JCTC Journal of Chemical Theory and Computation

Letter

pubs.acs.org/JCTC

Efficient Sampling of High-Dimensional Free-Energy Landscapes with Parallel Bias Metadynamics

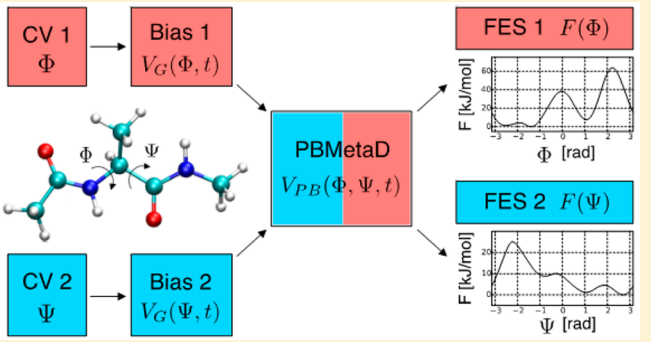
Jim Pfaendtner[†] and Massimiliano Bonomi*[,‡]

[†]Department of Chemical Engineering, University of Washington, Seattle, Washington 98195, United States

[‡]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

Supporting Information

ABSTRACT: Metadynamics accelerates sampling of molecular dynamics while reconstructing thermodynamic properties of selected descriptors of the system. Its main practical difficulty originates from the compromise between keeping the number of descriptors small for efficiently exploring their multidimensional free-energy landscape and biasing all of the slow motions of a process. Here we illustrate on a model system and on the tryptophan-cage miniprotein parallel bias metadynamics, a method that overcomes this issue by simultaneously applying multiple low-dimensional bias potentials.



Molecular dynamics (MD) simulations have played in recent years a fundamental role in molecular biology as well as in many other fields of science by providing insights into the structure and dynamics of complex systems at the atomistic level.[1] Unfortunately, exhaustive sampling by standard MD is in most cases computationally prohibitive because realistic models of biomolecules require introducing large numbers of particles and accurate interactions among them, which result in complex free-energy landscapes. Recently, the development of dedicated hardware[2] and distributed computing protocols[3] has alleviated some of these issues. However, typical time scales accessible to MD remain significantly shorter than those of interesting biomolecular processes and of many experimental techniques.

To extend sampling capabilities of MD simulations, several advanced methods have been proposed.[4,5] Metadynamics[6] (MetaD) is one of the most popular in the family of approaches based on the introduction of an external bias potential to accelerate sampling along selected coarse-grained descriptors of the system or collective variables (CVs). The MetaD bias potential is constructed adaptively during the simulation in order to discourage visiting regions of the CV space already explored. Additionally, the bias potential provides an estimate of the free energy as a function of the selected CVs once the simulation has converged. MetaD has been successfully applied to a wide range of problems in many areas of chemistry, physics, and biology, and its development is still actively in progress years after its original formulation.[7]

One of the main difficulties in MetaD lies in the choice of the CVs. Two conflicting requirements make this a frustrating task. On one hand, the set of CVs should include all of the slow modes of the systems, i.e., all those degrees of freedom that cannot be satisfactorily sampled in the time scale of the simulation and thus benefit most from the MetaD bias potential. On the other hand, the number of CVs should be kept as small as possible; otherwise a high-dimensional space has to be explored, and the efficiency of the exploration decreases exponentially with the number of CVs. To alleviate the problem of optimally choosing the CVs, several strategies have been proposed. Many of these are based on the combination of MetaD with replica exchange methods[8,9] (REM), such as the bias exchange MetaD[10] (BEM), the well-tempered ensemble,[11,12] and concurrent MetaD,[13] while others rely on complex CVs inspired by different dimensionality reductions tecniques.[14−16]

Here we introduce parallel bias metadynamics (PBMetaD), an enhanced sampling method that augments MD sampling by allowing the high barriers that characterize multidimensional free-energy profiles to be crossed in the time scale of typical simulations. This is achieved in a computationally efficient way compared to standard MetaD by simultaneuosly applying multiple parallel bias potentials in low dimensionality. The individual low-dimensional free energies are calculated directly from the bias potentials, and the full high-dimensional free energy can be easily recovered by standard reweighting techniques. We illustrate this method and compare it with BEM on a simple model system and on the 20-residue tryptophan-cage protein[17] (TRPC) simulated in implicit solvent. In the two cases, PBMetaD performed better or equally well compared to BEM, while always offering a greater flexibility and simplicity of use. In particular, in PBMetaD the number of replicas of the systems can be chosen independently from the number of CVs that need to be biased; we avoided the

Received: September 3, 2015

ACS Publications © XXXX American Chemical Society

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

need to introduce an additional parameter, the BEM exchange stride; and we simplified the analysis of the convergence and the reconstruction of high-dimensional free-energy landscapes.

First, we quickly review the theory of metadynamics. The MetaD history-dependent bias potential (V_G) is constructed in the space of a few selected CVs, which are functions S of the microscopic coordinates $\boldsymbol{R}$ of the system $S(\boldsymbol{R}) = (S_1(\boldsymbol{R}), ..., S_d(\boldsymbol{R}))$, as a sum of Gaussians deposited along the system trajectory in the CVs space. In the popular well-tempered variant[18] (WTMetaD), V_G is expressed at time t in terms of the histogram $N(S,t)$ collected during the biased simulation:

$$V_G(S, t) = k_B\Delta T \log\left(1 + \frac{\omega N(S, t)}{k_B\Delta T}\right) \tag{1}$$

where k_B is the Boltzmann constant and ω and ΔT are input parameters with the dimensions of an energy rate and a temperature, respectively. Equivalently, V_G can be written as a sum of Gaussians:

$$V_G(S, t) = \int_0^t dt'\, \omega(t') \exp\left(-\sum_{i=1}^{d} \frac{(S_i(\mathbf{R}) - S_i(\mathbf{R}(t')))^2}{2\sigma_i^2}\right) \tag{2}$$

where σ_i is the Gaussian width of the i-th CV, and, at variance with standard MetaD, the deposition rate decreases with simulation time:

$$\omega(t) = \omega \exp\left(-\frac{V_G(S, t)}{k_B\Delta T}\right) \tag{3}$$

In the long time limit, V_G converges to

$$V_G(S, t \to \infty) = -\frac{\Delta T}{T + \Delta T} F(S) + C \tag{4}$$

where T is the temperature of the system, $F(S)$ is the free energy as a function of the CVs, and C is an irrelevant additive constant.

We now introduce the basic PBMetaD equations. We consider for simplicity the case of two monodimensional WTMetaD potentials $V_G(S_1,t)$ and $V_G(S_2,t)$ acting on S_1 and S_2. The derivation below can be seamlessly extended to the case of multiple mono- or multidimensional WTMetaD bias potentials. If we apply simultaneously these two potentials as in concurrent metadynamics,[13] the bias potentials will not converge to the corresponding free energies $F(S_1)$ and $F(S_2)$ as in eq 4, because in general S_1 and S_2 are correlated. In this case, the bias deposited to one variable is in part felt by the other variable, thus leading to a "contamination" between bias potentials if all are acting at the same time during the simulation and therefore added together. To avoid this problem, in PBMetaD at any given time step either $V_G(S_1,t)$ or $V_G(S_2,t)$ is added to the underlying potential $U(\boldsymbol{R})$. To achieve this goal, we introduce a new discrete variable $\boldsymbol{\eta} = (\eta_1,\eta_2)$, which has the role of turning on and off each bias potential. This extended system samples the following probability distribution:

$$P_t(\mathbf{R}, \boldsymbol{\eta}) \propto \exp(-\beta(U(\mathbf{R}) + \eta_1 V_G(S_1, t) + \eta_2 V_G(S_2, t))) \tag{5}$$

where $\beta = 1/k_BT$ and $\boldsymbol{\eta}$ can be equal to either (1,0) or (0,1). Since $\boldsymbol{\eta}$ is a discrete variable, the probability distribution of eq 5 can be conveniently sampled by a Gibbs sampling scheme, in which coordinates are sampled by MD and $\boldsymbol{\eta}$ by Monte Carlo (MC). For the bias potentials to converge to the respective low-dimensional free energies, regardless of the fact that the variables are correlated or not, we should define each bias potential following eq 1, but in terms of the histogram of the subset of the conformations visited when the corresponding bias potential was active, or in other words conditionally on the value of $\boldsymbol{\eta}$:

$$V_G(S_1, t) = k_B\Delta T_1 \log\left(1 + \frac{\omega_1 N(S_1, \boldsymbol{\eta} = (1, 0), t)}{k_B\Delta T_1}\right)$$
$$V_G(S_2, t) = k_B\Delta T_2 \log\left(1 + \frac{\omega_2 N(S_2, \boldsymbol{\eta} = (0, 1), t)}{k_B\Delta T_2}\right) \tag{6}$$

Practically, during the simulation a new Gaussian is added using the recipe in eq 3 to either $V_G(S_1,t)$ or $V_G(S_2,t)$ based on the value of $\boldsymbol{\eta}$ at time t: when $\boldsymbol{\eta}$ = (1,0), $V_G(S_1,t)$ is updated; otherwise $V_G(S_2,t)$ is updated.

Up to this point, our approach is very similar to that of BEM: the system benefits from experiencing bias potentials acting on many different CVs without the need to explore and reconstruct a high-dimensional free-energy profile as in standard WTMetaD. However, while in BEM we need to exchange conformations between replicas that bias either S_1 or S_2, in this sampling scheme we have a single replica that switches, using a Metropolis criterion, the CV that is currently biased, and the corresponding bias potential that is updated. This implementation guarantees a greater simplicity of use and flexility since there is no requirement to introduce one replica of the system for each CV. Instead additional replicas, whenever computationally permitted, can be utilized in a multiple-walkers fashion,[19] in which several copies of the system perform a PBMetaD simulation and share the accumulated bias potentials.

In spite of these advantages, to maximize the simplicity and usability of the method, it is desirable not to introduce an additional layer of sampling. Recently, Galvelis and Sugita have proposed a way to improve BEM efficiency by introducing the infinite swapping and the Suwa–Todo algorithms.[20] Here we use a different approach and we eliminate completely the need of sampling the variable $\boldsymbol{\eta}$, i.e., to swap between the different CVs biased in our single-replica scheme. Owing to the fact that we are not interested in any thermodynamic property of $\boldsymbol{\eta}$, we can marginalize it:

$$P_t(\mathbf{R}) = \int d\boldsymbol{\eta}\, P_t(\mathbf{R}, \boldsymbol{\eta}) = P_t(\mathbf{R}, (1, 0)) + P_t(\mathbf{R}, (0, 1))$$
$$\propto \exp(-\beta(U(\mathbf{R}) + V_{PB}(S_1, S_2, t))) \tag{7}$$

where

$$V_{PB}(S_1, S_2, t) = -\frac{1}{\beta}\log(\exp(-\beta V_G(S_1, t)) + \exp(-\beta V_G(S_2, t))) \tag{8}$$

It should be noted that the effective (or marginal) bias potential as a function of S_1 and S_2 alone is clearly different from the one of concurrent metadynamics:[13]

$$V_{CM}(S_1, S_2, t) = V_G(S_1, t) + V_G(S_2, t) \tag{9}$$

This leaves only the task of determining how to update $V_G(S_1,t)$ and $V_G(S_2,t)$. Previously, the two biases were constructed during the simulation based on the value of the variable $\boldsymbol{\eta}$. Now that this variable has been marginalized, we cannot simply update $V_G(S_1,t)$ and $V_G(S_2,t)$ simultaneously using eq 3; otherwise the two bias potentials will not converge to the correct free-energy profiles (Figure S3). Instead we can

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

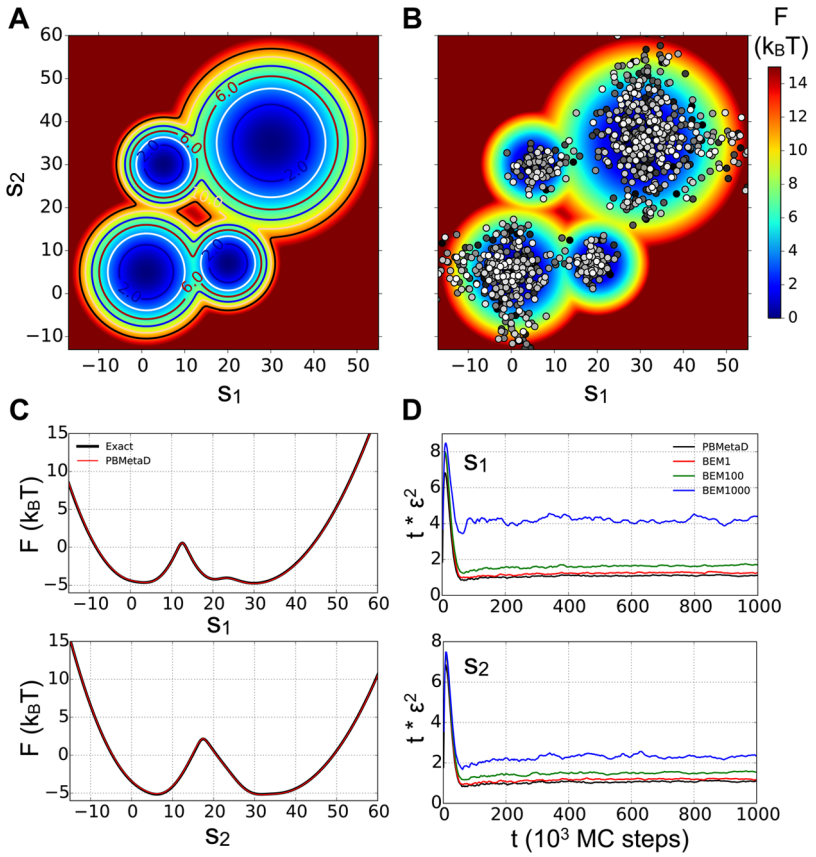


Figure 1. PBMetaD results on a model system. We verified the absence of systematic errors in the PBMetaD free-energy reconstruction and benchmarked its efficiency on a two-dimensional system characterized by the presence of four local minima, separated by high free-energy barriers (A). In the PBMetaD simulation the system repeatedly visited the four basins, as it can be seen by projecting the trajectory on the two-dimensional free energy and coloring each point based on the time step, ranging from white (beginning of the simulation) to black (end) (B). The reconstructed free energies were identical to the ones calculated analytically, for both CVs (C). The deviation of the PBMetaD free energies from the analytical solutions plotted as a function of time (D, black line) demonstrates the absence of systematic errors and the efficiency of the method in comparison with BEM (D, green, red, and blue lines), independently of the BEM exchange stride used.

make use of the conditional probability $P(\boldsymbol{\eta}|\mathbf{R})$ of observing a certain value of $\boldsymbol{\eta}$ given a configuration $\mathbf{R}$. This is trivially achieved since both the joint $P(\mathbf{R},\boldsymbol{\eta})$ and the marginal probability distributions $P(\mathbf{R})$ are known:

$$P(\boldsymbol{\eta}|\mathbf{R}) = \frac{P(\mathbf{R}, \boldsymbol{\eta})}{P(\mathbf{R})} \tag{10}$$

Note that this expression holds whether the two CVs are independent or not. Plugging eq 5, eq 7, and eq 8 into eq 10 yields

$$P(\boldsymbol{\eta} = (1, 0)|\mathbf{R}) = \frac{\exp(-\beta V_G(S_1, t))}{\exp(-\beta V_G(S_1, t)) + \exp(-\beta V_G(S_2, t))}$$
$$P(\boldsymbol{\eta} = (0, 1)|\mathbf{R}) = \frac{\exp(-\beta V_G(S_2, t))}{\exp(-\beta V_G(S_1, t)) + \exp(-\beta V_G(S_2, t))} \tag{11}$$

Since $\boldsymbol{\eta}$ can no longer be used to select which bias potential is updated, an alternate strategy is needed. Instead, we update both potentials, but with each Gaussian given an additional weight ("conditional weight") equal to the conditional probability of having the value of $\boldsymbol{\eta}$ "on" for that particular CV. Therefore, two new Gaussians are added on both S_1 and S_2 with the following rate:

$$\omega_1(t) = \omega_1 \exp\left(-\frac{V_G(S_1, t)}{k_B \Delta T_1}\right) P(\boldsymbol{\eta} = (1, 0)|\mathbf{R})$$
$$\omega_2(t) = \omega_2 \exp\left(-\frac{V_G(S_2, t)}{k_B \Delta T_2}\right) P(\boldsymbol{\eta} = (0, 1)|\mathbf{R}) \tag{12}$$

The conditional weights $P(\boldsymbol{\eta}|\mathbf{R})$ in eq 12 assign a higher additional weight to the Gaussian deposited on the CV that has the lower value of the bias potential. Similarly, in eq 8 the effective bias potential V_{PB} (S_1,S_2,t) is dominated by the minimum potential between V_G (S_1,t) and V_G (S_2,t). This in principle could be problematic if the two potentials are static and one of the two is significantly higher than the other. In that case, only one of the two potentials would be virtually active. However, in WTMetaD the two potentials constantly grow over time, thus leading to each conditional weight fluctuating between 0 and 1, although in general there is no assurance that the averages of the two weights over the entire simulation are identical (Figure S2E,F).

We show in the next section that through the use of the rescaling rule of eq 12 the two bias potentials exactly converge to the correct monodimensional free energies, and the final correction factor between bias potential and free energy is the same as in standard WTMetaD (eq 4). Furthermore, as in the long time limit each time-dependent potential becomes quasi-

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

static, we show in the next section that the full high-dimensional free energy can be easily recovered by applying a reweighting technique.

MODEL SYSTEM

To verify the absence of systematic errors in the free energy reconstructed by PBMetaD as well as benchmark its efficiency, we used a model system. The free energy of the model system is a function of two dimensionless collective variables S_1 and S_2 (Figure 1A), and it is characterized by the presence of four local minima, separated by barriers of several k_BT (eq S1 and Table T1). Moreover, the free energy is intrinsically two-dimensional; i.e., its monodimensional projections on one of the two variables merge together two independent minima (Figure 1C). Similar model systems have been extensiviley used in the past to study various properties of enhanced sampling methods.[10,21]

To sample this model system, we used a MC algorithm in which we added the PBMetaD bias potential defined by eq 8. For both CVs, we used Gaussians with initial height, width, and deposition stride equal to $0.1k_BT$, 0.3, and 1 MC step, respectively; the bias factor $\gamma = (T + \Delta T)/T$ was set to 12 and k_BT to 1. Each simulation was initialized in a random point of the CVs space and carried out for 5×10^6 MC steps. By inspecting the time series of S_1 and S_2 (Figure S1), it is clear that as the simulation proceeds and the Gaussian heights decrease (Figure S2), the system diffuses in the entire CV space and visits repeatedly the four local minima (Figure 1B). This behavior suggests the effectiveness of PBMetaD in overcoming high free-energy barriers in a multidimensional profile by applying multiple bias potentials in a lower dimensionality.

To assess whether the two monodimensional PBMetaD bias potentials converge to the correct free-energy profiles following eq 4, we monitored the error of our method defined as the root mean square deviation (RMSD) of the PBMetaD free-energy estimate from the exact analytical solution (eqs S2 and S3), averaged over 1000 independent MC simulations. Due to the fact that in PBMetaD the extent of free-energy exploration is regulated by the choice of ΔT, we limited our comparison to a region of the CVs space within 20 k_BT of the global minimum, as defined in the analytical free-energy profile. In the absence of systematic errors, the only source of error is the finite length t of the simulation, and thus the RMSD should decrease as $1/\sqrt{t}$, as it can be clearly seen in Figure 1D. If the conditional probability weight of eq 11 is not used in the Gaussian rescaling recipe (eq 12), a systematic error is introduced (Figure S3).

The long time behavior in Figure 1D provides an estimate of the prefactor in the time decay of the error, and thus it can be used to quantify the efficiency of the method, with lower prefactors corresponding to more efficient sampling methods. We used this criterion to compare PBMetaD with the popular BEM algorithm, in which two replicas were used to bias either S_1 or S_2. The parameters for the Gaussian width, initial height, CV temperature, and deposition stride were the same as in the PBMetaD simulations. The error was averaged over 1000 independent simulations, and "time" was defined as the number of MC steps multiplied by the number of replicas. Since in BEM the number of steps between two attempts at exchanging conformations (exchange stride) is a free parameter, we studied the dependency of the error on this quantity. BEM efficiency increases with decreasing exchange stride (Figure 1D, blue, red, and green lines), similarly to what was previously found in REM simulations,[22] and approaches the efficiency of PBMetaD (Figure 1D, black line). It should be noted that in real applications BEM efficiency has been shown to reach a pleatau when the exchange stride decreases.[23] In the limit of attempting an exchange every time step, PBMetaD is only slightly more efficient than BEM. However, in BEM applications to realistic systems, a drastic decrease of the exchange stride leads to a significant performance loss due to the parallel communication required at each exchange step, so that typically a value of the order of 1000−10000 time steps is used. In this regime, PBMetaD should prove particularly advantageous compared to BEM.

Another advantage of PBMetaD is the simplicity in reconstructing the high-dimensional free-energy profile. While in BEM the weighted histogram analysis method[24] (WHAM) should be used to combine simulations carried out with different bias potentials, in PBMetaD we have a single time-dependent bias potential $V_{PB}(S_1,S_2,t)$ that becomes quasi-static in the long time limit, so that a simpler reweighting technique can be utilized (eq S4 and Figure S4). Furthermore, more elaborate reweighting methods[25,26] developed for WTMetaD should be easily adapted to PBMetaD.

TRYPTOPHAN-CAGE MINIPROTEIN

After verifying on the model system that our approach reconstructs the underlying free energies without any systematic errors, we sought to demonstrate the applicability of the method in a case that is more representative of the types of applications for which we envision PBMetaD will be useful. The TRPC system is particularly interesting since it contains basic protein structural elements of secondary structure (helices) and a hydrophobic core. Following the original demonstration of the BEM approach,[10] we biased six generic CVs that describe global conformational change, a global degree of backbone dihedral rotations, contact formations for hydrogen bonds, and hydrophobic contacts, as well as the presence of both α and β secondary structure elements. An implicit solvent model was used to facilitate extended sampling and accurate comparison of CV-biased and unbiased simulations.

As a reference, we used an exhaustively sampled PT simulation (total simulation time > 70 μs) spanning a temperature range of 300—750 K. Three MetaD simulations were performed with identical CVs and parameters related to construction of the bias potential: BEM100, BEM1000, and PBMetaD. The 100 and 1000 simply refer to the exchange stride between biased replicas. Each BEM simulation used one replica per CV (six total replicas), and to ensure identical comparison, the PBMetaD simulation used six replicas within the multiple-walkers MetaD framework.[19] The total sampling time was ∼30 μs each for the three different MetaD simulations. The three MetaD simulations began from identical starting configurations (an ensemble of six randomly selected unfolded TRPC structures), and the PT simulation began from a similar ensemble of 12 randomly selected structures. Extensive details of the simulations including all of the parameters for the all of the simulations in GROMACS[27] and PLUMED[28] are provided in the Supporting Information.

Before using the exhaustive PT simulation as a reference for measuring the efficiency and properties of PBMetaD, we first ensured that the reference simulation was converged. Following suggested procedures,[29] we first analyzed the diffusion of the system through relevant collective variable space using a reconstruction of the continuous trajectories from the PT simulation. As shown in Figure S5, there is excellent diffusion

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

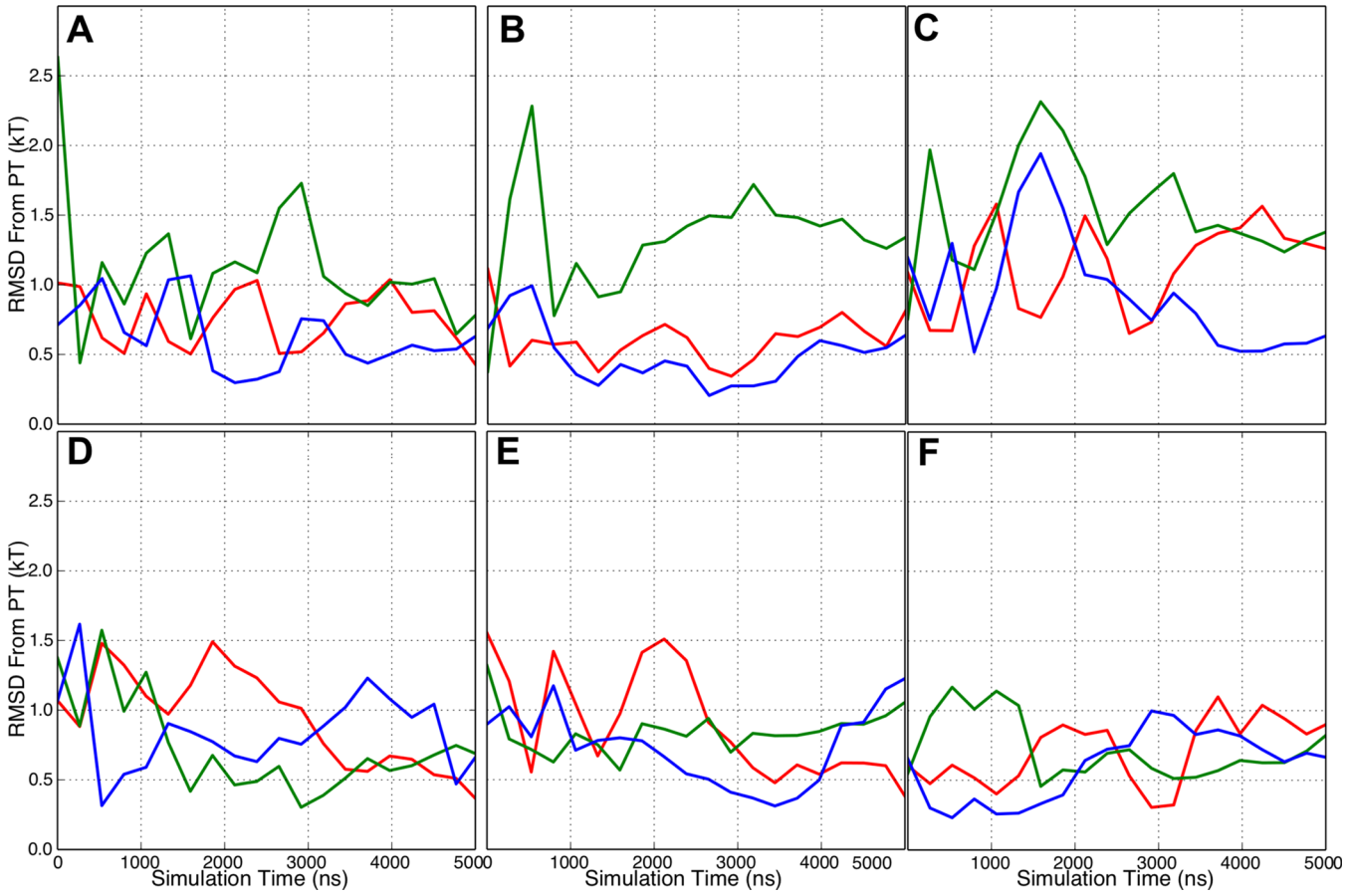


Figure 2. PBMetaD applied to the TRPC miniprotein. The lines show the convergence of the error between the monodimensional free energies obtained from PT and PBMetaD (blue), BEM100 (red), and BEM1000 (green). The convergence metric is the RMSD measured with the KL divergence, which was applied as described in the Supporting Information. The simulation time represents MD simulation time per replica in the implicit solvent model, with each curve representing 30 μs of aggregate sampling. The six panels represent the biased CVs: (A) hydrophobic contacts, (B) dihedral correlation, (C) hydrogen bonds, (D) α-helicity, (E) β-similarity, and (F) radius of gyration.

across all 12 replicas, with no obvious instances of a system becoming trapped in some point of the CV space. It should be noted that, due to the use of an implicit solvent model, the system expanded and collapsed extremely quickly once it reached the hot and cold replicas, respectively, thus making PT particularly efficient in this case (Figure S5). Following this, we analyzed the time dependence of the free-energy profiles of each of the six CVs as obtained from the 300 K replica. Throughout the last half the simulation the free-energy profiles remain constant (within a tiny fraction of k_BT) (Figure S6), while the simulation is diffusing through CV space, thus establishing convergence of the relevant observables from the simulations.

Next, we assessed the convergence of the three biased simulations. While for PBMetaD it is sufficient to monitor the CV diffusion along with the change in the estimated free energies with time (Figures S7 and S8), for the BEM simulations we also need to reconstruct the continuous trajectories, as for the PT case (Figures S9 and S10). Once convergence was established, we compared the final free energies obtained from the reference simulation with those obtained from the three biased simulations (Figure S11). As seen in the figure, there is not perfect agreement with the PT simulation. In each of the six free energies there are subtle differences between different features present in the CV-biased simulations that are not present in the free energies derived from PT, which are uniformly much more smooth. However, the different features are all on the order of k_BT. Some of the differences are systematic and reproduced across all three CV-biased simulations. Specifically, the free energies for hydrophobic contact and radius of gyration both show clear shoulders not present in the PT run at low values of the CVs. In all other cases there is no systematic difference between any of the simulations and the reference. In summary, there is excellent agreement between PT, BEM, and PBMetaD for these six CVs for implicity solvated TRPC. Note that, in the case of radius of gyration, all three CV-biased simulations reached significantly higher values of the CV. The PT run was not able to sample these high values, presumably due to the lack of sufficient thermal energy to compensate for the entropic driving force to remain in a coiled state.

Finally, to quantify the efficiency of PBMetaD in a real system we monitored the convergence of the CV-biased simulations by comparing the deviation from the exhaustive PT simulation using the same metric as in the preceding model system as well as the Kullback–Leibler divergence (KLDiv) method.[30] Owing to the strong favoring toward lower energies from the Boltzmann weight, the KLDiv results shown in Figure 2 are likely more representative of the type of convergence behavior one expects for a protein folding problem. The RMSD results, which weigh equally high- and low-energy states on the

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX

free energies, are also shown in Figure S12, where it can be seen that both metrics show similar convergence behavior.

Examination of the convergence profiles in Figure 2 provides for two conclusions. First, the BEM and PBMetaD simulations show basically identical noisy behavior over the 30 μs simulation. Second, the agreement with the reference free energies obtained from the exhaustive PT simulation is excellent, with nearly all of the final errors below k_BT and all 18 free energies possessing a final error less than 1.5 k_BT. We expect that in the limit of dramatically longer sampling or multiple independent simulations on this implicitly solvated TRPC system that the convergence behavior obtained in the model system (e.g., Figure 1) would be recovered, with PBMetaD appearing as an effective upper limit of the maximum accuracy of a BEM simulation.

In this letter we presented a method that is able to enhance sampling in systems characterized by the presence of multiple free-energy minima in a multidimensional space of selected degrees of freedom, or CVs, by applying in parallel several low-dimensional biases. We demonstrated its efficiency against the popular BEM approach in a model system and in the more realistic case of the 20-residue TRPC miniprotein. In both cases, PBMetaD also proved to be a more flexible and easier method to use. We showed the possibility of choosing a number of replicas of the systems independently from the number of CVs that need to be biased, which simplifies the application of our technique and the analysis of the results. In particular, we avoided the need to reconstruct the continuous replicas trajectories in order to assess the convergence of our method, as is instead necessary in BEM to detect whether replicas get trapped in local regions of the CVs space. Furthermore, we showed (Supporting Information) that the high-dimensional free energy can be recovered by a simple reweighting technique without the need to use the more complicated WHAM approach. Finally, PBMetaD is implemented in the popular enhanced sampling package PLUMED, thus enabling its application in a wide variety of problems in physics, chemistry, and biology.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jctc.5b00846.

Further details about the setup and analysis of the PBMetaD calculations on the model systems and on TRPC, including details on the form of the analytical potential, the reweighting procedure, and the CVs used in the TRPC system (PDF)

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: mb2006@cam.ac.uk.

Funding

J.P. acknowledges support of NSF Award CBET-1264459. This work was facilitated through the use of computational, storage, and networking infrastructure provided by the Hyak supercomputer system, supported in part by the University of Washington.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank Alessandro Laio, Carlo Camilloni, Giovanni Bussi, and Michele Vendruscolo for carefully reading the manuscript and giving several insightful comments. Alessandro Barducci is also acknowledged for useful discussions.

■ REFERENCES

(1) Frenkel, D.; Smit, B. *Understanding molecular simulation: From algorithms to applications*, 2nd ed.; Academic: San Diego, CA, USA, 2002; p xxii, DOI: 10.1016/B978-012267351-1/50002-X, 638 pp.

(2) Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. *Science* **2010**, *330*, 341−6.

(3) Beberg, A. L.; Ensign, D. L.; Jayachandran, G.; Khaliq, S.; Pande, V. S. Folding@home: Lessons from eight years of volunteer distributed computing. *Proceedings of the IEEE International Symposium on Parallel & Distributed Processing, 2009*, IPDPS '09; IEEE Computer Society: New York, 2009; pp 1−8, DOI: 10.1109/IPDPS.2009.5160922.

(4) Chipot, C.; Pohorille, A. *Free energy calculations: Theory and applications in chemistry and biology*; Springer: Berlin, New York, 2007; p *xviii*, 517 pp.

(5) Dellago, C.; Bolhuis, P. G. *Adv. Polym. Sci.* **2009**, *221*, 167−233.

(6) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562−12566.

(7) Barducci, A.; Bonomi, M.; Parrinello, M. *Wires Comput. Mol. Sci.* **2011**, *1*, 826−843.

(8) Sugita, Y.; Okamoto, Y. *Chem. Phys. Lett.* **1999**, *314*, 141−151.

(9) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435−13441.

(10) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553−4559.

(11) Bonomi, M.; Parrinello, M. *Phys. Rev. Lett.* **2010**, *104*, 190601.

(12) Deighan, M.; Bonomi, M.; Pfaendtner, J. *J. Chem. Theory Comput.* **2012**, *8*, 2189−2192.

(13) Gil-Ley, A.; Bussi, G. *J. Chem. Theory Comput.* **2015**, *11*, 1077−1085.

(14) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.

(15) Tribello, G. A.; Ceriotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5196−5201.

(16) Spiwok, V.; Kralova, B. *J. Chem. Phys.* **2011**, *135*, 224504.

(17) Neidigh, J. W.; Fesinmeyer, R. M.; Andersen, N. H. *Nat. Struct. Biol.* **2002**, *9*, 425−430.

(18) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.

(19) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533−3539.

(20) Galvelis, R.; Sugita, Y. *J. Comput. Chem.* **2015**, *36*, 1446−1455.

(21) Laio, A.; Rodriguez-Fortea, A.; Gervasio, F. L.; Ceccarelli, M.; Parrinello, M. *J. Phys. Chem. B* **2005**, *109*, 6714−6721.

(22) Sindhikara, D.; Meng, Y. L.; Roitberg, A. E. *J. Chem. Phys.* **2008**, *128*, 024103.

(23) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. *Curr. Phys. Chem.* **2012**, *2*, 79−91.

(24) Ferrenberg, A. M.; Swendsen, R. H. *Phys. Rev. Lett.* **1989**, *63*, 1195−1198.

(25) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615−21.

(26) Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2015**, *119*, 736−42.

(27) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(28) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604−613.

(29) Barducci, A.; Pfaendtner, J.; Bonomi, M. *Methods Mol. Biol.* **2015**, *1215*, 151−171.

(30) Sutto, L.; D'Abramo, M.; Gervasio, F. L. *J. Chem. Theory Comput.* **2010**, *6*, 3640−3646.

DOI: 10.1021/acs.jctc.5b00846
J. Chem. Theory Comput. XXXX, XXX, XXX−XXX