JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Designing Free Energy Surfaces That Match Experimental Data with Metadynamics

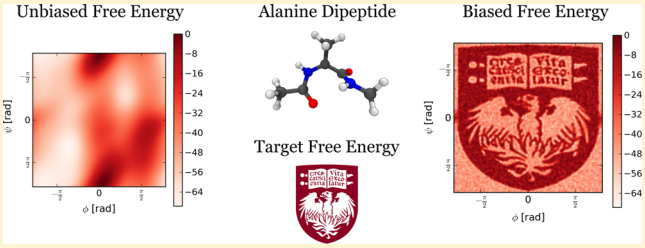
Andrew D. White, James F. Dama, and Gregory A. Voth*

Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, The University of Chicago, 5735 South Ellis Avenue, Chicago, Illinois 60637, United States

Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

Supporting Information

ABSTRACT: Creating models that are consistent with experimental data is essential in molecular modeling. This is often done by iteratively tuning the molecular force field of a simulation to match experimental data. An alternative method is to bias a simulation, leading to a hybrid model composed of the original force field and biasing terms. We previously introduced such a method called experiment directed simulation (EDS). EDS minimally biases simulations to match average values. In this work, we introduce a new method called experiment directed metadynamics (EDM) that creates minimal biases for matching entire free energy surfaces such as radial distribution functions and phi/psi angle free energies. It is also possible with EDM to create a tunable mixture of the experimental data and free energy of the unbiased ensemble with explicit ratios. EDM can be proven to be convergent, and we also present proof, via a maximum entropy argument, that the final bias is minimal and unique. Examples of its use are given in the construction of ensembles that follow a desired free energy. The example systems studied include a Lennard-Jones fluid made to match a radial distribution function, an atomistic model augmented with bioinformatics data, and a three-component electrolyte solution where ab initio simulation data is used to improve a classical empirical model.



INTRODUCTION

Creating physics-based molecular models that match experimental data is one of the major goals of molecular simulations. Such models give insight into the molecular mechanisms of experiments while also leading toward predictive models. The free energy along a collective variable, the potential of mean force (PMF), is one of the most important quantities from a physics-based molecular simulation. When that PMF or quantities calculated from it disagrees with experiments, how to modify the force field of the molecular simulation to rectify the disagreement is usually ambiguous. Modifying the force field to match a specific observable can inadvertently degrade the agreement of other observables. In this work, we show that these issues may be avoided by instead adding a bias that acts only along the collective variable that describes the observable. This leads to a minimal bias and exact match with experiment in that collective variable. This may be accomplished with a new method derived from metadynamics,[1] called experiment directed metadynamics (EDM), which can create biases that minimally modify a molecular simulation while giving exact agreement with a desired PMF. It should be noted immediately that, although EDM is derived from metadynamics, its use is not for calculating a PMF like metadynamics but instead modifying a PMF to match experimental data.

The idea of choosing the PMF of a collective variable by applying bias is similar to other ideas for matching experimental data in the literature.[2−4] Biasing force fields to match experimental data has been recently used as an alternative to traditional iterative construction of experimentally consistent force fields.[5−9] Some of these biasing methods maximize entropy and thus minimally disturb the statistical ensemble of a simulation.[6,7] The main use of these methods is when an approximately correct force field exists but the system of interest disagrees with experimental data. For example, one may have a general protein force field but the specific protein simulated violates some structural data from NMR experiments.[4] The force field can then be made to match the NMR experiments with the addition of a biasing term. The dynamic ensemble refinement of Lindorff-Larsen and co-workers is one biasing method that has been used extensively in the study of proteins with NMR data.[5,10] Pitera and Chodera proposed another minimal biasing method relying on single replicas[6] that was extended by White and Voth to create experiment directed simulations (EDS).[8] EDS uses a form of online stochastic gradient descent[11] to find a minimal biasing force and has been applied to model simple polymers and electrolyte solutions.

The minimal biasing methods described thus far match only average values of experimental data. For the treatment of densities, external biasing fields generated from experimental data are commonly used.[12] External biasing fields are generally not minimal biasing methods. They have seen success in NMR

Received: February 23, 2015

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

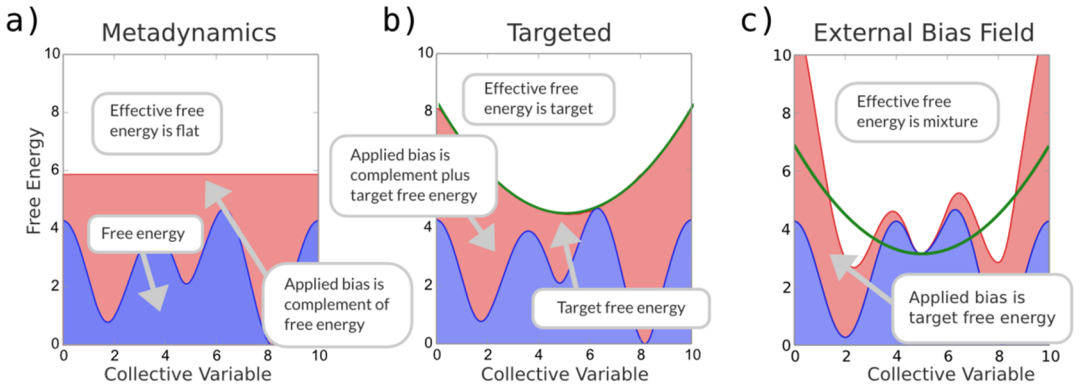


Figure 1. Comparison of metadynamics (a), targeted metadynamics (b), and external biases (c). The unbiased free energy of the system is shown in blue, and the bias is shown in red. The sum dictates the forces and is the effective free energy. Metadynamics replaces the free energy of the system with an effective free energy that is flat and allows diffusive exploration. Targeted metadynamics allows the choice of the effective free energy, shown in green, while still recovering the free energy, in blue. An external bias will add a targeted free energy, in green, to the unbiased free energy, in blue, creating a mixture of the two. The relative fraction of the two free energies is controlled by the strength of the external bias.

structure refinement,[13] X-ray crystallography,[12,14] and cryo-EM electron density fitting.[15] The final ensemble when experimental data is directly added to a simulation is a combination of the experimental data and the innate PMF given by the starting force field. A stronger biasing field relative to the force field tilts the combination toward the target density from experimental data but inadvertently creates higher forces and energies. A demonstration of this is shown in Figure 1c, where a harmonic potential is added to the potential energy to achieve a desired parabolic shape. Instead of a resultant parabolic free energy, a mixture of the harmonic potential and potential energy is the final ensemble. To achieve a closer fit with the harmonic potential, its relative weight must be increased, thereby creating large forces that can lead to energy conservation issues. In the context of fitting proteins to cryo-EM densities, the high forces can dampen protein fluctuations and distort structure such that additional harmonic restraints are usually added to maintain protein secondary structure.[15,16] This is not necessarily a disadvantage in the context of cryo-EM data, since often the goal is to find a candidate structure and not to study the biased ensemble. One way of avoiding the difficulties of strong forces is to apply the bias virtually through reweighting after a simulation, assuming sufficient sampling is acquired.[17,18] This method may be formulated with a minimum entropy principle as well.[19]

EDM, the method introduced here, has the ability to match a statistical distribution while still being a minimal bias. The trade-off between quality of fit and perturbation of the ensemble is optimal. This is not only important for experimental data from scattering experiments[20] and imaging experiments[21] but also data with uncertainty. In contrast to restraint or biasing methods, EDM is built from a free energy method, metadynamics, and indirectly uses an estimate of the PMF along a collective variable to morph the PMF to the chosen distribution. EDM morphs the PMF, whereas traditionally biasing methods create a superposition of the bias and PMF. Figure 1 provides a visual representation of the contrast between EDM and the external biasing field methods. EDM achieves an exact, convergent match with a target PMF; there is no superposition. In addition to being minimal and exact, EDM recovers an estimated unbiased PMF and the amount of energy added to the system to match the targeted PMF. This extra information gives a quantitative measure of how well the original model fits to experimental data and may be of use in distinguishing between force fields or even molecular structures. In contrast to methods relying on postprocessing,[17−19] EDM will sample directly from the minimally biased ensemble, thus avoiding the difficulties in reweighting a poorly sampled ensemble.

EDM is a variant of metadynamics, which is an enhanced sampling technique designed to overcome energy barriers in molecular dynamics simulations through the use of a history-dependent bias.[1,22−24] The metadynamics method creates a bias that is applied on a set of collective variables. These collective variables describe the phenomenon of interest in the simulation. Similar to the Wang−Landau histogramming algorithm, metadynamics gradually flattens a simulation until the effective free energy (the underlying free energy plus the bias) of the simulation is flat along the chosen set of collective variables.[25−27] The purpose of a flat effective free energy is to allow perfectly diffusive sampling along the set of collective variables. A flat effective free energy has no energetic barriers. In a recent article by Dama et al.,[28] it was shown that certain variants of metadynamics can be made to asymptotically converge to not just a flat distribution but to nearly any normalized distribution subject to a few mild conditions. The flat distribution provides the ideal sampling assuming diffusion is homogeneous[29,30] and thus other distributions have been relatively unexplored, aside from as a tool to focus sampling in different areas[31] or to optimize for inhomogeneous diffusion.[29,30] In this work, we describe how EDM may be formulated using this property of metadynamics. An overview of the relationship among effective free energy, bias energy, and system potential is shown in Figure 1.

There are three significant challenges that are addressed in this article to demonstrate the usefulness of EDM. The first challenge is to show under what conditions EDM is convergent and that it is a minimal bias. The second challenge is to demonstrate the method on model systems. One is the well-characterized alanine dipeptide in vacuum,[22,23] where we will show how the method is able to change the phi/psi dihedral PMF into one obtained from bioinformatics data. The other model system is a Lennard-Jones fluid, where we will demonstrate how it is possible to morph radial distribution functions using special region-specific metadynamics hill functions. This also requires reformulating the metadynamics algorithm to treat radial distribution functions directly, as opposed to past work[32] that used scalar descriptors of the radial

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

distribution functions. The last challenge is to show that EDM can be applied to large, practical systems and thus we apply it to a three-component electrolyte simulation.

The remaining sections of this article are organized as follows: the Theory section covers the convergence, end states, proof of minimal biasing of EDM, and a description of region-specific metadynamics hills. In the Results, EDM is applied to four model systems: a single particle in an idealized potential, alanine dipeptide in vacuum, a Lennard-Jones fluid, and a three-component electrolyte solution. In the Discussion, we outline what can and cannot be accomplished with EDM and compare it to other biasing methods. Summarizing remarks are given in the Conclusions.

THEORY

Assume that we are sampling from a constant NVT ensemble and wish to change the PMF $F(s)$ along a collective variable $s(\vec{r})$, which is a function that turns the position of particles into a vector or scalar. Examples of collective variables are the position of a single particle, the distance vector between two particles, and the radius of gyration of a group of particles. We may write the PMF constraint as

$$\int \mathrm{d}\vec{r}\,\delta\,(s(\vec{r}) - s')\,P(\vec{r}) = Q(s') \tag{1}$$

where $Q(s) = -Tk_\mathrm{B}\ln[F_\mathrm{t}(s)]$, $F_\mathrm{t}(s)$ is the target PMF, k_B is Boltzmann's constant, T is temperature, $\delta(\cdot)$ is the Dirac delta function, $P(\vec{r})$ is the probability distribution of the biased system, and the integral is understood to be over all position dimensions of the system. Following the entropy maximization procedure, as seen for example in Roux and Weare,[7] we wish to find an ensemble that has the lowest relative entropy to the unbiased ensemble but matches the constraint in eq 1. The relative entropy of the biased ensemble to the unbiased ensemble is

$$\Delta S_\mathrm{rel} = \int \mathrm{d}\vec{r}\;P(\vec{r})\ln\frac{P(\vec{r})}{P_\mathrm{O}(\vec{r})} \tag{2}$$

where $P_\mathrm{O}(\vec{r})$ is the unbiased ensemble probability distribution. This is a constrained optimization problem and can be solved via the method of Lagrange multipliers. The Lagrange functional is

$$\Lambda[P(\vec{r}),\,\lambda(s')] = \int \mathrm{d}\vec{r}\;P(\vec{r})\ln\frac{P(\vec{r})}{P_\mathrm{O}(\vec{r})} - \lambda(s') \times\left(\int \mathrm{d}\vec{r}\;\delta(s(\vec{r}) - s')\,P(\vec{r}) - Q(s')\right) \tag{3}$$

Taking a functional derivative with respect to $P(\vec{r})$ and finding its root yields the following expression for $P(\vec{r})$

$$P(\vec{r}) \propto P_\mathrm{O}(\vec{r})\,\mathrm{e}^{\lambda(s)} \propto \mathrm{e}^{(-U(\vec{r})/k_\mathrm{B}T)+\lambda(s)} \tag{4}$$

where normalization is omitted. This form of the biased ensemble indicates the bias must be additive with the potential energy and act only in the dimension of the target PMF. Although this is intuitive, it says that modifying force field parameters, which inadvertently create potential energy changes in dimensions other than the constraint dimension, creates a nonminimal bias. The Supporting Information contains more details and a second independent proof showing that if additional biasing dimensions are added to eq 4, then the relative entropy will increase. We also show in the Supporting Information that the bias in eq 4 is unique, provided there is sufficient sampling in the collective variable dimension and $Q(s) > 0$ (positive definite).

According to eq 4, satisfying eq 1 by biasing only in the collective variable dimension is sufficient for creating a unique, minimally biased ensemble that matches the target PMF. This can be done with a variant of metadynamics. Metadynamics is a history-dependent biasing procedure that changes the potential energy of the simulation according to the following update[1,23]

$$V(s,n+1) = V(s,n) + G(s,s_{n+1}) \tag{5}$$

where $V(s,n)$ is the bias added to the potential energy as a function of the collective variable s at step n. s_n is the observed value of the collective variable at step n and $G(s,s')$ is a suitable hill function. The hill function $G(s,s')$ is usually taken to be a Gaussian function: $G(s,s') = h\,\exp[-(s - s')^2/w]$, where w controls the length scale of the update process and h controls the strength of the bias. This generalizes to multidimensional Gaussians for a vector collective variable. The convergence result of Dama et al.[28] requires $G(s,s')$ to be positive semidefinite, and there must exist a function, $p(s)$, such that the equation $\int \mathrm{d}s'\;G(s,s')\,p(s) = C$ holds. A Gaussian function has this property with $p(s) = 1$. Although the update procedure of eq 5 does not converge,[33,34] one stationary point is when $V(s,n) = -F(s)$, meaning that we can inspect the bias added to the system to find the PMF. To see this is a stationary point, examine the expected amount of bias to be added at time n

$$E[\Delta V(s,n)] = \int \mathrm{d}s_{n+1}\;\mathrm{e}^{-[F(s_{n+1})+V(s_{n+1},n)]/k_\mathrm{B}T}G(s,s_{n+1}) = \int \mathrm{d}s_{n+1}\;G(s,s_{n+1}) = C \tag{6}$$

where we have used the current PMF of the simulation, $V(s,n) + F(s)$, to calculate the expectation and the last step on the right-hand side is by the property of the hill function (in this case, Gaussians). Since the expected change in bias is independent of the collective variable, there is a constant increase in energy everywhere, which does not change the PMF or the forces driving the system along the collective variable. Thus, this is a stationary point in bias update procedure.

Consider an alternative update procedure:

$$V(s,n+1) = V(s,n) + \mathrm{e}^{F_\mathrm{t}(s_{n+1})/k_\mathrm{B}T}\,G(s,s_{n+1}) \tag{7}$$

where $F_\mathrm{t}(s)$ is a target PMF. The stationary point becomes

$$F_\mathrm{t}(s) = V(s,n) + F(s) \tag{8}$$

The current PMF of the simulation at the stationary point is $F_\mathrm{t}(s)$, and we have morphed the PMF into the target. This update procedure does not converge, and the introduction of a tempering factor is required. The most commonly used tempering factor is $\mathrm{e}^{-V(s,n)/k_\mathrm{B}\Delta T}$,[22] which means the bias update at time n is more damped if there already exists bias at s_n. ΔT is an adjustable parameter that controls the rate of tempering. The update procedure with a target PMF becomes

$$V(s,n+1) = V(s,n) + \mathrm{e}^{-V(s_{n+1},n)/k_\mathrm{B}\Delta T + F_\mathrm{t}(s_{n+1})/k_\mathrm{B}T}\,G(s,s_{n+1}) \tag{9}$$

Following the stationary point argument above, we see that $F_\mathrm{t}(s) = [(T + \Delta T)/\Delta T]V(s,n) + F(s)$. Instead of arriving at $F_\mathrm{t}(s)$, the stationary PMF is $[\Delta T/(\Delta T + T)]F_\mathrm{t}(s) + [T/(\Delta T + T)]F(s)$. The stationary PMF is a mixture of the desired PMF and the underling PMF of the unbiased system. The mixture

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

can be tuned by increasing or decreasing tempering. This is also the case with well-tempered metadynamics without targeting, where, instead of a flattening the PMF, the PMF is compressed depending on the rate of tempering (Figure 2a). This variant

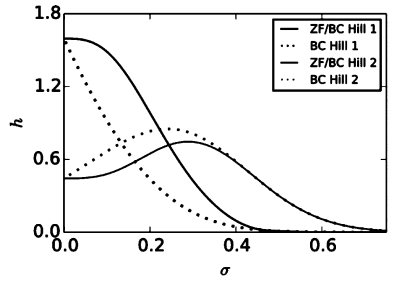


Figure 2. Comparison of Gaussian hills with the multiplicative McGovern–de Pablo boundary correction (BC) with the zeroed-force variant described in the Theory section (ZF/BC). Both are stationary hills, but the ZF/BC hills have no force at the boundary. Hill 1, the black lines, is centered at $\sigma = 0.0$, where σ is the hill width. Hill 2 is centered at $\sigma = 0.2$.

allows mixing of the target, and the PMF and may be viewed as sampling from a Bayesian posterior where the prior belief is the unbiased PMF of the system and the strength of that prior belief is controlled by ΔT (lower means stronger belief in the prior).

To remove the mixture in the well-tempered flavor of metadynamics, we must make the tempering independent of s_n. A simple solution is to follow the methods described in Dama et al. and use a global tempering with an optional threshold-based delay.[28,35] With global tempering, instead of using the local bias $V(s_n)$, we use an increasing term that is independent of s_n. We will use $\overline{V}_n$, the average bias over the collective variable domain, so that the choice of ΔT follows previous guidelines in literature. The update rule is

$$V(s,n+1) = V(s,n) + e^{-\overline{V}_n/k_B\Delta T}\, e^{F_t(s_{n+1})/k_B T}\, G(s,s_{n+1}) \tag{10}$$

This yields the following expected update

$$E[\Delta V(s)] = \int ds_{n+1}\, e^{-\overline{V}_n/k_B\Delta T}\, e^{F_t(s_{n+1})/k_B T}\, e^{-[F(s_{n+1})+V(s_{n+1},n)]/k_B T}\, G(s,s_{n+1}) \tag{11}$$

It is clear that the expectation of the average bias is constant; thus, we arrive at the simpler stationary point of traditional metadynamics in eq 8. After convergence, we have a simulation that evolves according to $F_t(s)$, and because we know the bias applied to achieve this, we can infer $F(s)$ using eq 8. This is the update rule of EDM and is used in all simulations.

Given a reference radial distribution function $h(r)$, EDM can morph the radial distribution of a system $g(r)$ into $h(r)$. $g(r)$ is the unbiased radial distribution function that could be calculated without biasing. Take $g(r,n)$ to be the radial distribution function at step n. The collective variable s in this calculation is the pairwise distance between particles, which at each time takes as many values as there are particle pairs. The bias update rule is

$$V(s,n+1) = V(s,n) + \frac{1}{M}\sum_{i<j}\frac{G(s,s_{n+1}^{ij})}{(s_{n+1}^{ij})^2\, h(s_{n+1}^{ij})} \tag{12}$$

where M is the number of pairs and s_n^{ij} is the distance between particle i and j at step n. Again, following Dama et al., this update procedure converges when the expectation of successive updates is a constant. That may be written as

$$E[\Delta V(s,n)] = \frac{1}{M}M\frac{4\pi}{Q}\int dr\; G(s,r)\,\frac{r^2\, g(r,n)}{r^2 h(r)} = C \tag{13}$$

where Q is the system volume and $M4\pi r^2 g(r,n)/Q$ is the probability of the sum of the pairwise distances. One way for this equation to be constant is if $g(r,n) = h(r)$. Thus, there exists a stationary point in the targeted metadynamics procedure that yields a potential $V(r) + U(r)$ that corresponds to $h(r)$. The existence of other stationary points is ruled out by the Henderson uniqueness theorem for pair-correlation functions.[36] The proof of a stationary point can be extended to a proof of convergence with the introduction of a global tempering factor

$$V(s,n+1) = V(s,n) + e^{-\overline{V}_n/k_B\Delta T}\frac{1}{M}\sum_{i<j}\frac{G(s,s_{n+1}^{ij})}{(s_{n+1}^{ij})^2\, h(s_{n+1}^{ij})} \tag{14}$$

which is the update rule for radial distribution functions in EDM. The convergence can be seen by noting that $G(s,s_{n+1}^{ij})$ is a valid hill function so that the methods of Dama et al. apply with $p_w = p_b\, h(r)/r^2$.[28]

It may occur that the target PMF may not span the domain that the simulation may explore. For example, one may have a radial distribution function from ab initio data that does not extend as far as a classical simulation radial distribution function. This may be resolved by using a hill function that has support only on the interval of the target PMF. Such hills were introduced by Crespo for one dimension[37] and extended to multiple dimensions by McGovern and de Pablo[38] for the case of a simulation with hard boundaries. The latter authors showed that a normal hill function may be restricted to a finite domain by using a multiplicative correction. The traditional Gaussians, however, will exert a nonzero force at the boundary even when the multiplicative boundary correction is used. The boundary force may be set to zero by following the reflection strategy of Crespo or by introducing a region that smoothly varies from Gaussians to a constant

$$G_b(s,s') = G(s,s') + [G(s',B) - G(s,s')] \times [d(s,B)^3 - 3d(s,B)^2 + 1] \tag{15}$$

where B is the boundary and $d(s,B)$ is the distance to the nearest boundary. Equation 15 should be divided by its integral over s' in order to be stationary (the McGovern–de Pablo multiplicative correction). Figure 2 shows a comparison of hills generated via eq 15 and normal Gaussian hills. Notice that the derivative, and hence force, is 0 at the boundary. The vertical lines indicate where the hill was added, and the left of the plot is a boundary. The hill equations expanded in 1 dimension and its derivatives are given in Equations S28–S32. A comparison of the results from these zeroed-force boundary-corrected hills vs boundary-corrected hills is given in Figure S1. Overall, we observe little difference between zeroed-force and normal hill types in their convergence and behavior aside from improved stability at the boundary; the zeroed-force hills simply have a smooth transition region to zero force, whereas normal hills abruptly change to zero force just past the biasing boundary.

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

METHODS

Targeted metadynamics is implemented as a fork of the PLUMED simulation engine plug-in[39] and as a LAMMPS[40] plugin for the radial distribution function EDM. The LAMMPS version makes use of neighbor lists and MPI to scale to larger systems well. A scaling plot may be seen in Figure S2. The version of PLUMED with this implementation and the LAMMPS plugin may be obtained by request from the authors. The implementation also includes an additional Python program that allows construction of target distributions in the PLUMED grid file format. Target distributions may be constructed in Python, as the result of a PLUMED simulation, a radial distribution function, or as PNG files in the case of 2D target distributions. Calculating free energies from a targeted metadynamics simulation requires removal of the bias introduced from the targeted distribution that is also done with the additional Python program. As described above, in the case of biasing radial distribution function, M hills are added at each time step (the summation in eq 14), which is time-consuming. Instead, in our implementation, hills are added stochastically according to an adjustable "density of hills" ξ. At each step, the probability of adding a hill is ξ/M, and its height is scaled by M/ξ, resulting in an expected equivalent hill height of 1 at each metadynamics step. The ΔT parameter described above is specified with a bias factor that is $(T + \Delta T)/T$.

The model one-dimensional cosine system was made up of soft-cosine potentials

$$U(x) = \sum U_i(x)$$
$$U_i(x) = \begin{cases} \cos\left(\frac{\pi(x - x_i)}{2}\right), & |x - x_i| < 2 \\ 0 \end{cases} \quad (16)$$

The x_i coefficients are 0, 1, 3, 3.2, 4, 5, 6.5, 6.5, 6.5, 9, and 10. A single particle was placed in this potential with unit mass and simulated from 0 to 10 length units with periodic boundary conditions. The system was evolved using a Nosé−Hoover thermostat[41,42] with a temperature of 1.0 and a time constant of 0.3 time units. The system was simulated with a time step of 0.01 time units for 1 million steps except for the bounded targeted simulation in Figure 3e, which was simulated for 12 million steps. EDM was used for all targeted simulations with an initial hill height of 0.02 energy units with the particle position being the collective variable. The parameter σ was 0.1 length units, and hills were deposited with a stride of 25 steps. The bias factor was 10 energy units. The simulations were conducted in the LAMMPS simulation engine.[40]

The unit mass Lennard-Jones fluids were simulated in the constant NVT ensemble with a Langevin thermostat[43] at a temperature of $1.5\varepsilon^*$ and a time constant of 1.0 $(\varepsilon^*/m\sigma^{*2})^{1/2}$ for 250 000 steps with a time step of 0.001 $(\varepsilon^*/m\sigma^{*2})^{1/2}$. Following the NVT ensemble simulation, radial distribution function data was collected for the last 100 000 steps in the constant NVE ensemble. The Lennard-Jones potential was used with a cutoff of $4\sigma^*$. The targeted simulation used EDM with a hill height of $1 \times 10^{-3}\varepsilon^*$, a width of $0.025\sigma^*$, a bias factor of 15, a stride of 25 steps, a threshold of $1\varepsilon^*$, and a density of hills of 0.25. The target was applied only from $r = 0.28\sigma^*$ to $3.0\sigma^*$ with the use of the previously described boundary-corrected hills.

The alanine dipeptide in vacuum system was simulated using Gromacs 4.6.1[44] with the CHARMM27[45] force field in a constant NVT ensemble. The Bussi−Donadio−Parrinello velocity-rescaling thermostat was used with a time constant of 5 picoseconds.[46] The nonbonded forces were cutoff at 1 nm, and particle mesh Ewald summation was used for long-range electrostatics with a grid constructed according to an error of tolerance 1×10^{-5} kJ/mol.[47] All hydrogen-participating bonds were constrained using the LINCS algorithm.[48] The phi and psi dihedral angles were biased with a width of 0.35 radians. The parameters for the global-tempered metadynamics runs were a hill height of 0.2 kJ/mol, a bias factor of 10, and a stride of 120 fs. The parameters for EDM were the same as the global-tempered metadynamics run.

Complete details of the electrolyte simulation are given in White and Voth[8] and Jorn et al.[49] A total of 201 ethylene carbonate, 15 lithium ions, and 15 hexafluorophosphate ions (1 M $LiPF_6$) were simulated in a 3 nm cube at a temperature of 300 K. Radial distribution functions were calculated after a 2 ns constant NVT equilibration period for 2 ns in the constant NVE ensemble. The EDM parameters were a hill height of 0.01 kcal/mol, a stride of 1 ps, a bias factor of 5, a hill width of 0.025 Å, a threshold of 1 kcal/mol and a density of hills of 1. EDM was applied only on the interval 1.68 to 5.0 Å, and all radial distribution functions were normalized to that interval.

RESULTS

One-dimensional systems can be plotted simply and are intuitive to understand. Consider a single particle in a constant temperature periodic system with a potential energy shown in Figure 3a. The various panels in Figure 3 show a comparison of well-tempered metadynamics simulations[22] and EDM simulations. Figure 3a,c,e,f shows the effective PMF driving the single particle in a 1D potential. The blue area is the potential energy function, and the red area is the bias after running a metadynamics algorithm for 1 million steps. The sum of the blue and red areas is the effective PMF felt by the particle. Figure 3b,d demonstrates how to calculate the true PMF (blue area) using the bias applied. The target is shown as the green area, and the negative bias, as the red area. Figure 3a is the result of a well-tempered metadynamics simulation and shows the compressing property of well-tempered metadynamics, where the degree of compression is $T/(\Delta T + T)$ (eq 8). Since the compression is known, the PMF is recovered exactly as shown in Figure 3b. Traditional and well-tempered metadynamics may be thought of as EDM where the target is a horizontal line. By contrast, Figure 3c shows the effective PMF from an EDM simulation. The green curve is the target PMF and the effective PMF matches it. Using eq 8, Figure 3d shows that the true PMF of the system can be recovered. The green and red areas show that adding the negative of the bias with the targeted PMF yields the true PMF (dotted blue line). Figure 3e shows the system may target other 1D PMFs. Finally, Figure 3f shows that we can target partial PMFs by turning off EDM outside of an interval and using the appropriate boundary-corrected hills described in the Theory section. EDM effectively arrives at a targeted PMF while simultaneously recovering the true PMF of the system of interest.

A more realistic system, the alanine dipeptide in vacuum, is shown in Figure 4. The target PMF is derived from bioinformatics data[50] and will bias the simulation so that it samples the conformations of alanine in proteins. The applied bias can also show the discrepancy between the inherent PMF of alanine in vacuum and when alanine is a protein constituent. The alanine dipeptide in vacuum is an often studied system and

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

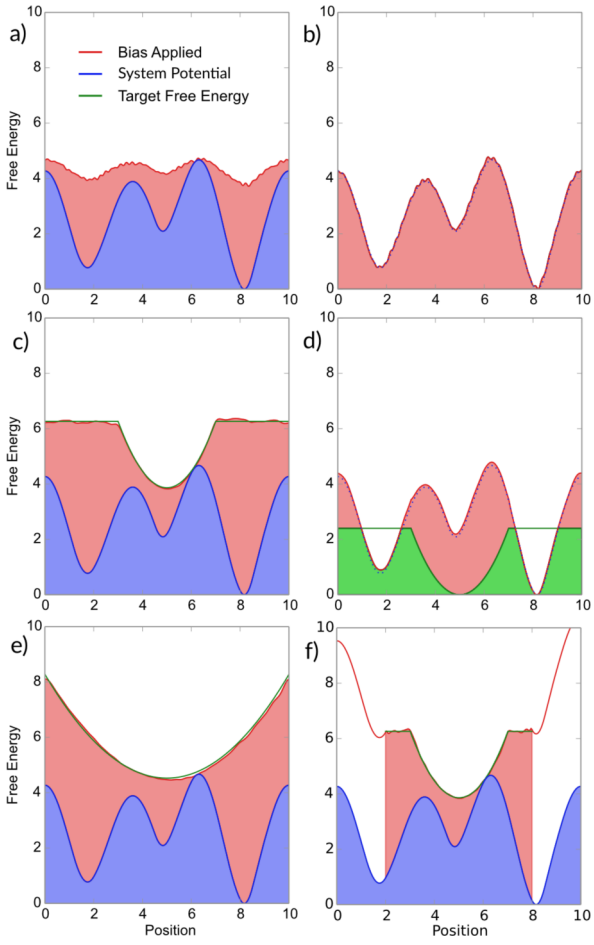


Figure 3. Comparison of targeted metadynamics (c–f) and well-tempered metadynamics (a, b). (a) Resultant effective PMF of a well-tempered metadynamics simulation, which yields the true PMF (b). (c) Resultant effective PMF of a targeted simulation, which matches the target PMF shown in green. (d) Calculation of the true PMF from the targeted bias. (e–f) Different target types possible. In (f), the red line above shows the effective PMF of the unbiased region, which is the same as the system potential. These model simulations are unitless.

has been well-characterized in previous free energy methods literature.[22,23,35] The free energy surface of the φ/ϕ angles of alanine dipeptide is shown in Figure 3a, as determined from a standard well-tempered metadynamics calculation with 15 ns of simulation time. EDM was used to create a minimally biased simulation that matches the PMF derived from bioinformatics shown in Figure 4c. The bioinformatics data is the conditioned triplet distribution for the φ/ϕ angles of alanine with glycines at its N- and C-terminus. This is a good approximation of the alanine dipeptide since glycines provide no steric hindrance. The bioinformatics data is from Ting and co-workers, where they considered 2,020 protein structures to create neighbor-dependent Ramachandran plots.[50] The probability distribution was modified, however, since the original data had probabilities separated by 25 orders of magnitude. The probabilities were rescaled so that their magnitude spanned the same orders of magnitude as that in the PMF in Figure 4a, about 10 orders of magnitude (60 kJ/mol). The EDM simulation resulted in the effective PMF shown in Figure 4d, which was derived via a histogram of the last 5 ns of the simulation. The effective PMF matches the target in Figure 4c and thus the simulation is evolved according to the bioinformatics PMF instead of the

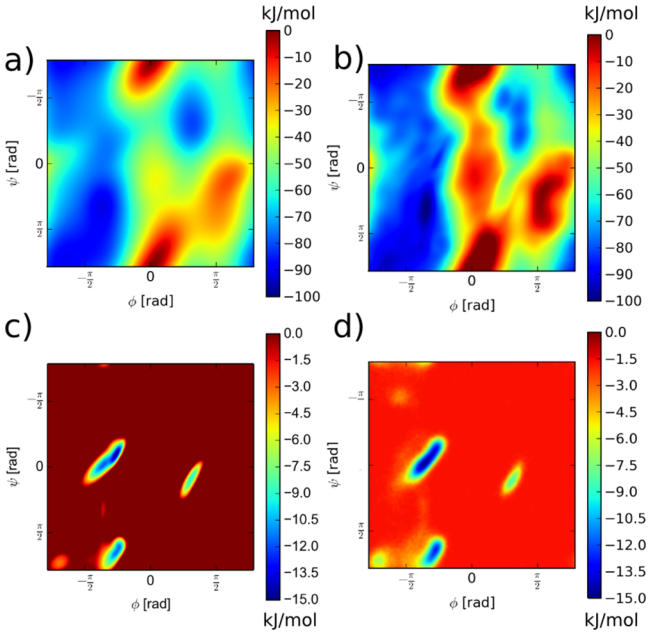


Figure 4. Results of well-tempered metadynamics and EDM on alanine dipeptide in vacuum. PMF calculated using (a) well-tempered metadynamics and (b) EDM. (c) Target PMF and (d) effective PMF of the EDM simulation.

vacuum PMF. As a point of interest, through eq 8 the untargeted PMF can be recovered and is shown in Figure 4b. It matches most of the features from the PMF of the well-tempered metadynamics run in Figure 4a. Not all features are recovered because the target PMF is causing the simulation to avoid its free energy peaks, resulting in poor sampling in the red regions of Figure 4c. These results demonstrate how to build a model by combining bioinformatics data with atomistic simulations.

One of the most important observables in molecular simulations is the radial distribution function. This is a transformation of the pair-probability density and can be used as a target probability distribution as shown in eq 14. An 864 particle Lennard-Jones system was simulated with parameters σ = 0.9, ε = 1.2 with a targeted radial distribution function generated from a Lennard-Jones system with σ = 1.0, ε = 1.0. The targeted radial distribution function and the radial distribution function of the unbiased σ = 0.9, ε = 1.2 are shown in Figure 5. The black dashed vertical line indicates the boundary of the targeted radial distribution; metadynamics was

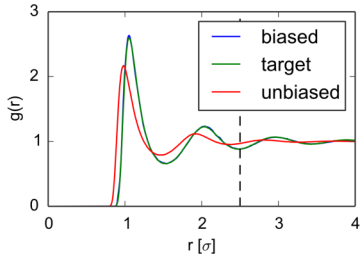


Figure 5. Targeted metadynamics of a Lennard-Jones fluid with 864 particles. The unbiased line is moving to the target radial distribution function. Note that the biased simulation line is entirely covered by the target line. The vertical dashed gray line indicates the boundary of the EDM simulation.

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX–XXX

not applied to the right of the line. Using boundary-corrected hills with EDM, the biased radial distribution closely matches the target radial distribution in Figure 5. One observation from these results is that the hill width should be less than the distance between the rise of the first peak and the top of the first peak of the radial distribution function. If the hill width is wider, then the bias oscillates between repulsion and attraction. This may be observed in Figure S3, which shows the same simulations as Figure 5 with a variety of hill width choices. These results demonstrate an important use of EDM in the targeting of radial distribution functions.

The final system considered is a three-component electrolyte system composed of ethylene carbonate, lithium ions, and hexafluorophosphate. The formation of a solid electrolyte interphase (SEI) at the cathode is one of the most important microscopic properties for the performance and durability of a lithium battery.[51,52] Characterizing the SEI experimentally is challenging,[53] and this is an area where simulations can shed insight into its structure and atomic details. This electrolyte solution contains the liquid components of the SEI. Following our earlier work,[8] we consider again the model of Jorn et al.,[49] which is a traditional molecular dynamics simulation with a classical two-body pair-potential. Jorn et al. parametrized a model that gives a radial distribution function quite different than their *ab initio* molecular dynamics simulations and the more complex polarizable force field from Borodin and Smith.[54] In ref 9, this model was "repaired" by biasing coordination numbers and moments with EDS. With EDM, it is instead possible to repair the radial distribution function directly.

The results from EDM simulation are shown in Figure 6. The targeted distribution is the lithium−ethylene carbonate carbonyl oxygen radial distribution function shown in Figure 6a. The target radial distribution function comes from the data of Jorn and et al.,[49] which overlaps well with the data of Borodin and Smith.[54] Thus, we also compare the diffusion data and the lithium−fluorine radial distribution function with their more detailed polarizable model. Indeed, after improvement of the lithium−ethylene carbonate carbonyl oxygen radial distribution function, there is improvement in the other properties, including the lithium self-diffusion in Figure 6c, which is improved further than that in ref 9. This connection between improvement of structural properties and dynamics by biasing a single radial distribution function was also seen in ref 9 and has been explored in past literature.[55,56] The lithium−fluorine radial distribution function did not improve in Figure 6, and an additional simulation with two biases was conducted with the same parameters, except that two independent biases were applied to the lithium−ethylene carbonate carbonyl oxygen radial distribution function and the lithium−fluorine radial distribution function. This simulation is farther away from the unbiased system since two biases are applied, but it is still the unique ensemble that is as close to the unbiased system while being consistent with both radial distribution functions. The results of that simulation may be found in Figure S4, where near exact agreement for both radial distribution functions is seen. The electrolyte simulation results provide a demonstration of how available data from more detailed methods or experiments can be used to improve models without iterative simulation.

The resulting bias obtained from EDM gives insight into how to modify potentials, since the resulting bias is the unique minimal modification to obtain a desired radial distribution function. An example of this is shown in Figure 7. Figure 7 shows the bias potential and the radial distribution function of the lithium−ethylene carbonate carbonyl oxygen. The largest amount of bias is in the repulsion area, which was responsible for moving the first rise of the radial distribution function from 1.68 to1.76 Å. This indicates that the Lennard-Jones radius of this interaction should be increased. Past the repulsion area, the bias gradually pushes away lithium ions to broaden and lower the first peak. It was observed that EDM with a cutoff of 2.5 Å was sufficient to reshape the radial distribution function, but the diffusion data was not improved (results not shown). This led to the choice of the 5 Å EDM cutoff. The requisite long-range interaction shows that some modification to the Lennard-Jones

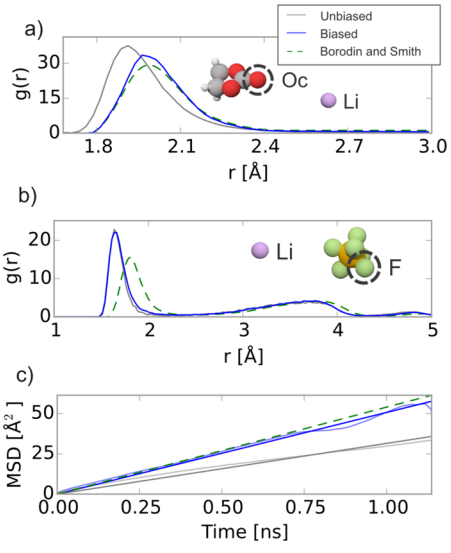


Figure 6. Simulation results from an electrolyte solution containing lithium, ethylene carbonate, and hexafluorophosphate. (a) Lithium−ethylene carbonate carbonyl oxygen radial distribution function, which was targeted to the Borodin and Smith reference data[54] shown in the dashed line green. (b) No improvement in the lithium−fluorine radial distribution function, since it was not biased, and (c) improvement in the mean-square displacement self-diffusion of the lithium ions. A doubly biased simulation may be seen in Figure S4, which shows good agreement with both radial distribution functions.

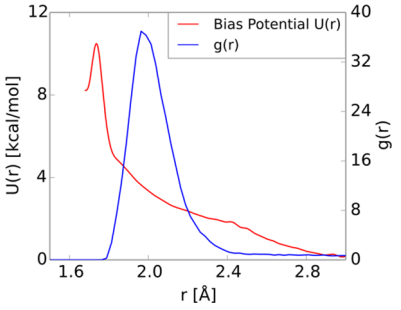


Figure 7. Lithium−ethylene carbonate carbonyl oxygen radial distribution function and EDM bias from the electrolyte simulation. The red line is the bias that caused the radial distribution function in blue to match the target Borodin and Smith data.[54] The bias has a strong repulsive core, a sloping section which lowers the peak of the radial distribution function, and a long-range repulsion. It is not attractive in any region.

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

energy parameter or partial charge of the two atoms would be required to affect the same change as EDM.

DISCUSSION

EDM is *not* a new method for improving convergence or accuracy of free energy calculations. It is a method for choosing a free energy surface of a system. It has three advantages over the traditional addition of a harmonic restraint. The first is that EDM converges exactly to the target distribution. This is possible only by calculating the free energy first, nullifying it, and replacing it with the target free energy. This leads to the second advantage: the free energy of the unbiased/untargeted system becomes known in the process of EDM metadynamics. Finally, knowing the free energy of both the untargeted and targeted systems means that it is possible to calculate how much energy was added to achieve the target free energy. This yields a quantitative thermodynamic measure of how close different force fields or structures are to a target free energy or experimental value.

EDM does not require exhaustive sampling of the chosen collective variable in order to correctly modify the PMF. The sampling of EDM will follow the target PMF so that the system will not cross high-energy barriers present in the target PMF. It is possible to choose collective variables that have very long autocorrelation times even after biasing, have slow orthogonal collective variables, have regions with zero-derivatives (ie, no force) in the collective variables, or have infinite energy barriers between regions. Such pathological collective variables may not be morphed by EDM to the target PMF due to how slowly they are sampled. Choosing good collective variables is outside the scope of this work, but it has been covered extensively in the free energy sampling literature.[23]

The difference between EDM and experiment directed simulation[8] (EDS) is that the first is for modifying whole probability distributions or PMFs and EDS is for modifying expectation values with minimal changes to the statistical ensemble. For example, if we want to set the average distance between two amino acids on different proteins but do not know their distribution, then EDS would provide a minimal bias that disturbs the entropy of the system least. As another example, consider constructing a coarse-grained single-site model of a benzene molecule. If the model does not have the correct radial distribution function, then EDM would be best used to modify until it is correct. After the simulation, the EDM bias could be added to the pair-potential, resulting in a new pair-potential that gives the correct radial distribution function. EDM applied to a radial distribution is similar to iterative Boltzmann inversion (IBI),[57] which is a widely used method for modifying a force field to match a target radial distribution function. In IBI, a simulation is run with a candidate force field and the potential is updated according to

$$\Delta V(r,n) = \alpha k_B T \ln[g(r,n)/h(r)] \tag{17}$$

where $V(r)$ is the pairwise potential and α is an adjustable parameter. Following that notation, the EDM rule may be written as

$$\mathrm{E}[\Delta V(r,n)] = \mathrm{e}^{-\bar{V}_n/k_B\Delta T}\frac{4\pi}{Q}\int \mathrm{d}r'\, G(r,r')\,\frac{g(r,n)}{h(r)} \tag{18}$$

Aside from the prefactor, the main difference is that the EDM rule uses the ratio between the target and current radial distribution functions directly and it smoothens it using the hill function. The other benefit is that the method converges through global tempering and requires a single simulation, whereas IBI requires iterative simulation and updates of the potential. Finally, it is worth noting that this connection to IBI also corresponds to a deeper connection between EDS and EDM. IBI is a gradient descent procedure, whereas EDM is a stochastic version of it. If not for the hill function $G(r,r')$, then EDM would simply be a stochastic gradient descent version of IBI and would therefore be exactly analogous to EDS. The hill function serves to regularize the gradient estimates, showing that EDM is a very natural relative of EDS.

CONCLUSIONS

Experiment directed metadynamics (EDM) is a new method to choose the probability distribution, also known as potential of mean force (PMF), of collective variables in a molecular simulation. EDM finds the unique, minimal bias so that the original properties of the molecular simulation are as unchanged as possible. The ability to design the PMF of collective variables for a system provides a direct method for improving molecular models with ab initio, bioinformatics, or experimental data. EDM simultaneously estimates and morphs the PMF so that the resulting ensemble has no mixing between the unbiased PMF and target PMF. This is a property shared with other biasing techniques such as harmonic restraint ensembles and EDS.[8] EDM provides two additional quantities: the PMF of the unbiased system and the amount of energy required to achieve the target PMF. The PMF of the unbiased system is usually of interest, and energy required to reach the target PMF provides a thermodynamic measure for comparing experimental data, force fields, or even different molecular structures. EDM has been shown to converge for the general case of multidimensional collective variables and radial distribution functions. EDM has also been shown to be the unique minimal bias for matching a target PMF. Compared to EDS, the EDM method is best when an entire PMF is known and experiment directed simulation is best when average values are known. EDM has been implemented and tested in Gromacs 4.6.1,[44] LAMMPS June 2014,[40] and CP2K 2.5[58] via a forked version of the PLUMED 1.3 plugin.[39] A second parallelized implementation is available in LAMMPS specifically for biasing radial distribution functions. Both are available via request from the authors.

ASSOCIATED CONTENT

Supporting Information

A more detailed derivation of eq 4, a proof that adding biasing dimensions beyond eq 4 increases relative entropy, the conditions for uniqueness, and boundary-corrected and region-specific hill equations. The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00178.

AUTHOR INFORMATION

Corresponding Author

*E-mail: gavoth@uchicago.edu.

Funding

We gratefully acknowledge the Office of Naval Research (ONR award N00014-13-1-0058) and the support of the U.S. Department of Energy through the LANL/LDRD Program for this work. G.A.V. also thanks the Los Alamos National Laboratory (LANL) Center for Nonlinear Studies (CNLS) for

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

a Stanislaw M. Ulam Distinguished Scholar Award in 2014. A.D.W. and J.F.D. were visitors to the CNLS during this period. A.D.W. was supported in part by a Yen Fellowship from the University of Chicago Institute for Biophysical Dynamics.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge the University of Chicago Research Computing Center for providing computing resources for this research.

■ REFERENCES

(1) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(2) Best, R. B.; Vendruscolo, M. Determination of protein structures consistent with NMR order parameters. *J. Am. Chem. Soc.* **2004**, *126*, 8090−8091.
(3) Islam, S. M.; Stein, R. A.; Mchaourab, H. S.; Roux, B. Structural refinement from restrained-ensemble simulations based on EPR/DEER data: application to T4 lysozyme. *J. Phys. Chem. B* **2013**, *117*, 4740−4754.
(4) De Simone, A.; Montalvao, R. W.; Dobson, C. M.; Vendruscolo, M. Characterization of the interdomain motions in hen lysozyme using residual dipolar couplings as replica-averaged structural restraints in molecular dynamics simulations. *Biochemistry* **2013**, *52*, 6480−6486.
(5) Lindorff-Larsen, K.; Best, R. B.; DePristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005**, *433*, 128−132.
(6) Pitera, J. W.; Chodera, J. D. On the use of experimental observations to bias simulated ensembles. *J. Chem. Theory Comput.* **2012**, *8*, 3445−3451.
(7) Roux, B.; Weare, J. On the statistical equivalence of restrained-ensemble simulations with the maximum entropy method. *J. Chem. Phys.* **2013**, *138*, 084107.
(8) White, A. D.; Voth, G. A. Efficient and minimal method to bias molecular simulations with experimental data. *J. Chem. Theory Comput.* **2014**, *10*, 3023−3030.
(9) Boomsma, W.; Ferkinghoff-Borg, J.; Lindorff-Larsen, K. Combining experiments and simulations using the maximum entropy principle. *PLoS Comput. Biol.* **2014**, *10*, e1003406.
(10) Frank, A. T.; Law, S. M.; Brooks, C. L., III A simple and fast approach for predicting ^{1}H and ^{13}C chemical shifts: toward chemical shift-guided simulations of RNA. *J. Phys. Chem. B* **2014**, *118*, 12168−75.
(11) Duchi, J.; Hazan, E.; Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res.* **2011**, *12*, 2121−2159.
(12) Brunger, A. T.; Nilges, M. Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Q. Rev. Biophys.* **1993**, *26*, 49−125.
(13) Dolenc, J.; Missimer, J. H.; Steinmetz, M. O.; van Gunsteren, W. F. Methods of NMR structure refinement: molecular dynamics simulations improve the agreement with measured NMR data of a C-terminal peptide of GCN4-p1. *J. Biomol. NMR* **2010**, *47*, 221−235.
(14) Trabuco, L. G.; Villa, E.; Schreiner, E.; Harrison, C. B.; Schulten, K. Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods* **2009**, *49*, 174−180.
(15) Chan, K. Y.; Trabuco, L. G.; Schreiner, E.; Schulten, K. Cryo-electron microscopy modeling by the molecular dynamics flexible fitting method. *Biopolymers* **2012**, *97*, 678−686.
(16) Chan, K. Y.; Gumbart, J.; McGreevy, R.; Watermeyer, J. M.; Sewell, B. T.; Schulten, K. Symmetry-restrained flexible fitting for symmetric EM Maps. *Structure* **2011**, *19*, 1211−1218.
(17) Olsson, S.; Frellsen, J.; Boomsma, W.; Mardia, K. V.; Hamelryck, T. Inference of structure ensembles of flexible biomolecules from sparse, averaged Data. *PLoS One* **2013**, *8*, e79439.
(18) Rozycki, B.; Kim, Y. C.; Hummer, G. SAXS ensemble refinement of ESCRT-III CHMP3 conformational transitions. *Structure* **2011**, *19*, 109−116.
(19) Beauchamp, K. A.; Pande, V. S.; Das, R. Bayesian energy landscape tilting: towards concordant models of molecular ensembles. *Biophys. J.* **2014**, *106*, 1381−1390.
(20) Skinner, L. B.; Huang, C. C.; Schlesinger, D.; Pettersson, L. G. M.; Nilsson, A.; Benmore, C. J. Benchmark oxygen-oxygen pair-distribution function of ambient water from X-ray diffraction measurements with a wide Q-range. *J. Chem. Phys.* **2013**, *138*, 074506.
(21) Sousa, D. R.; Stagg, S. M.; Stroupe, M. E. Cryo-EM structures of the actin:tropomyosin filament reveal the mechanism for the transition from C- to M-state. *J. Mol. Biol.* **2013**, *425*, 4544−4555.
(22) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 8.
(23) Barducci, A.; Bonomi, M.; Parrinello, M. Metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2011**, *1*, 826−843.
(24) Sutto, L.; Marsili, S.; Gervasio, F. L. New advances in metadynamics. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 771−779.
(25) Wang, F. G.; Landau, D. P. Efficient, multiple-range random walk algorithm to calculate the density of states. *Phys. Rev. Lett.* **2001**, *86*, 2050−2053.
(26) Wang, F.; Landau, D. P. Determining the density of states for classical statistical models: a random walk algorithm to produce a flat histogram. *Phys. Rev. E* **2001**, *64*, 16.
(27) Junghans, C.; Perez, D.; Vogel, T. Molecular dynamics in the multicanonical ensemble: equivalence of Wang−Landau sampling, statistical temperature molecular dynamics, and metadynamics. *J. Chem. Theory Comput.* **2014**, *10*, 1843−1847.
(28) Dama, J. F.; Parrinello, M.; Voth, G. A. Well-tempered metadynamics converges asymptotically. *Phys. Rev. Lett.* **2014**, *112*, 6.
(29) Trebst, S.; Huse, D. A.; Troyer, M. Optimizing the ensemble for equilibration in broad-histogram Monte Carlo simulations. *Phys. Rev. E* **2004**, *70*, 046701.
(30) Singh, S.; Chiu, C.-c.; de Pablo, J. Flux tempered metadynamics. *J. Stat. Phys.* **2011**, *145*, 932−945.
(31) Lindahl, V.; Lidmar, J.; Hess, B. Accelerated weight histogram method for exploring free energy landscapes. *J. Chem. Phys.* **2014**, *141*, 044110.
(32) Tribello, G. A.; Cuny, J.; Eshet, H.; Parrinello, M. Exploring the free energy surfaces of clusters using reconnaissance metadynamics. *J. Chem. Phys.* **2011**, *135*, 114109 .
(33) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. Assessing the accuracy of metadynamics. *J. Phys. Chem. B* **2005**, *109*, 6714−6721.
(34) Bussi, G.; Laio, A.; Parrinello, M. Equilibrium free energies from nonequilibrium metadynamics. *Phys. Rev. Lett.* **2006**, *96*, 090601 .
(35) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. Transition-tempered metadynamics: robust, convergent metadynamics via on-the-fly transition barrier estimation. *J. Chem. Theory Comput.* **2014**, *10*, 3626−3633.
(36) Henderson, R. L. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A* **1974**, *49*, 197−198.
(37) Crespo, Y.; Marinelli, F.; Pietrucci, F.; Laio, A. Metadynamics convergence law in a multidimensional system. *Phys. Rev. E* **2010**, *81*, 4.
(38) McGovern, M.; de Pablo, J. A boundary correction algorithm for metadynamics in multiple dimensions. *J. Chem. Phys.* **2013**, *139*, 084102 .
(39) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R. A.; Parrinello, M. PLUMED: a portable plugin for free-energy calculations with molecular dynamics. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

(40) Plimpton, S. Fast parallel algorithms for short-range molecular dynamics. *J. Comput. Phys.* **1995**, *117*, 1−19.
(41) Hoover, W. G. Canonical dynamics—equilibrium phase-space distributions. *Phys. Rev. A* **1985**, *31*, 1695−1697.
(42) Shinoda, W.; Shiga, M.; Mikami, M. Rapid estimation of elastic constants by molecular dynamics simulation under constant stress. *Phys. Rev. B* **2004**, *69*, 8.
(43) Schneider, T.; Stoll, E. Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* **1978**, *17*, 1302−1322.
(44) Berendsen, H. J. C.; Vanderspoel, D.; Vandrunen, R. Gromacs—a message-passing parallel molecular-dynamics implementation. *Comput. Phys. Commun.* **1995**, *91*, 43−56.
(45) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.
(46) Bussi, G.; Donadio, D.; Parrinello, M. Canonical sampling through velocity rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.
(47) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G.; Smooth, A. Particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577−8593.
(48) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. LINCS: a linear constraint solver for molecular simulations. *J. Comput. Chem.* **1997**, *18*, 1463−1472.
(49) Jorn, R.; Kumar, R.; Abraham, D. P.; Voth, G. A. Atomistic modeling of the electrode−electrolyte interface in Li-ion energy storage systems: electrolyte structuring. *J. Phys. Chem. C* **2013**, *117*, 3747−3761.
(50) Ting, D.; Wang, G. L.; Shapovalov, M.; Mitra, R.; Jordan, M. I.; Dunbrack, R. L. Neighbor-dependent Ramachandran probability distributions of amino acids developed from a hierarchical Dirichlet process model. *PLoS Comput. Biol.* **2010**, *6*, e1000763.
(51) Verma, P.; Maire, P.; Novak, P. A review of the features and analyses of the solid electrolyte interphase in Li-ion batteries. *Electrochim. Acta* **2010**, *55*, 6332−6341.
(52) Xu, K.; von Cresce, A. Interfacing electrolytes with electrodes in Li ion batteries. *J. Mater. Chem.* **2011**, *21*, 9849−9864.
(53) Winter, M. The solid electrolyte interphase—the most important and the least understood solid electrolyte in rechargeable Li batteries. *Z. Phys. Chem.* **2009**, *223*, 1395−1406.
(54) Borodin, O.; Smith, G. D. Quantum chemistry and molecular dynamics simulation study of dimethyl carbonate: ethylene carbonate electrolytes doped with $LiPF_6$. *J. Phys. Chem. B* **2009**, *113*, 1763−1776.
(55) Bretonnet, J. L. Self-diffusion coefficient of dense fluids from the pair correlation function. *J. Chem. Phys.* **2002**, *117*, 9370−9373.
(56) Laghaei, R.; Nasrabad, A. E.; Eu, B. C. Pair correlation functions and the self-diffusion coefficient of Lennard-Jones liquid in the modified free volume theory of diffusion. *J. Phys. Chem. B* **2005**, *109*, 21375−21379.
(57) Reith, D.; Putz, M.; Muller-Plathe, F. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.* **2003**, *24*, 1624−1636.
(58) VandeVondele, J.; Krack, M.; Mohamed, F.; Parrinello, M.; Chassaing, T.; Hutter, J. QUICKSTEP: fast and accurate density functional calculations using a mixed Gaussian and plane waves approach. *Comput. Phys. Commun.* **2005**, *167*, 103−128.

DOI: 10.1021/acs.jctc.5b00178
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX