methods involve generalized ensembles, where the target Boltzmann distribution is modified to facilitate sampling, after which the Boltzmann statistics can be recovered using reweighting techniques.[6] A typical example is the multicanonical, or flat-histogram, method,[7] where one reduces trapping in local minima in some chosen reaction coordinate by constructing an ensemble in which the distribution of this coordinate is uniform. This idea forms the basis for various adaptive bias methods, such as Wang−Landau[8] and multihistogram methods,[9] and has been extended for use with molecular dynamics simulations under names such as local-elevation[10] and metadynamics.[11]

The main advantage of the multicanonical ensemble, which was used in a protein simulation for the first time already 20 years ago,[12] is that it compensates for the entropic differences of different states. While the method has proven useful in both coarse-grained and atomic-level molecular simulations,[13−15] there are some indications that efficiency could be improved further by taking the roughness of the energy landscape into account. Assuming diffusive dynamics along the reaction coordinate, Trebst et al. proposed a method for generalized ensemble Monte Carlo (MC) simulations that compensates for the fact that the diffusion coefficient varies along the energy coordinate, and demonstrated that this corresponds to minimizing the time it takes the system to diffuse from high to low energies.[16] With the Ising model as a testbed, it was shown that this method, compared to flat-histogram sampling, leads to faster equilibration and smaller statistical errors of the

Received: September 25, 2013
Published: January 14, 2014

estimated density of states. For molecular systems, this idea has since been employed to maximize the temperature flow in replica exchange simulations,[17,18] and to improve the efficiency in simulations based on metadynamics[19,20] and multicanonical molecular dynamics.[21]

While the diffusion-optimized ensemble approach shows great promise, the practical application of it has been hindered by the serious challenge of estimating the diffusion coefficient accurately, especially at early stages of a simulation when only limited statistics are available. Here, we present a solution to this problem. We propose replacing the standard global estimation procedure with a local estimation technique that was previously introduced in the context of trajectory analysis and reaction coordinate optimization.[22−24] To make the technique applicable for use in the context of enhanced sampling, we present simple techniques for determining the relevant sampling interval and for avoiding boundary artifacts. This forms the basis of a novel weight update scheme that leads to faster convergence in the estimation of the diffusion-optimized ensemble. The resulting method produces optimized ensembles more precisely while requiring only a fraction of the data utilized by previous methods.

The paper consists of two parts. First, we give a detailed description of the proposed method and compare its performance to the current state-of-the-art on two test systems. Second, we illustrate the scope and robustness of the method by applying it to two challenging biological simulation problems.

## 2. METHODOLOGY

The methodology in this paper will be presented in the context of MC-based simulations, although it can be applied in other contexts as well.[19−21] For simplicity, we choose the energy $E$ as reaction coordinate, but more general choices are possible (see below).

In a generalized ensemble defined by the weight function $w(E)$, the energy distribution $P(E)$ is given by

$$P(E) = \frac{w(E)g(E)}{Z} \quad Z = \int w(E)g(E)\mathrm{d}E \tag{1}$$

where $g(E)$ is the density of states. The canonical ensemble corresponds to the choice $w(E) = \exp(-E/kT)$, where $k$ is Boltzmann's constant and $T$ is temperature. In the multicanonical approach,[7] all energies are made equally likely by taking the weight function to be $w(E) = 1/g(E)$.

Given any proposed weight function $w(E)$, microstates $S$, distributed according to $P(S) \propto w(E(S))$, can be generated by the Metropolis−Hastings method. In each iteration of such a simulation, the system is perturbed by a random move, taking the system from a microstate $S_1$ with energy $E_1$ to a microstate $S_2$ with energy $E_2$. The new microstate $S_2$ is subject to an accept/reject step, where the acceptance probability is given by

$$P_{\mathrm{acc}}(S_1 \to S_2) = \min[1, w(E_2)/w(E_1)] \tag{2}$$

where we for simplicity assume that the probabilities of suggesting the moves $S_1 \to S_2$ and $S_2 \to S_1$ are the same. The algorithm fulfills detailed balance, thus ensuring proper sampling of the desired distribution $P(S) \propto w(E(S))$. From such simulations for any chosen weight function $w(E)$, canonical averages can be determined by reweighting techniques.[6]

The freedom in the choice of $w(E)$ leaves room to optimize the computational efficiency. The multicanonical form of $w(E)$, which gives a flat energy histogram, is a natural first guess. The mobility in $E$ in a Metropolis−Hastings simulation is, however, generally position-dependent even if $P(E)$ is flat, and it might therefore be advantageous to let the system spend more time in regions of low mobility. Trebst et al. proposed a strategy to achieve this, based on the assumption that the motion along the chosen reaction coordinate $E$ can be modeled as one-dimensional diffusion.[16] With this assumption, the mean first-passage time $\tau$ for round trips across the energy interval of interest, $E_{\min} < E < E_{\max}$, can be estimated as[25]

$$\tau = \int_{E_{\min}}^{E_{\max}} \frac{\mathrm{d}E}{D(E)P(E)} \tag{3}$$

where $D(E)$ is the position-dependent diffusion coefficient. Assuming $D(E)$ to be independent of the weight function, Trebst et al. proposed to optimize this expression with respect to $P(E)$, subject to the constraint that $P(E)$ must be normalized. Using a Lagrange multiplier to account for the constraint, the task becomes to optimize the integral

$$\int_{E_{\min}}^{E_{\max}} \left( \frac{1}{D(E)P(E)} + \lambda P(E) \right) \mathrm{d}E \tag{4}$$

As is easily verified, the solution to this problem is given by

$$P_{\mathrm{opt}}(E) \sim \frac{1}{\sqrt{D(E)}} \tag{5}$$

To obtain this energy distribution, the weight function has to be chosen as

$$w_{\mathrm{opt}}(E) \sim \frac{1}{g(E)\sqrt{D(E)}} \tag{6}$$

Optimizing the rate of round trips thus yields a weight function, $w_{\mathrm{opt}}(E)$, that differs from the multicanonical one by a factor $D(E)^{-1/2}$. This factor implies that the optimized ensemble indeed puts more emphasis on regions of slow diffusion, compared to the multicanonical ensemble. The optimized and multicanonical ensembles coincide in the special case of constant $D(E)$.

In order to implement the optimized-ensemble approach, one has to estimate the product $g(E)D(E)^{1/2}$. This can be accomplished by using, for instance, the Wang−Landau method[8] to estimate $g(E)$, and performing a separate calculation of the position-dependence of the diffusion coefficient. Various techniques have been proposed for estimating this quantity. The standard approach is to consider the mean square displacement, often calculated on a sliding window over the trajectory.[26−32] Trebst et al. proposed an alternative technique, where states in the simulation are labeled with the boundary that was last visited, and the distribution of these labels is analyzed as a function of the reaction coordinate.[16] This approach is currently the standard technique for diffusion-optimized enhanced sampling and will be covered in detail in the next section. A Bayesian estimation procedure was proposed by Hummer, allowing for self-consistent estimation of free energies, diffusion coefficients and their uncertainties, and making it possible to incorporate prior information on for instance the smoothness of the diffusion coefficient.[33] More recently, other Bayesian approaches have made it possible to simultaneously estimate other kinetic

quantities.[34,35] Finally, Krivov developed an estimation strategy based on the analysis of the frequency of the trajectory crossing 'cuts' at different reaction coordinate values.[22−24] For the purpose of enhanced sampling, this method is particularly convenient, as the density of states and the diffusion coefficient are estimated together, potentially reducing error levels compared to the case where the two factors are estimated independently. As we will demonstrate, it also means that the estimation procedure can be seen as a natural generalization of the histogram estimation in the standard multicanonical ensemble.

The degree of variation of the diffusion coefficient will depend strongly on the chosen reaction coordinate, the set of moves (for MC simulations) and nature of the studied system. There is a substantial literature exploring diffusivity in molecular simulation, with reports of variation factors anywhere between unity and an order of magnitude.[31,32,35,36] This sets the scale for the efficiency improvements one can expect to obtain using eq 6.

**2.1. Diffusion Coefficient from Labeled Walkers.** Trebst et al.[16] analyzed the flux along the reaction coordinate by labeling the current simulation state as either an up- or a down-walker (Figure 1). With this representation, it is possible

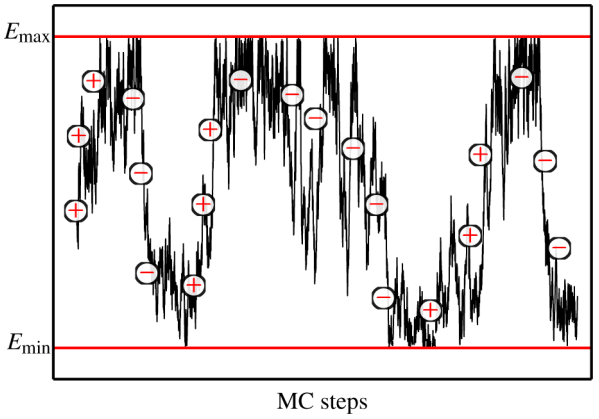



**Figure 1.** Description of simulation trajectories in terms of up- and down-walkers. At any given point in time, the system is labeled as an up-walker (+) or a down-walker (−) depending on which of the two boundaries it visited most recently. Hence, for an up-walker, $E_{min}$ and $E_{max}$ act as reflecting and absorbing boundaries, respectively; the label '+' is left unchanged upon visits to $E_{min}$, but changes to '−' when $E_{max}$ is reached. It is counted as one round trip when the system moves from one boundary to the other and back again. Plot generated using matplotlib.[37]

not only to estimate the rate of round trips but also the diffusion coefficient, $D(E)$, can be inferred from how labeled and unlabeled walkers are distributed. Let $Z_H(E)$ denote the histogram of energy for unlabeled walkers, and let $f(E)$ be the fraction of all walkers with energy $E$ that are currently labeled down. In steady state, the currents of up- and down-walkers are position-independent and balance each other, which in particular implies that[16]

$$D(E) = \frac{1}{\tau}\frac{1}{Z_H(E)\mathrm{d}f/\mathrm{d}E} \tag{7}$$

This relation can be used to formulate a scheme for iterative determination of the optimized weights (eq 6), in which new weights $w'(E)$ are estimated from simulations with trial weights $w(E)$ as

$$\ln w'(E) = \ln w(E) + 1/2[\ln \mathrm{d}f/\mathrm{d}E - \ln Z_H(E)] \tag{8}$$

This update-rule has a similar form as that of the standard multicanonical ensemble, $\ln w'(E) = \ln w(E) - \ln Z_H(E)$.

The labeled walker (LW) technique for estimating $D(E)$ and thereby $w_{opt}(E)$ has been successfully applied to simulations of various systems, including Ising models[16] and atomic-level protein models.[20,21] It is, however, a global procedure that requires the availability of a reasonably high number of round trips, which implies that the simulations are already fairly well converged (Supporting Information (SI) Figure S1). Furthermore, the method turns out to be quite sensitive to the precise location of the boundaries $E_{min}$ and $E_{max}$ (SI Figure S2). For the purpose of enhanced sampling, these features are undesirable. It would be convenient if optimized weights could be estimated at early stages of a simulation, and if the estimation procedure was relatively insensitive to how much of the energy landscape had been explored so far. A step in this direction was taken by Nadler et al.[18] by devising a method that estimates $D(E)$ from mean-first passage times across subintervals of the interval of interest. Here, we take it one step further by proposing an essentially local estimation of $D(E)$.

**2.2. Diffusion Coefficient from a Transition Histogram.** The diffusion coefficient estimation technique by Krivov[22−24] combines the ordinary energy histogram, $Z_H(E)$, with a second histogram denoted by $Z_C(E)$, which counts the number of transitions that cross a 'cut' at $E$ (Figure 2a).

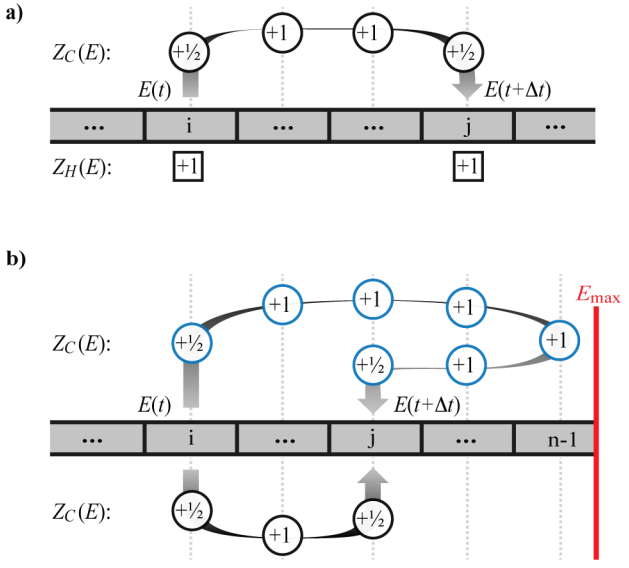



**Figure 2.** Illustration of $Z_C(E)$ update schemes: (a) Standard scheme along the reaction coordinate $E$, (b) extension using a ring topology. The two possible paths from $i$ to $j$ are selected with equal probability. The dotted lines represent the 'cuts' at each bin.

Specifically, from a reaction-coordinate time series $E_i = E(i\Delta t)$ $(i = 1,2,...)$, $Z_C(E)$ is calculated as

$$Z_C(E) = 1/2 \sum_i \theta[(E_i - E)(E - E_{i+1})] \tag{9}$$

where $\theta(x)$ is the Heaviside step function. Assuming diffusive dynamics, which generally requires the sampling interval $\Delta t$ to


dx.doi.org/10.1021/ct400844x | J. Chem. Theory Comput. 2014, 10, 543−553

be neither too large nor too small (see below), one finds that $D(E)$ can be expressed as[22]

$$D(E) = \frac{\pi}{\Delta t}\frac{Z_C(E)^2}{Z_H(E)^2} \tag{10}$$

Note that corrections due to a nonzero gradient of the free energy are of the second order. Using this form for $D(E)$, the reweighting rule of eq 8 can be rewritten as

$$\ln w'(E) = \ln w(E) - \ln Z_C(E) \tag{11}$$

This update scheme is similar in form to its multicanonical analogue, $\ln w'(E) = \ln w(E) - \ln Z_H(E)$, and suggests an alternative interpretation of the diffusion-optimized ensemble. Instead of a flat $Z_H(E)$, as in the multicanonical method, the diffusion-optimized choice of weights (eq 6) leads to an ensemble where the transition histogram $Z_C(E)$ is flat.

In a simulation, the $Z_C(E)$ histogram can readily be accumulated, which makes eq 11 a computationally convenient method for generating new weights. Empirically, we find that this method produces significantly less noisy weight estimates, compared to eq 8; the fluctuations in $Z_C(E)$ tend to be much smaller than those in $Z_H(E)$ and $df/dE$. These factors lead to a robust estimation of the new weights with eq 11. The diffusivity analysis underlying this method has previously been successfully used for analyzing molecular simulation trajectories, and constructing optimal reaction coordinates,[22−24] but has, to our knowledge, never before been applied in the context of enhanced sampling. We will refer to this technique as the transition histogram (TH) method.

**2.3. Performance Comparison on a 1D Model System.** We illustrate the performance of the two procedures for determining the diffusion coefficient on a simple one-dimensional model system, adapted from previous work.[38] Here, the state of the system is a number $x$ between 0 and 1, discretized as $x_i = i/100$, $i = 0,1,...,100$. The system is assigned a potential energy $U(x)$ with a double-well shape, $U(x)/kT = 2\cos(2\pi x - \pi)$. We study the system under MC dynamics. In each MC update, a new trial state is generated by adding to the old state a random step drawn from a Gaussian distribution with zero mean and a standard deviation of 0.015. The trial state is subject to a Metropolis−Hastings accept/reject step.

We study the motion of the system along the coordinate $p_{fold}(x)$, which is an optimal reaction coordinate for this system[38] and essentially is the ratio between the first passage time from 0 to $x$ and that for crossing the entire interval from 0 to 1. This coordinate can be calculated as[38]

$$p_{fold}(x) = \int_0^x Z_C^{-2}(x')Z_H(x')dx' \Big/ \int_0^1 Z_C^{-2}(x')Z_H(x')dx' \tag{12}$$

Using the same simulation trajectories, we compute the diffusion coefficient along the $p_{fold}$ coordinate in two ways, with eq 7 or eq 10, respectively. The derivative $df/dE$ in eq 7 is estimated by linear regression of several neighboring points, where the number of points is reduced with increasing statistics. The results obtained with the two methods are compared in Figure 3. Both methods seem to converge toward the same $D(p_{fold})$ curve with increasing sample size, but the convergence is considerably faster with the TH method (eq 10) than it is with the LW technique (eq 7).

**2.4. Ring-Shaped Reaction Coordinate.** The results in the previous section suggest that the TH procedure for

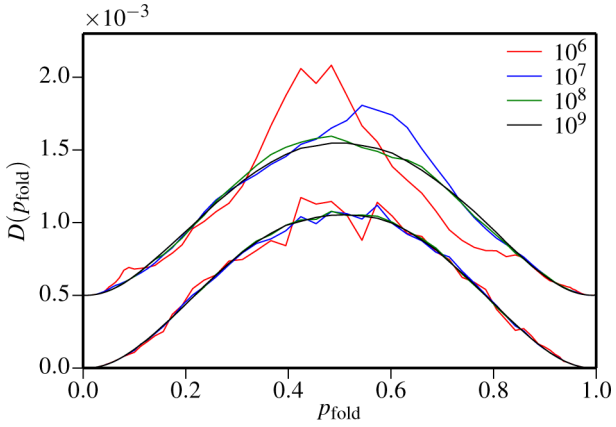



**Figure 3.** Diffusion coefficient along the $p_{fold}$ coordinate (eq 12) in the toy model of Section 2.3, as obtained using eq 7 and eq 10, respectively. Red, blue, green, and black curves represent results based on $10^6$, $10^7$, $10^8$, and $10^9$ MC steps, respectively. The lower-lying curves are the results obtained using eq 10 with a sampling interval of $\Delta t = 16$ MC steps (see Figure 5 below). The curves obtained with eq 7 are shifted upward by 0.0005 for visual clarity. Plot generated using matplotlib.[37]

estimating the diffusion coefficient[22−24] has advantages compared to the LW method,[16] which is currently the standard technique used to derive diffusion-optimized ensembles.[17−21,39] There are, however, two issues that need to be resolved before the TH technique can be applied more generally for enhanced sampling. The primary problem is that unlike the global LW procedure,[16] the TH method does not automatically dictate a specific time scale at which to analyze the process. In cases where the dynamics is not fully diffusive, the diffusion coefficient will generally depend on how frequently statistics are recorded, and we thus need a robust procedure for determining the relevant sampling interval $\Delta t$. The second issue is related to boundary artifacts. When registering statistics at larger time intervals $\Delta t$, one will increasingly underestimate $Z_C$ at the boundaries, since any transitions going to the boundary and back will not be properly registered. This effect can be illustrated by the cut free energy profile $F_C(p_{fold})/kT = -\ln Z_C(p_{fold})$ for the 1D model system described in the previous section. With larger sampling interval, $\Delta t$, the behavior becomes increasingly irregular close to the boundaries (Figure 4a). We will address this second issue first and return to the determination of the sampling interval in the next section.

A solution to the boundary artifact problem for the TH method was recently described in the literature.[38] Here, we follow a different strategy,[40] which is simpler from an implementation perspective and has worked well for us in practice. The boundary artifact problem can be understood by considering that eq 10 is derived under the assumption that $Z_H(E)$ (or, correspondingly, the free energy profile) is constant on the distance of mean displacement, an assumption that is violated at the boundary (with $Z_H = 0$ in the forbidden region). This consideration, however, also suggest a way to rectify the issue. Given a coordinate $E$ with boundaries at points $E_{min}$ and $E_{max}$, another exact copy of the coordinate, $E'$, is created and the two coordinates are joined at the boundaries ($E_{min}$ with $E'_{min}$ and $E_{max}$ with $E'_{max}$) to form a coordinate with ring topology. Whenever a trajectory reaches one of the boundaries either $E$ or $E'$ is selected as current coordinate with equal probability of 0.5. The dynamics on such a constructed system with the

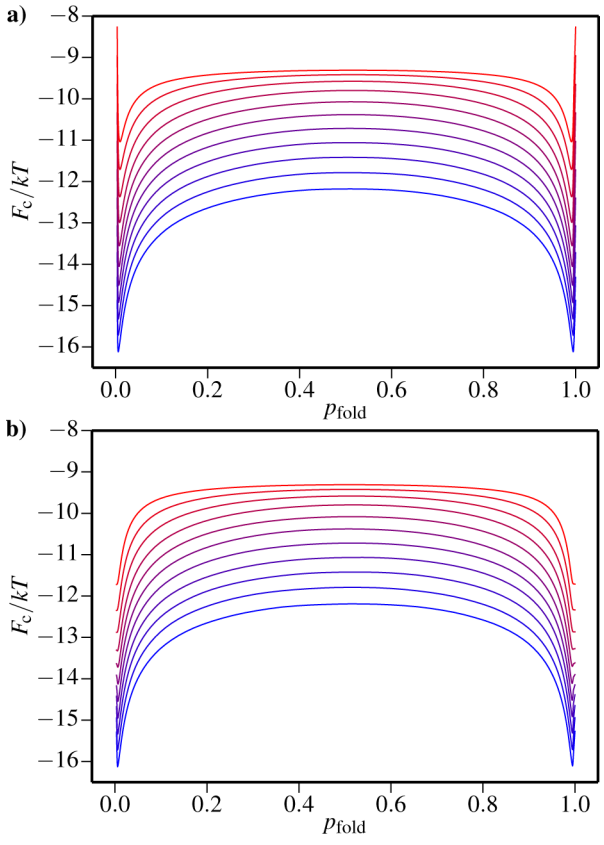



**Figure 4.** Cut free-energy profile $F_C(p_{fold})/kT = -\ln Z_C(p_{fold})$ in the toy model of Section 2.3, as obtained using different sampling intervals $\Delta t$ = 1, 2, 4,...,$2^{10}$ MC steps. Lower-lying curves correspond to smaller $\Delta t$. The $Z_C(p_{fold})$ counting was done using (a) the normal coordinate and (b) the ring-topology coordinate (see Figure 2). Plots generated using matplotlib.[37]

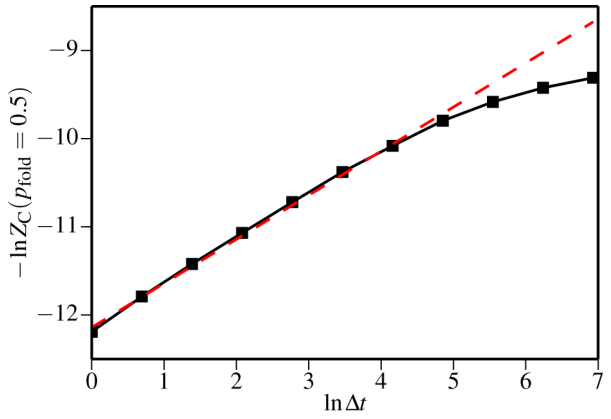



**Figure 5.** Dependence of $-\ln Z_C(p_{fold} = 0.5)$ on the sampling interval $\Delta t$ (in MC steps) in the toy model of Section 2.3. The straight line corresponds to a linear dependence of $-\ln Z_C(p_{fold} = 0.5)$ on $\ln \Delta t$ with slope 1/2, as expected for diffusive dynamics. This behavior is to a good approximation observed for small to intermediate $\Delta t$, while deviations, as expected, can be seen at large $\Delta t$. Plot generated using matplotlib.[37]

extended configuration space is not modified, since the second coordinate is an exact copy of the first one. Since the system has no boundaries, there are no boundary artifacts. Implementation-wise, it is not necessary to model the additional regions explicitly. Instead, we use reflection at the boundaries. During the sampling interval, $\Delta t$, we register whether the simulation visits either $E_{min}$ or $E_{max}$. If it does, we randomly and with equal probability select among two options: (i) register the move as normal (Figure 2b, black), or (ii) register the move as if it reflected in the boundary (Figure 2b, blue). If multiple boundary visits have occurred within $\Delta t$, this selection rule is applied multiple times, registering complete traversals of the entire range if both boundaries have been observed. Using the system from the previous section as an example, we see that this technique successfully removes most of the artifacts at the boundary (Figure 4b).

**2.5. Choice of Sampling Interval.** In addition to the boundary artifacts, the results in Figure 4 illustrate another point. The shift of the $-\ln Z_C(p_{fold})$ curves in Figure 4 with increasing sampling interval $\Delta t$ provides an indication of whether or not the assumption of diffusive dynamics is valid. In the case of fully diffusive dynamics, $-\ln Z_C(p_{fold})$ will vary as 1/2 $\ln \Delta t$ with $\Delta t$ (from eq 10 with constant $D(E)$). To illustrate this more clearly, we plot the relationship between $-\ln Z_C(p_{fold})$ and $\ln \Delta t$ for a single value, $p_{fold}$ = 0.5 (Figure 5). Not surprisingly, deviations from diffusive behavior can be seen for large $\Delta t$. In fact, for very large $\Delta t$, $-\ln Z_C(p_{fold})$ should become independent of $\Delta t$. There is, however, a substantial $\Delta t$ regime where the results agree very well with the behavior expected for diffusive dynamics.

In general, for systems with subdiffusive dynamics, it is clear that the diffusion coefficient will depend strongly on the choice of sampling interval $\Delta t$. Intuitively, we wish to determine a time scale that best reflects the overall dynamics in the system, in order to optimize the flow across the range of the reaction coordinate. Interestingly, the LW procedure does not have a parameter analogous to $\Delta t$, as this time scale is inherently encoded in the global nature of the method.

To determine $\Delta t$, we follow a heuristic approach, in which we apply eq 10 to estimate the diffusion coefficient for different $\Delta t$, $D_{\Delta t}(E)$. Using the $D_{\Delta t}(E)$ profiles, we select $\Delta t$ by maximizing a theoretical round-trip speedup factor, $R(\Delta t)$, for the optimized ensemble relative to the flat-histogram ensemble. Assuming eq 3, this speedup factor is calculated as

$$R(\Delta t) = \int_{E_{min}}^{E_{max}} \frac{dE}{D_{\Delta t}(E) P_{flat}(E)} \Big/ \int_{E_{min}}^{E_{max}} \frac{dE}{D_{\Delta t}(E) P_{opt,\Delta t}(E)} \quad (13)$$

where $P_{flat}(E)$ and $P_{opt,\Delta t}(E)$ are the energy distributions for the two ensembles. The subscript $\Delta t$ indicates that the optimized distribution depends on $D_{\Delta t}(E)$ (eq 5). Equation 13 for the speedup factor relies on the assumption of diffusive dynamics and is only approximate. By forming the ratio between round-trip times, we expect, however, that these uncertainties will partly cancel out. We furthermore expect $R(\Delta t)$ to decrease toward unity for large $\Delta t$, where the position-dependence of $D_{\Delta t}(E)$ becomes small.

To illustrate this procedure, we apply it to the standard ferromagnetic Ising model on a 32 × 32 square lattice. Using multicanonical simulations, we estimate $D_{\Delta t}(E)$, $P_{opt,\Delta t}(E)$ and the speedup factor $R(\Delta t)$ for different $\Delta t$. As can be seen from Figure 6a, the shape of the distribution $P_{opt,\Delta t}(E)$ shows a noticeable $\Delta t$ dependence. To find a favorable shape with respect to round-trip rate, we maximize $R(\Delta t)$ in eq 13. The maximum occurs near $\Delta t = 32 \times 10^3$ MC sweeps (Figure 6b). For this $\Delta t$, $P_{opt,\Delta t}(E)$ exhibits a clear peak between low and high energies. A similarly located peak occurs with the LW


dx.doi.org/10.1021/ct400844x | J. Chem. Theory Comput. 2014, 10, 543−553

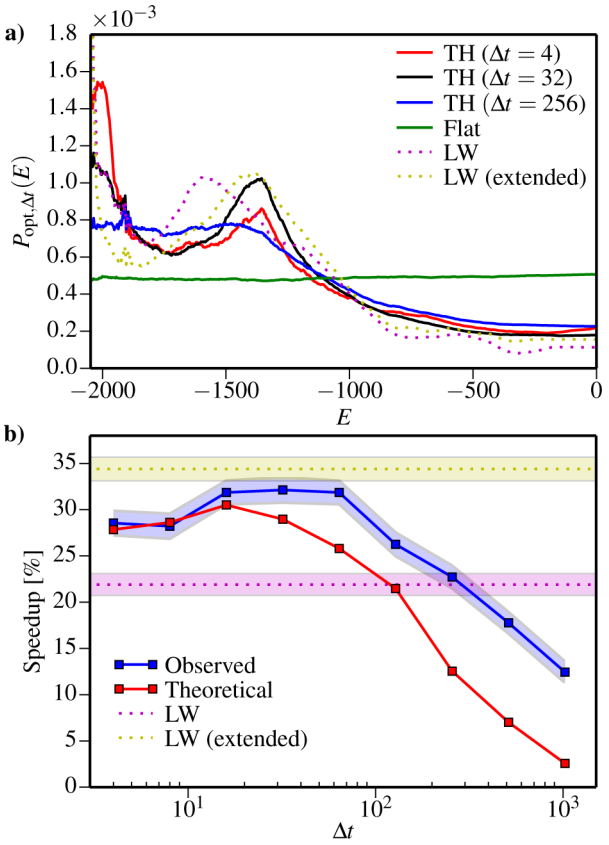



**Figure 6.** (a) Energy distribution in the diffusion-optimized ensemble, $P_{\mathrm{opt},\Delta t}(E)$, for the 2D Ising model on a 32 × 32 lattice, as obtained using the TH procedure for three different sampling intervals: $\Delta t = 4$ (red), $\Delta t = 32$ (black), and $\Delta t = 256$ (blue). Here, the time unit is $10^3$ MC sweeps, where one sweep consists of one attempted move per spin. The green line shows the energy histogram from a flat-histogram simulation. All weight functions are estimated from a single initial multicanonical run with the program MUNINN[41] consisting of $2 \times 10^5$ MC sweeps, of which the first $10^5$ were discarded for thermalization (for an illustration of the run-time evolution in this run, see Figure S3). For comparison, diffusion-optimized weights are also estimated using the LW method, eq 8: 'LW' denotes a simulation based on the same initial run as for the TH method ($1 \times 10^5$ MC sweeps), while 'LW (extended)' denotes weights estimated on a longer initial simulation ($3 \times 10^5$ MC sweeps), illustrating that more statistics are needed for this method (see SI Figure S1). (b) To determine the relevant sampling interval, a set of 32 independent runs of length $10^6$ MC sweeps is carried out for each weight function, and the speedup in round-trip rate relative to the flat-histogram ensemble is plotted against the sampling interval $\Delta t$ used for determining the optimized weights. Red and blue squares represent theoretical (eq 13) and observed values, respectively, while the shaded areas represent the standard error. The theoretical and simulated results for the speedup factor both suggest a maximum near $\Delta t = 32$. The two LW cases from (a) are included for reference, illustrating that the TH and LW methods converge to similar ensembles when choosing the optimal $\Delta t$ for the TH method but that more simulation data are needed for the LW method. Plots generated using matplotlib.[37]

method, and in the optimized ensemble determined by Trebst et al. for the same model.[16] In agreement with this previous work, our analysis thus suggests that placing higher statistical weight on the intermediate energies promotes transitions between low and high energies. To verify this conclusion, we measure the actual round-trip speedup factor through simulations with the optimized weight functions obtained for different $\Delta t$. The actual speedup factors turn out to agree quite well with the ones predicted using eq 13 with $E_{\max} = 0$ and $E_{\min} = -2048$ (Figure 6b). In particular, the maximal real and predicted speedups are similar, ∼ 32%, and occur at approximately the same $\Delta t$, and also correspond to the speedup obtained with the LW method. This suggests that maximizing $R(\Delta t)$ in eq 13 can indeed be considered as a relevant criterion for determining optimal weights. Finally, Figure 6 also demonstrates that while the TH and LW methods generally converge to the same solution for this system, more simulation data is necessary in order to obtain converged estimates for the LW procedure (Figure 6, LW vs LW (extended)).

## 3. BIOLOGICAL APPLICATIONS

The previous sections introduced a robust new technique for estimating diffusion optimized ensembles. We now investigate the potential of the method in two more challenging sampling problems. Both problems are concerned with the simulation of biological macromolecules. In the first case, we consider a fully atomistic simulation of the reversible folding of a small peptide, the $\beta$-hairpin-forming GB1p. In the second, we study the behavior of a coarse-grained model of peptide aggregation. In both cases, we compare the performance of the flat-histogram method to that of diffusion-optimized sampling, calculating the diffusion coefficient using both our new TH-based method and the LW method. We follow a three-step procedure. First, an initial multicanonical simulation is performed by using the MUNINN multihistogram program,[41] which iteratively refines the weight function $w(E)$ using the generalized multihistogram equations,[9] eventually converging to $1/g(E)$. Second, based on data from the initial run, diffusion-optimized and flat-histogram weights are constructed according to either $\ln w_{\mathrm{TH}}(E) = \ln w(E) - \ln Z_{\mathrm{C},\Delta t}(E)$, $\ln w_{\mathrm{LW}}(E) = \ln w(E) + 1/2[\ln \mathrm{d}f/\mathrm{d}E - \ln Z_{\mathrm{H}}(E)]$ or $\ln w_{\mathrm{flat}}(E) = \ln w(E) - \ln Z_{\mathrm{H}}(E)$, respectively. For the TH method, the sampling interval, $\Delta t$, is chosen so as to maximize the theoretical speedup factor in eq 13. If the initial simulation is carried out using multiple parallel threads, $Z_{\mathrm{C},\Delta t}(E)$ and $Z_{\mathrm{H}}(E)$, and the LW statistics are accumulated from all threads. In the third and last step, simulations are performed with the computed weight functions $w_{\mathrm{TH}}(E)$, $w_{\mathrm{LW}}(E)$, and $w_{\mathrm{flat}}(E)$, and the sampling efficiencies are compared.

**3.1. All-Atom MC Simulation of GB1p with Different Reaction Coordinates.** The GB1p peptide (GEWTYDDATKTFTVTE) is a 16-residue fragment of the GB1 protein domain. This segment forms the second of two $\beta$-hairpins in the GB1 domain (Figure 7a) and was demonstrated to form a $\beta$-hairpin on its own as well.[42] As a well characterized example of a minimal $\beta$-structure unit, GB1p has become a widely used model system in computational studies; numerous methods and force fields have been benchmarked against thermodynamic and kinetic data for this peptide.[32,43−52] The folding of a hairpin involves the formation of long-range contacts, which makes it a more challenging problem compared to helix folding.

The folding thermodynamics of GB1p was well captured by all-atom MC simulations with the program PROFASI,[53] in which reversible and cooperative folding to the $\beta$-hairpin state was observed and good quantitative agreement was found with experimental melting properties.[50,54] This program has also been used to simulate larger proteins, including the mixed $\alpha + \beta$ Top7 protein with >90 residues.[39] The PROFASI software contains an efficient force field that is parametrized to implicitly

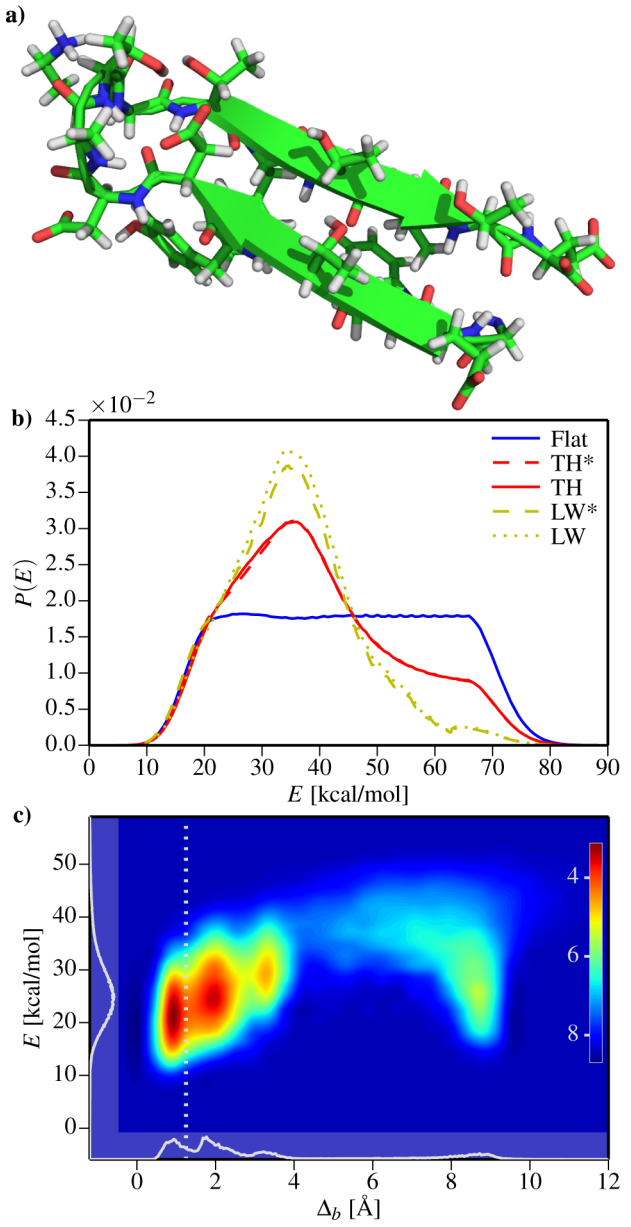



**Figure 7.** (a) Structure of the GB1p hairpin as it appears in the native state of GB1 (PDB code 1GB1), illustrated using PyMOL.[55] (b) Energy distributions in the TH-based (red) and LW-based (green) diffusion-optimized ensembles, compared to the flat-histogram ensemble (blue). The weight functions (LW* and TH*) are illustrated by a dashed line while the observed histograms in simulation are shown using a solid line. (c) Free-energy landscape $F(\Delta_b,E)$ calculated as a function of backbone RMSD, $\Delta_b$, and energy, $E$, at $T$ = 279 K. The bottom and left gray areas show the projections on each of the reaction coordinates. The vertical line represents the low-RMSD cut ($\Delta_b$ = 1.25 Å) used when counting round trips. Plot generated using matplotlib.[37]

include solvent effects. Both this force field and the PROFASI move set have been described in detail in previous work.[50,54]

Using PROFASI in combination with the MUNINN multihistogram method, we conduct GB1p simulations using the different weight schemes. We determine the weight functions of the ensembles, $w_{TH}(E)$, $w_{LW}(E)$, and $w_{flat}(E)$, by an initial multicanonical simulation that covers an energy interval corresponding to the temperature range 279−367 K. This calculation is run on eight parallel threads, each executing $10^9$ MC steps. Having found $w_{TH}(E)$, $w_{LW}(E)$, and $w_{flat}(E)$, we conduct three batches of simulations with exactly the same settings, one for each weight function, to compare the dynamics in the two ensembles. Each batch consists of 24 independent runs of length $4 \times 10^9$ MC steps.

The energy distribution obtained with $w_{flat}(E)$ is to a very good approximation uniform (Figure 7b). The use of $w_{TH}(E)$ and $w_{LW}(E)$ leads to a modified distribution, demonstrating that the calculated diffusion coefficient is indeed position-dependent. The diffusion-optimized ensembles assign highest statistical weight to energies just above the range observed in the folded state (Figure 7b).

Canonical averages at a given temperature can be extracted from generalized ensemble simulations by reweighting. The free energy $F(\Delta_b,E)$ at $T$ = 279 K, calculated as a function of backbone RMSD, $\Delta_b$, and energy, $E$, shows that GB1p samples a few different local free-energy minima in the simulations (Figure 7c). The two main minima correspond to $\beta$-hairpin states and are found at $\Delta_b \sim 1.0$ Å and $\Delta_b \sim 2.0$ Å, respectively.

To compare the relative efficiency of the methods, we measure the time required for the system to move back and forth between the folded and unfolded states. Roughly, we expect the computational effort to scale as $1/\tau_{rt}$ with the round-trip time $\tau_{rt}$. We define round trips in terms of $\Delta_b$. Using rather strict criteria, we say that the system is folded if $\Delta_b < 1.25$ Å and unfolded if $\Delta_b > 12.0$ Å. With the 'boundaries' defined this way, we compute the average time for round trips in $\Delta_b$ by using the up/down-walker representation of Section 2.1. Let $\tau_{up}$ denote the average length of an up-walker trajectory segment, or the average survival time of such a walker, and let $\tau_{down}$ be the corresponding quantity for down-walkers. The average round-trip time can then be found as $\tau_{rt} = \tau_{up} + \tau_{down}$. By applying this method to our GB1p simulations, we find that the TH-based diffusion-optimized ensemble yields a 14% reduction of $\tau_{rt}$, compared to the flat-histogram ensemble (Table 1). In the diffusion-optimized case, $\tau_{down}$ = 170 and $\tau_{up}$ = 39 in million MC steps. The corresponding flat-histogram values are 199 and 38 million MC steps, respectively. The reduction in $\tau_{rt}$ for the diffusion-optimized ensemble thus comes from a reduced $\tau_{down}$, indicating faster folding, whereas $\tau_{up}$ remains approximately the same as in the flat-histogram runs.

Interestingly, while the TH and LW methods agree on the main features of the weight function, there are some discrepancies in amplitude. In general, we expect some degree of discrepancy between the methods, arising from the fact that the underlying process is subdiffusive and that the two methods approximate it by a diffusive process in distinct ways (either locally or globally). The global LW analysis, when fully converged, will presumably produce weights that are most compatible with the (global) target of a minimal round trip time. In practice, however, it is difficult to converge this quantity, and observed discrepancies will therefore typically be at least partially due to lack of convergence. In the current case, we see that the speedup obtained by the LW method is actually somewhat smaller than that of the TH method (Table 1), indicating that the increased robustness of the TH method is in this case preferable over the theoretical advantages of the LW method.

The speedup of 14% is rather modest. One plausible explanation for this result is that energy is a poor reaction coordinate for the system in question. One obvious problem is that several distinct structural states are mapped to the same energy values (Figure 7c). This makes it difficult to distinguish

**Table 1. Round-Trip Statistics from GB1p Simulations with the Flat-Histogram and Diffusion-Optimized Ensembles, with Energy, $E$, or Backbone RMSD, $\Delta_b$, as the Reaction Coordinate[a]**

| | flat | | TH | | LW | |
|---|---|---|---|---|---|---|
| reaction coordinate | $N$ | $\tau_{rt}$ | $N$ | $\tau_{rt}$ | $N$ | $\tau_{rt}$ |
| $E$ ($96 \times 10^9$ MC steps) | 388 | 237 ± 10 | 443 | 209 ± 8 | 400 | 230 ± 10 |
| $\Delta_b$ ($24 \times 10^9$ MC steps) | 80 | 265 ± 26 | 137 | 168 ± 11 | | |

[a]$N$ denotes the number of observed round trips. The average round-trip time $\tau_{rt}$ is given in million MC steps. In the calculations with $E$ as the reaction coordinate, a total of 96 billion MC steps were collected for each ensemble, using 24 threads. The simulations with $\Delta_b$ as the reaction coordinate contained a total of 24 billion MC steps for each ensemble, generated using 16 threads. For comparison, the final column contains results obtained with the LW method. We were unable to obtain converged weights for the LW method in the $\Delta_b$ case (SI Figure S4), presumably due to the large barrier and the limited number of observed round trips.

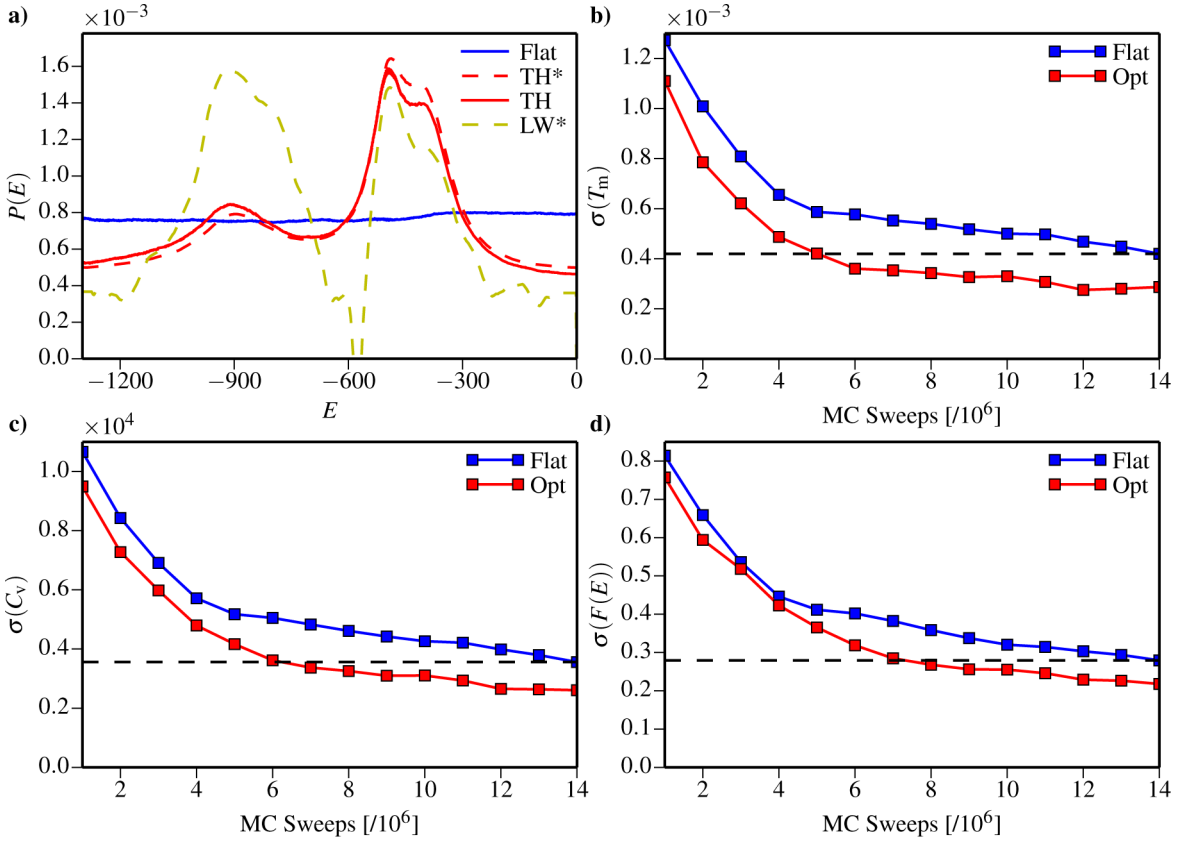



**Figure 8.** (a) Weight-histogram (dashed) and observed histogram (solid) for TH-based diffusion optimized ensemble (red), the LW-based optimized ensemble (yellow) and the flat-histogram method (blue). For the LW method, convergence was not obtained (see SI Figure S5b) and no simulation was run. Statistical error on (b) the aggregation temperature, $\sigma(T_m)$, (c) the heat capacity, $\sigma(C_v)$, and (d) the free energy, and $\sigma(F(E))$, against simulation length in the diffusion-optimized and flat-histogram simulations. For precise definitions of $\sigma(C_v)$ and $\sigma(F(E))$, see the main text. The horizontal dashed lines indicate final error levels with the flat-histogram method, which are reached 2−3 times faster with the diffusion-optimized method. Plots generated using matplotlib.[37]

between these states and to resolve the transitions between them, which in turn leads to poorly defined diffusion coefficients for these regions. The question of determining optimal reaction coordinates is a whole topic in itself[24,56,57] and beyond the scope of this paper. We, however, briefly investigate this potential explanation by considering the RMSD as a reaction coordinate, which should provide a clearer distinction between the observed local minima (Figure 7c). A shortcoming of the RMSD coordinate is its insensitivity to atom collisions, which means that unphysical conformations can have a relatively low RMSD. We therefore construct a hybrid weight function, consisting of a product between a standard constant temperature Boltzmann weight and the generalized-ensemble weight $w_{gen}$, so that the acceptance rate becomes

$$P_{acc}(S_1 \rightarrow S_2) = \min\left[1, \frac{w_{gen}(\Delta_b(S_2))e^{-E(S_2)/kT}}{w_{gen}(\Delta_b(S_1))e^{-E(S_1)/kT}}\right] \quad (14)$$

where $E$ is the PROFASI energy function, $S$ represents the state of the system (GB1p), and the temperature is set to $T = 300$ K. This procedure has parallels to metadynamics, where collective variables are combined with a force field in a similar manner. It should be noted that although the density of states is only recorded over the $\Delta_b$ reaction coordinate, the hybrid form of eq 14 induces a density of states that is measured with respect to the Boltzmann distribution of $E$

$$g_E(\Delta_b) = \int_S e^{-E(S)/kT}\delta(\Delta_b(S) - \Delta_b)dS \quad (15)$$


dx.doi.org/10.1021/ct400844x | J. Chem. Theory Comput. 2014, 10, 543−553

It is this redefined density of states that is used to estimate the flat-histogram weights. Likewise, the $Z_C$ statistics have an underlying bias induced by $E$.

Following the same procedure, we again perform an initial flat-histogram simulation, this time with $\Delta_b$ as the reaction coordinate, from which we estimate both diffusion-optimized and flat-histogram weights, restricting ourselves to the range 0 Å $\leq \Delta_b \leq$ 13 Å. We proceed as before, calculating weight functions $w_{gen,TH}$, $w_{gen,LW}$, and $w_{gen,flat}$. We could not obtain converged values for $w_{gen,LW}$ from the available initial simulation data (see SI Figure S4), and we therefore had to refrain from performing a diffusion-optimized simulation using weights obtained from the LW method. This problem is presumably related to the higher energy barrier observed for this reaction coordinate, and the fewer round trips observed with this improved reaction coordinate. In contrast, our TH method remains robust in this case. For $w_{gen,TH}(\Delta_b)$ and $w_{gen,flat}(\Delta_b)$, we perform simulations with eq 14 as the acceptance criterion, measuring the round-trip time $\tau_{rt}$, defined as above. As shown in Table 1, a significantly larger speedup (~58%) is obtained for this case. Comparing all our GB1p simulations, the round-trip rate is highest with $\Delta_b$ as reaction coordinate and diffusion-optimized weights, and we find the TH method to be superior to the LW method for this system.

**3.2. Stick Model.** The second application for our algorithm is a minimal structure-based model for peptide aggregation into amyloid fibrils,[58] in which the peptides are represented by unit-length sticks on a cubic lattice and interact by hydrogen bonding and hydrophobicity forces. Amyloid formation is known to occur with reproducible sigmoidal kinetics,[59,60] possibly indicating a nucleation-dependent process. The 'stick model' can be studied for system sizes sufficiently large to reproduce this property, and has the advantage over more detailed models of permitting studies of the interplay between aggregate length and width in fibril nucleation. At the temperature at which fibrillation sets in, the energy distribution has a clearly bimodal shape, signaling the coexistence of a high-energy phase dominated by small aggregates and a low-energy phase where large fibrillar aggregates are present.[58] Here, we apply our method to a 256-peptide instance of this model. For this system size, the high- and low-energy phases are separated by a free-energy barrier of height > $3kT$. This makes the system a challenge to simulate, and thus a well suited target for generalized-ensemble methods.

We simulate this model using the same setup as for GB1p, starting by estimating the weight functions $w_{TH}(E)$, $w_{LW}(E)$, and $w_{flat}(E)$ through an initial multicanonical calculation (Figure 8a). There is a substantial difference between the weights obtained using the LW and TH methods, with the LW method assigning much more weight to the left peak. However, as for the GB1 simulations, the LW-algorithm failed to converge within the allocated simulation time (see SI Figure S5), and at least part of this discrepancy is therefore presumably of a statistic nature. Due to the zero-weight regions induced by the lack of convergence, it was not possible to perform a simulation in a LW diffusion optimized ensemble.

For $w_{TH}(E)$ and $w_{flat}(E)$, we performed production simulations, each consisting of 64 independent runs of length $1.6 \times 10^8$ MC sweeps, where one sweep consists of 256 elementary moves. The move set consists of single-peptide translations and rotations (60%) and a Swendsen–Wang-type cluster move (40%).[58,61] The runs are started from random initial conditions, and the first $0.2 \times 10^8$ MC sweeps of every run are discarded for thermalization.

In the calculations with $w_{flat}(E)$ as weight function, the energy distribution is, as expected, essentially flat (Figure 8a). The energy distribution obtained with the $w_{opt}(E)$ weights has two peaks (Figure 8a). The highest peak is located at the free-energy barrier separating the high- and low-energy phases.[58] This transition region is thus assigned additional weight by the diffusion-optimized algorithm. The second, smaller peak occurs within the fibrillar low-energy phase.

We evaluate the relative performance of the two methods by comparing statistical errors and their evolution with increasing sample size. Using the jackknife method,[62] we compute statistical errors after every 20 million MC sweeps. We focus on three quantities: (i) the fibrillation temperature, $T_m$, (ii) the heat capacity, $C_v$, and (iii) the free energy $F(E) = -kT \ln P(E)$. We define $T_m$ as the maximum of the heat capacity and determine the standard error of the mean $T_m$, denoted by $\sigma(T_m)$. The estimation of heat capacity is very sensitive to the quality of the sampling, and often serves as a benchmark in assessing sampling efficiency.[8,9,15] We computed the standard error on $C_v$ at 200 temperatures in the interval 0.66–0.69, which contains the fibrillation temperature, $T_m \approx 0.671$. We take the average of these errors as our heat capacity uncertainty, $\sigma(C_v)$. Similarly, we define the uncertainty in the free energy, $\sigma(F(E))$, as the average of the standard errors at all the 1300 energies in the interval studied, calculated using $T = T_m$.

Figure 8 shows how the errors $\sigma(T_m)$, $\sigma(C_v)$, and $\sigma(F(E))$ evolve with increasing simulation time. All three properties converge faster with the diffusion optimized method than they do with the flat-histogram technique. The final error levels in the flat-histogram simulations are reached 2–3 times faster with the diffusion-optimized method. As in the GB1p case, we find that the construction of the diffusion-optimized ensemble is greatly facilitated by the TH method, which is statistically more robust than the LW scheme.

## 4. DISCUSSION

This paper introduces a robust technique for the estimation of diffusion coefficients for use in diffusion-optimized molecular simulations. This area has recently received considerable attention with reports of spectacular gains in efficiency both in the context of Monte Carlo[18] and molecular dynamics simulations.[20,21] One important challenge that has hindered a broad adoption of diffusion-optimized ensembles is the difficulty in obtaining reliable estimates of the diffusion coefficient. We present a method that dramatically improves the precision of the estimation procedure, requiring considerable fewer data than the currently used approach.

The proposed weight update rule is remarkably simple to use, and appears as a natural generalization of the standard multicanonical equivalent. While the multicanonical ensemble seeks to flatten the histogram of states, the proposed method will instead flatten the histogram of transitions through each state. We demonstrate that in cases where the diffusion coefficient is constant, the proposed update rule reduces to the standard multicanonical equivalent. This suggests that, in principle, the diffusion-optimized ensemble can be broadly applied as a general improvement over the multicanonical ensemble. In practice, the diffusion optimized ensemble has not yet replaced its multicanonical counterpart for two main reasons: (a) the estimation of diffusion optimized weights has been unreliable when dealing with limited simulation data, and

(b) in molecular simulation, the dynamics are typically not fully diffusive, and it has been unclear to which extent this violation of the diffusion assumption affects the performance. We believe that the technique presented in this paper solves the first problem. The second problem is harder to address exhaustively. On one hand, the TH represents a measure of local mobility, and as such, it should be meaningful to flatten this quantity even when dynamics are not fully diffusive, despite the fact that one in this case lacks the coupling to the round trip time (eq 3). On the other hand, we expect that the method would break down in cases of extreme subdiffusivity. Empirically, we have in this paper demonstrated performance gains on several subdiffusive problems and have not observed any cases where our method was outperformed by the multicanonical technique.

In the current work, optimized weights are estimated from a relatively converged multicanonical ensemble. A natural extension is to estimate them iteratively during the course of a simulation. Compared to current diffusion estimation approaches, our method should have significantly fewer problems in estimating reliable diffusion coefficients in early stages of the simulation, when only few (or no) round trips have occurred. We are currently working on a variant of the multihistogram method[9,41] to investigate the potential of such an approach. Given that our results on the GB1p system demonstrated a strong dependence on using a good reaction coordinate, an interesting avenue for future research would be to optimize the reaction coordinate simultaneously with the simulation weights. The TH method has previously been demonstrated to work well for this task.[22−24] Even greater benefits might be obtained by simultaneously simulating on multiple reactions coordinates, as is routinely done in the field of metadynamics. A direct generalization of our approach to two-dimensional diffusion is not straightforward. However, a potentially potent solution would be to use a bias-exchange approach, where each reaction coordinate is represented by a separate replica.[63] By iteratively tuning each replica toward the diffusion-optimized ensemble associated with its reaction coordinate, we envisage that the sampling efficiency for the full replica-exchange ensemble can be significantly improved compared to the flat histogram approach.

## ■ ASSOCIATED CONTENT

### Supporting Information

Figures showing detailed comparison of the TH method and the LW method performance for the 2D Ising model, including sensitivity to the length of simulation and location of boundaries; weight estimation with the TH method and LW method for the GB1P protein simulation (with RMSD as reaction coordinate); weight estimation with the LW-method on the stick model system. This material is available free of charge via the Internet at http://pubs.acs.org.

## ■ AUTHOR INFORMATION

### Corresponding Author

*E-mail: wb@bio.ku.dk.

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

P.T. was supported by the Lundbeck Foundation. K.L.L. and W.B. were supported by a Hallas−Møller stipend from the Novo Nordisk Foundation. Part of the simulations were performed on resources provided by the Swedish National Infrastructure for Computing (SNIC) at the LUNARC facility. We thank Jes Frellsen for valuable input on generalized ensembles.

## ■ REFERENCES

(1) Kirkpatrick, S.; Gelatt, C. D., Jr.; Vecchi, M. P. *Science* **1983**, *220*, 671−680.
(2) Swendsen, R. H.; Wang, J. S. *Phys. Rev. Lett.* **1986**, *57*, 2607−2609.
(3) Hukushima, K.; Nemoto, K. *J. Phys. Soc. (Jap.)* **1996**, *65*, 1604−1608.
(4) Marinari, E.; Parisi, G. *Europhys. Lett.* **1992**, *19*, 451−458.
(5) Lyubartsev, A. P.; Martsinovski, A. A.; Shevkunov, S. V.; Vorontsov-Velyaminov, P. N. *J. Chem. Phys.* **1992**, *96*, 1776−1783.
(6) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.
(7) Berg, B. A.; Neuhaus, T. *Phys. Lett. B* **1991**, *267*, 249−253.
(8) Wang, F.; Landau, D. P. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(9) Ferkinghoff-Borg, J. *Eur. Phys. J. B* **2002**, *29*, 481−484.
(10) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695−708.
(11) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(12) Hansmann, U. H. E.; Okamoto, Y. *J. Comput. Chem.* **1993**, *14*, 1333−1338.
(13) Bachmann, M.; Janke, W. *Phys. Rev. Lett.* **2003**, *91*, 208105.
(14) Taylor, M. P.; Paul, W.; Binder, K. *J. Chem. Phys.* **2009**, *131*, 114907.
(15) Jónsson, S. Æ.; Mohanty, S.; Irbäck, A. *J. Chem. Phys.* **2011**, *135*, 125102.
(16) Trebst, S.; Huse, D. A.; Troyer, M. *Phys. Rev. E* **2004**, *70*, 046701.
(17) Trebst, S.; Troyer, M.; Hansmann, U. H. E. *J. Chem. Phys.* **2006**, *124*, 174903.
(18) Nadler, W.; Meinke, J. H.; Hansmann, U. H. E. *Phys. Rev. E* **2008**, *78*, 061905.
(19) Singh, S.; Chiu, C.-C.; de Pablo, J. J. *J. Stat. Phys.* **2011**, *145*, 932−945.
(20) Singh, S.; Chiu, C.-C.; de Pablo, J. J. *J. Chem. Theory Comput.* **2012**, *8*, 4657−4662.
(21) Jiang, P.; Yasar, F.; Hansmann, U. H. E. *J. Chem. Theory Comput.* **2013**, *9*, 3816−3825.
(22) Krivov, S. V.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 13841−13846.
(23) Krivov, S. V. *J. Phys. Chem. B* **2011**, *115*, 12315−12324.
(24) Krivov, S. V. *J. Phys. Chem. B* **2011**, *115*, 11382−11388.
(25) Nadler, W.; Hansmann, U. H. E. *Phys. Rev. E* **2007**, *75*, 026109.
(26) Berne, B. J.; Borkovec, M.; Straub, J. E. *J. Phys. Chem.* **1988**, *92*, 3711−3725.
(27) Woolf, T. B.; Roux, B. *J. Am. Chem. Soc.* **1994**, *116*, 5916−5926.
(28) Socci, N. D.; Onuchic, J. N.; Wolynes, P. G. *J. Chem. Phys.* **1996**, *104*, 5860−5868.
(29) Hummer, G.; García, A. E.; Garde, S. *Phys. Rev. Lett.* **2000**, *85*, 2637−2640.
(30) Favrin, G.; Irbäck, A.; Samuelsson, B.; Wallin, S. *Biophys. J.* **2003**, *85*, 1457−1465.
(31) Chahine, J.; Oliveira, R. J.; Leite, V. B. P.; Wang, J. *Proc. Natl. Acad. Sci. U.S.A.* **2007**, *104*, 14646−14651.
(32) Yang, S.; Onuchic, J. N.; García, A. E.; Levine, H. *J. Mol. Biol.* **2007**, *372*, 756−763.
(33) Hummer, G. *New J. Phys.* **2005**, *7*, 34.
(34) Türkcan, S.; Alexandrou, A.; Masson, J.-B. *Biophys. J.* **2012**, *102*, 2288−2298.
(35) Comer, J.; Chipot, C.; González-Nilo, F. D. *J. Chem. Theory Comput.* **2013**, *9*, 876−882.
(36) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 1088−1093.

(37) Hunter, J. D. *Comput. Sci. Eng.* **2007**, *9*, 90−95.
(38) Krivov, S. V. *J. Chem. Theory Comput.* **2012**, *9*, 135−146.
(39) Mohanty, S.; Meinke, J. H.; Zimmermann, O. *Proteins* **2013**, *81*, 1446−1456.
(40) Krivov, S. V. *Phys. Rev. E* **2013**, *88*, 062131.
(41) Frellsen, J. Probabilistic methods in macromolecular structure prediction. Ph.D. thesis, University of Copenhagen, 2011; http://muninn.sourceforge.net/.
(42) Blanco, F. J.; Rivas, G.; Serrano, L. *Nat. Struct. Mol. Biol.* **1994**, *1*, 584−590.
(43) Dinner, A. R.; Lazaridis, T.; Karplus, M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 9068−9073.
(44) Roccatano, D.; Amadei, A.; Di Nola, A.; Berendsen, H. J. *Protein Sci.* **1999**, *8*, 2130−2143.
(45) Zhou, R.; Berne, B. J.; Germain, R. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 14931−14936.
(46) Zagrovic, B.; Sorin, E. J.; Pande, V. *J. Mol. Biol.* **2001**, *313*, 151−169.
(47) Kussell, E.; Shimada, J.; Shakhnovich, E. I. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 5343−5348.
(48) Bolhuis, P. G. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12129−12134.
(49) Wei, G.; Mousseau, N.; Derreumaux, P. *Proteins* **2004**, *56*, 464−474.
(50) Irbäck, A.; Mohanty, S. *Biophys. J.* **2005**, *88*, 1560−1569.
(51) Chen, J.; Im, W. W.; Brooks, C. L. C. *J. Am. Chem. Soc.* **2006**, *128*, 3728−3736.
(52) Han, W.; Schulten, K. *J. Chem. Theory Comput.* **2012**, *8*, 4413−4424.
(53) Irbäck, A.; Mohanty, S. *J. Comput. Chem.* **2006**, *27*, 1548−1555.
(54) Irbäck, A.; Mitternacht, S.; Mohanty, S. *PMC Biophys.* **2009**, *2*, 2.
(55) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific: San Carlos, CA, 2002.
(56) Best, R. B.; Hummer, G. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6732−6737.
(57) Peters, B.; Bolhuis, P. G.; Mullen, R. G.; Shea, J.-E. *J. Chem. Phys.* **2013**, *138*, 054106.
(58) Irbäck, A.; Jónsson, S. Æ.; Linnemann, N.; Linse, B.; Wallin, S. *Phys. Rev. Lett.* **2013**, *110*, 058101.
(59) Hellstrand, E.; Boland, B.; Walsh, D. M.; Linse, S. *ACS Chem. Neurosci.* **2010**, *1*, 13−18.
(60) Giehm, L.; Otzen, D. E. *Anal. Biochem.* **2010**, *400*, 270−281.
(61) Swendsen, R. H.; Wang, J.-S. *Phys. Rev. Lett.* **1987**, *58*, 86−88.
(62) Miller, R. G. *Biometrika* **1974**, *61*, 1−15.
(63) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.