

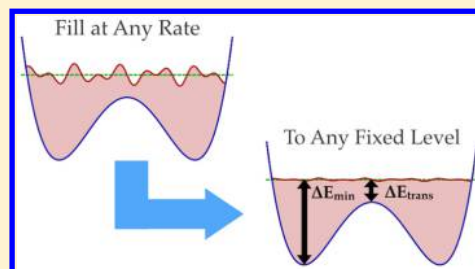
# Exploring Valleys without Climbing Every Peak: More Efficient and Forgiving Metabasin Metadynamics via Robust On-the-Fly Bias Domain Restriction

James F. Dama,<sup>†,‡</sup> Glen M. Hocky,<sup>†</sup> Rui Sun,<sup>†</sup> and Gregory A. Voth<sup>\*,†,‡</sup>

<sup>†</sup>Department of Chemistry, James Franck Institute, Institute for Biophysical Dynamics, and Computation Institute, The University of Chicago, Chicago, Illinois 60637, United States

<sup>‡</sup>Center for Nonlinear Studies, Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545, United States

**ABSTRACT:** Metadynamics is an enhanced sampling method designed to flatten free energy surfaces uniformly. However, the highest-energy regions are often irrelevant to study and dangerous to explore because systems often change irreversibly in unforeseen ways in response to driving forces in these regions, spoiling the sampling. Introducing an on-the-fly domain restriction allows metadynamics to flatten only up to a specified energy level and no further, improving efficiency and safety while decreasing the pressure on practitioners to design collective variables that are robust to otherwise irrelevant high energy driving. This paper describes a new method that achieves this using sequential on-the-fly estimation of energy wells and redefinition of the metadynamics hill shape, termed metabasin metadynamics. The energy level may be defined *a priori* or relative to unknown barrier energies estimated on-the-fly. Altering only the hill ensures that the method is compatible with many other advances in metadynamics methodology. The hill shape has a natural interpretation in terms of multiscale dynamics, and the computational overhead in simulation is minimal when studying systems of any reasonable size, for instance proteins or other macromolecules. Three example applications show that the formula is accurate and robust to complex dynamics, making metadynamics significantly more forgiving with respect to CV quality and thus more feasible to apply to the most challenging biomolecular systems.



## 1. INTRODUCTION

Although in principle observing nature carefully over infinite time scales could be sufficient to reveal all physics, it is more practical to design experiments that investigate specific questions by intentionally varying physical parameters in a controlled manner. Similarly in computational modeling, when direct simulation of natural processes by molecular dynamics<sup>1–3</sup> is infeasible, specially designed simulations can nonetheless reveal key physics in far less simulation time.<sup>4–6</sup> The adaptive enhanced sampling method metadynamics<sup>7–9</sup> is one such approach specifically designed for the determination of potentials of mean force (PMFs) by promoting transitions between long-lived metastable states. Metadynamics has been widely applied across chemistry from materials science to biochemistry, but it is still quite young theoretically, with a rigorous proof of convergence appearing only one year ago.<sup>10</sup>

Metadynamics works by using a choice of reduced coordinates called collective variables (CVs) to iteratively build a bias that increases the rates of transitions between metastable energy wells; increased transition rates imply decreased sampling autocorrelation and thus improved PMF estimates. Other adaptive methods of the same generation and similar philosophy include the Adaptive Biasing Force<sup>11,12</sup> and Wang–Landau<sup>13</sup> algorithms, and it is descended from the Local Elevation Method.<sup>14</sup> The degree to which a bias can actually promote those transitions, however, depends on how well the CVs capture the true reaction coordinates. When the CVs are

imperfect, the results may not approach the true PMF rapidly, and this is common enough that it is often considered to be the single most relevant limitation preventing application of metadynamics to the study of complex systems.<sup>9,15–18</sup> Furthermore, the cost of building the bias also depends on the complexity of the CVs, scaling with the volume of CV space—i.e., exponentially with CV number. All enhanced sampling methods that rely on CVs share these drawbacks to greater or lesser extents.<sup>6,19,20</sup> This paper describes a new variant of metadynamics, metabasin metadynamics (MBMe-taD), that is designed to suffer less from the use of poor quality CVs and to make more practical the use of larger numbers of CVs by judiciously restricting the bias's domain in CV space. However, in order to discuss the features of metadynamics that we wish to improve with our new method, we must first compare to a more venerable alternative, window-based umbrella sampling.<sup>21,22</sup>

Window-based umbrella sampling is stratified sampling applied to simulation.<sup>21</sup> It is accomplished by running many simulations with different energetic biases that keep each simulation restrained within a different small region, or window, of CV space. These windows are constructed so that the sampled distributions in the windows overlap with one another to cover all of the phase space of interest in a given

Received: September 21, 2015

Published: November 3, 2015

investigation; this typically involves choosing 1) a scale for each stratified dimension of CV space to set the separation of window centers from one another and 2) a single energy scale to set the strengths of the restraint. Once these are chosen, one then runs simulations in each window, perhaps with some form of replica exchange among the biased walkers.<sup>6,22</sup> After this, the sampling in all of the windows forms a patchwork covering of the CV space that can then be sewn together into a single overall PMF estimate using a method such as Gaussian process regression,<sup>23</sup> the weighted histogram analysis method,<sup>24</sup> or the multistate Bennett acceptance ratio.<sup>25</sup>

Metadynamics, on the other hand, is an auxiliary distribution sampling approach. It functions by iteratively building a bias away from previously visited points to accelerate escape from metastable basins and thereby decrease autocorrelation of sampling on the CV space. In tempered metadynamics,<sup>26,27</sup> the type of metadynamics that converges asymptotically,<sup>10</sup> the bias is constructed by adding Gaussian hills of bias of progressively shrinking height at regular intervals. Overall, a typical case requires choosing 1) the length scale of the hill in each dimension, 2) a single hill energy scale, and 3) a single energy scale for the convergence rate of the bias. After these are chosen, one then runs one or more simulations under an iteratively growing bias, again with the possibility of replica exchange.<sup>28–30</sup> Once this is done, one can calculate PMFs via their direct connection to the bias in the asymptotic regime or via one of several nonequilibrium reweighting estimators.<sup>31–33</sup>

When stratifying based on a set of CVs is not enough to ensure good sampling over the CV range, one says that there are hidden slow variables (HSVs). When these are present, replica exchange can be inefficient, and sampling in the windows often does not accurately represent true Boltzmann sampling on accessible simulation time scales.<sup>22,34</sup> The free energy differences between windows spaced far apart are pieced together using information from windows between them, making the errors induced by poor sampling in a single window nonlocal. This nonlocality follows the topology of the PMF in a way such that sampling errors in transition regions have large effects on relative free energy estimates between the basins they connect, while *errors in basin regions and high energy regions have a more local effect.*<sup>35</sup> However, unless the HSVs prevent replica exchange or lead to obvious unphysical features in a PMF, the erroneous sampling can easily be mistaken for correct.

The presence of HSVs in metadynamics, on the other hand, typically causes sampling hysteresis that is clearly visible in the bias.<sup>9,17,28</sup> When the CVs are open for modification, this is an advantage because it gives a clear signal that the CVs need improvement. As in window-based umbrella sampling, the effect on PMF estimates is nonlocal, but, unlike in that method, *the nonlocal effects from HSV sampling in high-energy regions can affect the estimates of the PMF across all other regions.* Specifically, the nonequilibrium bias can drive undesired changes in HSVs that never relax on the time scale of simulation, for instance driving undesired boiling of solvent or an irreversible fluctuation of a protein conformation. That spoils the sampling, making it useless for prediction. Therefore, although hysteresis in the presence of HSVs is a desired behavior in regions of CV space relevant to the process of interest, it is an inconvenience in high-energy regions. It is especially troublesome in complex systems with sensitive dynamical environments and many potential HSVs, e.g., almost all proteins.

Regardless of the presence of HSVs, the simplest form of windowed umbrella sampling scales poorly with the dimensionality of CV space because one typically covers the space with the fixed-volume windows, and the cost of covering space with fixed-volume sets scales exponentially with dimension. This can be ameliorated somewhat by taking a metadynamics-like approach in which one runs simulations only in windows in low-energy regions encompassing the main basins of interest, i.e., a metabasin, gradually learning the shape of the metabasin along the way.<sup>36</sup> The metabasin manifold has a lower volume—and often a lower effective dimensionality—than its bounding box, so covering it with windows is less prohibitive than covering the full span of the CV space. No advance knowledge of the metabasin's shape is required, only an energy scale of interest.

In this paper we propose to emulate this approach for metadynamics. However, our goal and approach differ. Our primary goal with MBMetaD is to prevent driving in high-energy regions to prevent the excitation of HSVs that irreversibly alter sampling. Greater efficiency with many CVs is another benefit and an important one, but it is currently a less pressing issue in the application of metadynamics than the problem of irreversible driving. The strategy differs from adaptive umbrella sampling because metadynamics already acts to fill metabasins from the lowest point out—an initialization logic—and our change is therefore rather to cause metadynamics to drive escape no further and converge once it has filled the metabasin—a termination logic.

The most common approach to this termination in the literature has been to self-limit by slowing down the addition of bias, for instance in self-healing umbrella sampling,<sup>37</sup> flat histogram metadynamics,<sup>38</sup> and well-tempered metadynamics.<sup>26</sup> However, though these can be effective, in such methods the time scales of updating the bias and the energy scale of the self-limiting are tightly coupled, which can make it impossible to simultaneously choose acceptable values for both in complex systems. Most often, one must simply choose to add energy very slowly or, similarly, decrease the rate of energy addition very rapidly to prevent spoiling the dynamics. In either case, this can make the simulations prohibitively expensive. There is therefore a need for a scheme that takes a different approach, sampling a specific energy scale regardless of how quickly bias is added, to empower researchers to make more practically efficient investigations of sensitive systems with metadynamics.

In this paper, we solve the longstanding problem of designing metadynamics to flatten only a region of arbitrary shape, without restricting sampling to that region, in order to design a metadynamics that flattens only up to a chosen energy level and no further regardless of how quickly new bias is added. The approach we take is compatible with all tempering strategies,<sup>10,26,27,39</sup> bias exchange,<sup>29</sup> multiple walkers,<sup>40</sup> and more,<sup>18,30,41,42</sup> functioning as a modular enhancement that can and should be used together with other advances in metadynamics methodology. The task of filling a finite domain has been a pernicious challenge throughout the history of metadynamics.<sup>43–45</sup> When used without proper understanding it can suffer from boundary artifacts. Solutions are available for 1D intervals,<sup>43,44</sup> but the first simple, implemented solutions for rectangular boxes with hard wall boundary conditions in any number of dimensions were first published by McGovern and de Pablo only two years ago.<sup>45</sup> We independently derived the same corrections in the course of studying the convergence of metadynamics, and here we extend the McGovern–de Pablo

approach to time-dependent domains of arbitrary shape with open boundaries. Our method consists of this generalized boundary-corrected metadynamics together with adaptive rules for defining time-dependent domains that represent the metabasins of genuine physical interest in practical metadynamics applications. These rules do not require any *a priori* estimate of barrier heights in the system. Section 2 details this new method, Section 3 presents results and discussion of three example applications to biomolecular simulation, and Section 4 concludes the paper.

## 2. METHODS

Tempered metadynamics is proven to converge like a specific quasiequilibrium differential equation that depends only on the method and not on underlying system dynamics.<sup>10</sup> Therefore, to describe the design of the new approach that converges on a finite, open-boundary domain, we first define a new shape of hill that gives rise to a differential equation that flattens the energy on any domain. Second, because the domain of interest will only rarely be known ahead of time, we next discuss how to adaptively define these domains over the course of simulation using either a known target basin energy level or an adaptively learned energy level corresponding to the minimal metabasin containing several known points of interest.

In the original metadynamics,<sup>7</sup> the adaptive bias is constructed as a sum of Gaussian hills centered at previously visited points in CV space, i.e., in 1D

$$V(s, t_n) = \sum_{i=0}^n h e^{-(s-s_i)^2/2\sigma^2} \quad (1)$$

where  $h$  is an energy-valued hill height,  $\sigma$  is a CV-valued hill width, and the  $s_i$  are the sequence of points in CV space visited at each  $t_i$ . To ensure convergence one must sequentially decrease the hill height, for instance using the well-tempered metadynamics (WTMetaD)<sup>26</sup> rule

$$V(s, t_n) = \sum_{i=0}^n h e^{-(s-s_i)^2/2\sigma^2} e^{-V(s_i, t_{i-1})/\Delta T} \quad (2)$$

where  $\Delta T$  is a parameter that controls the rate of height decrease. A general theoretical form for metadynamics that can be proven to converge<sup>10</sup> is

$$V(s, t_n) = \sum_{i=0}^n G(s, s_i) w[V(s, t_n)] e^{-V(s_i, t_{i-1})/\Delta T} \quad (3)$$

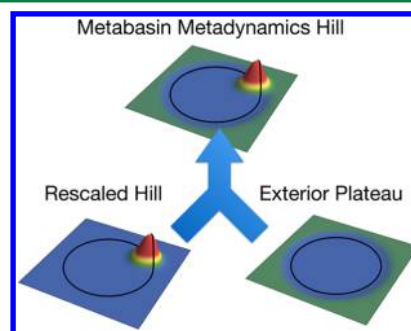
where  $G(s, s')$  is called the hill kernel and encapsulates the hill shape, and  $w[V(s, t_n)]$  is called the global tempering rule and serves as a complement or substitute to the WTMetaD rule for decreasing the hill sizes as time goes on; see for example transition-tempered metadynamics (TTMetaD)<sup>27</sup> or the global tempering used in experiment directed metadynamics.<sup>39</sup> The new method of this paper is modular with respect to the various tempering rules and only involves redefinition of  $G(s, s')$ .

Specifically, consider the case that one knows of some hill kernel  $G_0(s, s')$  that could flatten an entire CV space but would like to flatten a domain  $D$  of that CV space instead. Then the MBMetaD rule is to use a hill with two parts

$$G_{MB}(s, s') = G_{Int}(s, s') + G_{Ext}(s, s') \quad (4)$$

where  $G_{Int}(s, s')$  are hills that would flatten the interior of the domain if it had hard walls, and  $G_{Ext}(s, s')$  raises the bias level of

the exterior in such a way to exactly match the exterior bias level to the bias level of the domain boundary while also exactly counteracting the parts of  $G_{Int}(s, s')$  that slope into the exterior of the domain. Both are based only on  $G_0(s, s')$  and the domain  $D$ . These hills are illustrated in Figure 1.



**Figure 1.** An illustration of a metabasin metadynamics hill function (top, eq 4) on the unit circle near the domain boundary (black lines denote  $r = 1$ ) and its decomposition into a rescaled hill (left, eq 7) and a low plateau outside of the domain (right, eq 9). The hill is based on a Gaussian of width  $1/\sqrt{10}$  placed at  $r = 0.9$ . The effects of adding rescaled hill and exterior plateau contributions evenly over all points on the unit circle are exact complements, leading to bias updates that are flat everywhere when sampling is flat on the unit circle.

For the first part, we use a close relative to the multiplicative McGovern–de Pablo<sup>45</sup> rule

$$G_{McGdP}(s, s') = G_0(s, s')/I(s; D, G_0) \quad (5)$$

where  $I(s; D, G_0)$  is a boundary-normalized integral of  $G_0(s, s')$  over  $D$  for its  $s'$  argument, i.e.

$$I(s; D, G_0) \equiv \int_D ds' G_0(s, s') \left( \frac{\int_{\delta D} ds \int_D ds' G_0(s, s')}{\int_{\delta D} ds} \right)^{-1} \quad (6)$$

with  $\delta D$  the boundary of domain  $D$ . This integral does not have a closed form solution for domains of general shape, but for Gaussian hills on a rectangular domain it is given by the sum of error functions provided in the paper of McGovern and de Pablo.<sup>45</sup> The rule eq 5 has the undesirable property that as  $s'$  approaches the boundary of  $D$  from the interior, the values of  $G(s, s')$  in the exterior diverge to infinity, making it unsuitable for metadynamics on an open boundary domain. This occurs because  $I(s; D, G_0)$  goes to zero quickly as  $s$  goes beyond the boundary, so an expedient solution is to modify  $I(s; D, G_0)$  such that it falls to a fixed value greater than zero. In MBMetaD, therefore, our form for the first term in eq 4 is

$$G_{Int}(s, s') = G_0(s, s')/f(I(s; D, G_0)) \quad (7)$$

with  $I(s; D, G_0)$  as above and in the implementation described in this paper  $f(x)$  is the simple piecewise polynomial  $f(x) = x$  for  $x \geq 1$  and  $f(x) = x + 0.5(1-x)^2$  for  $x < 1$ . This choice of  $f$  is continuous and differentiable even at  $x = 1$ , ensuring that all bias forces will be continuous.  $I(s; D, G_0)$  is constant in the interior of the domain and decays to zero outside of the domain, so dividing by  $f(I(s; D, G_0))$  has the effect of leaving hills in the interior unchanged and causing them to bulge out of the domain subtly more than they otherwise would when close to its boundary.

Using only that first part of the hill, the differential equation from the convergence paper<sup>10</sup>

$$\frac{dV(s, t)}{dt} = r(t) \int ds' G(s, s') p_{wt}(s') \quad (8)$$

would predict that the interior of the region would be flattened, but the bias inside would grow indefinitely, forming a steep-walled, flat-topped mesa in the PMF. This would drive all sampling out of the metabasin of interest and is clearly undesirable. The second term in eq 4 ensures that this does not occur. Specifically, we use

$$G_{Ext}(s, s') = (1 - I(s; D, G_0)) / f(I(s; D, G_0)) \frac{\int_{\delta D} ds G_{Int}(s, s')}{\int_{\delta D} ds} \quad (9)$$

which plateaus in the exterior of the region with a plateau height equal to the boundary average value of the first part of the new bias. An example is illustrated in the bottom right of Figure 1. Because its height is set to match the bias added to the boundary, it ensures that the exterior bias level always matches the average bias level of the domain boundary. In other words, this part of the bias update pulls sampling into the domain everywhere along the boundary exactly enough to cancel out how the first part of the hill would push sampling out of the domain close to the new hill. Moreover, the plateau is perfectly shaped so that flat quasiequilibrium sampling on the interior of the domain leads to an exactly flat increase of the bias everywhere in CV space. Finally, this term, eq 9, is only nonzero when a hill is added close to the boundary ( $s'$  is close to the boundary), so MBMetaD functions just like normal metadynamics away from those boundaries.

While this scheme might appear technical at first glance, it has a simple multiscale physical interpretation. On a coarse scale, we look at the system through a two-state lens: either the sample is inside or outside the domain. On a fine-grained scale, we look at the points in CV space to a resolution comparable to the length scale  $\sigma$  of the original hill. The metabasin rule, then, corresponds to looking at the CV space using a topology in which all the points inside the domain are resolved to a fine-grained level, while all of the points outside are collapsed into one generic state. The first part of the hill is added to a fine-grained neighborhood of the interior, while the second is added to the entire coarse-grained exterior state—also a single neighborhood, but using the multiscale topology described above instead of the usual metric topology of the CV space.

It is often argued that bias hills should match the Green's functions for the dynamics, in at least a loose sense,<sup>14,32,46,47</sup> and this is no exception. The simplest way to understand this new hill shape in those terms is to recognize that these hills are also like Green's functions, but where boundary conditions have been adjusted so that the walls are only partially reflective and any particles that exit the boundary from the interior must be balanced by particles crossing the boundary from the exterior. In the exterior, the constant profile in the far field corresponds to a well-mixed, fixed concentration boundary condition at infinity. The second part is thus related to the quasistationary distribution of a well-mixed domain exterior and therefore retains a natural connection to dynamics.

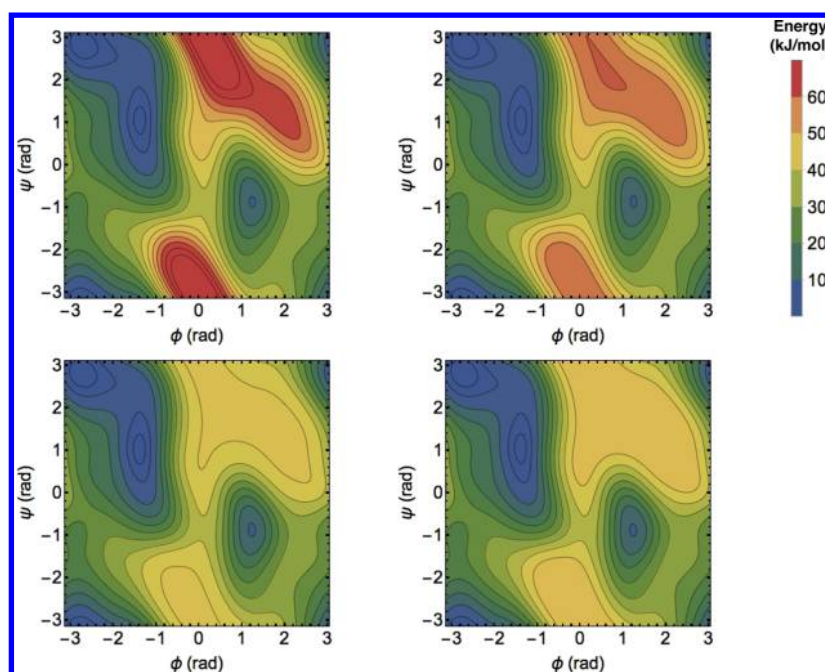
Up to now we have spoken in terms of a single connected domain, but the development above applies to any disconnected domain just as easily. Simply divide the domain  $D$  into its  $n$  components  $D_i$  and whenever a sample would be

added inside a domain  $D_i$ , use the above definitions with  $D_i$  in the place of  $D$ . In the discussion of its physical meaning, consider a  $2^n$ -state coarse-graining rather than a two-state coarse-graining; the new set of states is the Cartesian product of all the sets of two states (interior and exterior) per domain.

When one's goal is to flatten a PMF in a domain without causing discontinuities, the only domains that can be flattened completely are those with an isoenergetic boundary. Therefore, it is natural to define valid domains by defining a boundary free energy and letting the domain be the set of all CV points with a lower free energy than that boundary energy. Since the zero of free energy is arbitrary, these definitions must be made in terms of an energy difference rather than an energy alone. For instance, in this paper, we will show the use of domains like “all points less than 45 kJ above the minimum free energy”, a known basin free energy level, and “all points less than 15 kJ above the transition barrier between points A and B”, an implicitly defined metabasin energy level. Of course, the choice of domains  $D_i$  is rarely possible to make before a simulation. It is often one of the things one would like to discover using metadynamics. For that reason, we also add an on-the-fly mechanism to discover  $D_i$  defined in terms of simpler criteria like the examples above.

We solve this problem much as in TTMetaD,<sup>27</sup> where we used the bias as an on-the-fly free energy estimator. However, in this case we seek metabasins rather than barriers, and now the bias will not properly approximate the free energy anywhere outside of a previously defined domain. Therefore, we use the nonequilibrium umbrella sampling estimator of Branduardi, Bussi, and Parrinello<sup>32</sup> to approximate the domain locations on-the-fly rather than the bias alone. This allows for the discovery of extensions of a domain outside a previously chosen domain, possibly including new disconnected components, regardless of initial conditions. The free energy estimates provide minima, transition barriers, and on-the-fly region selections as described in the next paragraph.

The recipes above have straightforward numerical implementations.  $I(s; D, G_0)$  can be calculated by numerical integration on any domain  $D$  simply by summing many hills  $G_0(s, s')$  evenly over  $D$ , and once it is available  $f(I(s; D, G_0))$  and  $I(s; D, G_0)/f(I(s; D, G_0))$  are simple to calculate. The nonequilibrium umbrella sampling estimator is evaluated using a running histogram accumulated during simulation alongside the bias, free energy minima are trivial to calculate, and the transition barriers are found via breadth-first path search as in TTMetaD.<sup>27</sup> These minima and barriers, plus an offset, define the target region energy. The region of interest is then the total set of grid points under that free energy in the current on-the-fly estimate. That set of grid points is then split into connected domains with a standard flood-fill connected components labeling algorithm.<sup>48</sup> Once these domains are defined, we compute  $I(s; D, G_0)$  for each domain and then continue the simulation. Our implementation updates the domain at fixed time intervals with an option not to update the domain each time but to instead also wait until the free energy estimate in the exterior of the region has changed by some set amount. With tempering, the latter corresponds to a logarithmic update schedule that cuts down on unnecessary computational overhead at late times when the domain definition is stable. Each change in domain definition only affects future hill addition, so the previously deposited bias continues to be used unchanged.



**Figure 2.** Bias-based free energy estimates averaged over 32 runs of 8 ns for four different metadynamics choices designed to give different levels of resolution of the high-energy regions. Clockwise from top left: WTMetaD, TTMetaD, 15 kJ/mol transition-referenced MBMetaD, 42.5 kJ/mol minimum-referenced MBMetaD. All methods use parameters as defined in the text. Contours are placed every 5 kJ/mol. The estimates are all but identical in low-energy regions despite stark differences in the high-energy regions. WTMetaD is least controllable and gives the most resolution of high energy regions, TTMetaD is more controllable but would give more and more resolution of high energy regions given additional convergence time, and both forms of MBMetaD are fully controllable, converging without resolving the high energy regions.

Our implementation is a public fork of the PLUMED2 package.<sup>49</sup> It provides MBMetaD with a minimum based domain or a transition barrier based domain (with transition barriers defined as described in the paper introducing TTMetaD<sup>27</sup>), using only one additional required input number, the energy offset from the transition or minimum. Two further parameters to control the update schedule and one controlling accuracy of the numerical integration are also available for fine-tuning, but these have general defaults and can be considered optional.

### 3. RESULTS AND DISCUSSION

Our purpose with MBMetaD is to fundamentally change the trade-offs of metadynamics with respect to other methods by giving it a self-limiting mechanism, and thus the remainder of this paper focuses on investigating the self-limiting behavior in three examples. First, we show that it functions as intended all the way through PMF convergence for a simple model biomolecule with imperfect CVs, alanine dipeptide, without negatively affecting convergence efficiency or accuracy. Second, we show that it correctly self-limits in a more complex example, monomeric actin, where the self-limiting is used to achieve more stable and reproducible exploration of conformational space. Finally, we apply it to a membrane transport protein, CIC-ec1, using mediocre CVs to examine its failure modes and in particular show that it retains metadynamics' attractive property of displaying hysteresis in the domain of interest (rather than harder to recognize errors) even as it prevents the nonequilibrium driving from spoiling the dynamics in other domains.

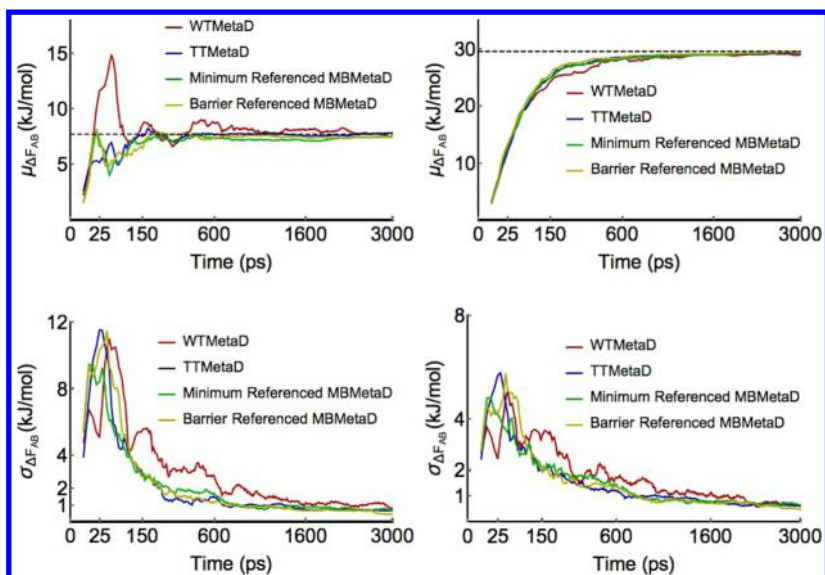
Our focus in these examples is the self-limiting behavior and reproducibility rather than convergence efficiency or accuracy; self-limiting behavior aimed at improving reproducibility is the

sole novel feature introduced in MBMetaD, and the method otherwise behaves much like whatever other metadynamics method it augments. Though the accuracy of metadynamics in practice does deserve further study, it is not a focus of this particular paper.

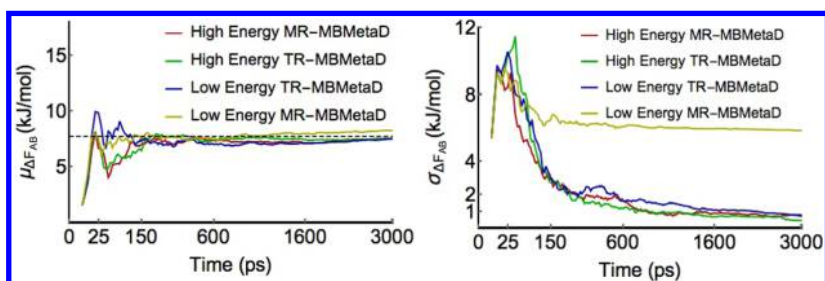
**3.1. Alanine Dipeptide.** A standard first test case for new metadynamics methods is blocked alanine dipeptide *in vacuo* with  $\phi$  and  $\psi$  dihedral CVs because the long literature on metadynamics provides excellent guidance on transferring knowledge gained with this model to understanding of the method in more complex systems.<sup>7,26,32</sup> Alanine dipeptide provides a simple case of a system with imperfect separation of time scales between CV dynamics and other relaxations and is thus a minimal test for any adaptive enhanced sampling method that is meant to be robust to memory in the CV dynamics.

These tests modeled alanine dipeptide using the Amber03 force field<sup>50</sup> and at a constant temperature of 300 K using a Langevin thermostat with a drag of 5 amu/ps. The dynamics were integrated in Gromacs 4.6.1<sup>51–54</sup> patched with a customized version of PLUMED2<sup>49</sup> and using a stochastic dynamics leapfrog algorithm with a time step of 2 fs, particle-mesh Ewald summation,<sup>55</sup> and SHAKE constraints on all bonds.<sup>56</sup> We performed 32 simulations for each method and parameter set presented. Each simulation began from the same  $C_{7eq}$  conformation, but each set of 32 runs began from 32 distinct initial velocities pseudorandomly generated from distinct random seeds and used 32 distinct Langevin thermostat seeds.

We compared WTMetaD, TTMetaD, and transition-tempered MBMetaD using hill height, width, and rate parameters and tempering parameters as in our previous work introducing TTMetaD.<sup>27</sup> The height was 1.2 kJ/mol, the width 0.35 radians, hills were deposited every 120 fs, the



**Figure 3.** Convergence of MBMetaD bias compared to TTMetaD and WTMetaD. Figures on the left show convergence results for the estimated free energy difference between points  $(-1.25, 1.0)$  and  $(1.25, -1.0)$ , the two primary basins, while figures on the right show convergence results for the estimated free energy difference between points  $(-1.25, 1.0)$  and  $(2.1, -1.5)$ , the deeper basin and the lowest free energy barrier point between the two basins. Figures on top show the averaged free energy difference estimates as a function of time, while the lower figures show the standard deviation of the estimates as a function of time. All methods are as described in the text, with the statistics calculated over 32 runs per method. The MBMetaD methods in these figures used energy offsets of 42.5 kJ/mol (minimum referenced) and 15 kJ/mol (transition referenced). For these free energy difference estimates inside the metabasin domain, MBMetaD incurs no convergence accuracy penalty.



**Figure 4.** Comparison of the convergence of MBMetaD biases using different domain specifications. The MBMetaD methods in these figures used energy offsets of 42.5 kJ/mol (high energy minimum referenced), 35 kJ/mol (low energy minimum referenced), 15 kJ/mol (high energy transition referenced), and 7.5 kJ/mol (low energy transition referenced). The trueness of the MBMetaD methods are all fairly consistent, but the precision of transition-referenced MBMetaD is more robust with respect to choice of domain specification due to the early time cancellation of errors discussed in the text.

TTMetaD parameter  $\Delta T$  was  $2k_B T$ , and the transition wells for TTMetaD were  $(-1.25, 1.25)$  and  $(1.0, -1.25)$  in  $(\phi, \psi)$  CV space; for WTMetaD  $\Delta T$  was  $4 k_B T$ . When using metabasin metadynamics we compared the minimum-based domain definition with relative energy levels of 35.0 and 42.5 kJ/mol and transition-based domain definition with relative energy levels of 7.5 and 15.0 kJ/mol. We chose these values in order to roughly match the two methods based on the approximation that the apparent transition barrier relative to the minimum in these CVs is 27.5 kJ/mol; this makes for clear and direct comparison between the two different styles of domain specification. The higher energy values were chosen as a representative test of the method. The lower energy values were chosen to illustrate how the two domain definitions differ in behavior as the energy excess above the transition barrier becomes low and the PMF is barely flattened over the barrier between the transition wells.

The results for higher energy limits confirm that MBMetaD with either domain definition retains the accuracy and efficiency of TTMetaD in the low-energy regions it is designed to study

while giving up all resolution of the high-energy regions it is designed not to explore. Figure 2 shows the replicate-averaged estimated PMFs of regular TTMetaD and the two MBMetaD variants after 8 ns of simulation for comparison; it is evident that the low-lying contours are essentially indistinguishable among the three methods, while the high-energy contours present in the first are simply missing in the latter two. Also as expected, the contours outside of the domain but close enough to be biased are somewhat random and are not optimized over the course of convergence. Figure 3 compares the convergence rates for these methods on the domain of interest in terms of trueness and precision.

However, with lower energy limits, the transition-referenced domain definitions become clearly superior to minimum-referenced domains. This is shown in Figure 4. This is because in the normal case that CVs are imperfect, hidden barriers in HSVs can cause the bias to grow too high in one basin before exploring the next basin.<sup>9</sup> When this occurs in minimum-referenced MBMetaD, the energy differences from the minimum can therefore appear higher than they should and

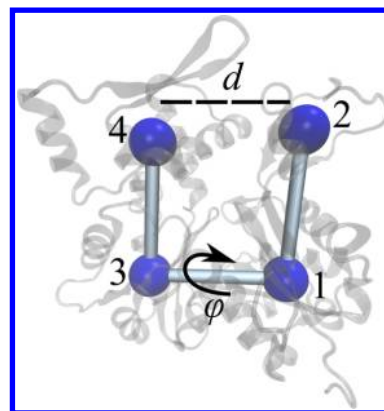
the estimated domains then become too small. If the intended domain was not much larger than the true metabasin of interest, the bias may not fully flatten all the way between the states of interest until the overestimation corrects itself. On the other hand, when similar overestimation occurs in transition-referenced MBMetaD, the transition barrier estimate also overshoots the true value, so that the domain is made larger by the same amount that the previous error would make it smaller. This systematic cancellation of errors ensures greater stability and better flattening before convergence and, thus better, less autocorrelated sampling and faster convergence. We always recommend using the transition-based rule for adaptive domain restriction when seeking to connect basins.

In these tests, the desired regions of interest are single large connected domains including both wells. In the high energy tests, the on-the-fly regions begin with large connected domains including both wells and small disconnected satellite domains due to sampling noise in the free energy estimate. As the simulation proceeds and the free energy estimate is refined, new satellites spring up, some disappear, and some connect with and absorb into the main domain. The behavior is the same in the low energy transition-referenced simulation as well. However, in the low energy minimum-referenced simulations, the regions often begin with one primary domain around the basin the simulation is initialized in. In these runs, when a transition to a new well occurs there is a delay before the umbrella sampling estimator registers the discovery of a new low-energy region, but then a domain begins to grow around the new basin. Depending on vagaries of sampling, these domains can either remain separate or join together before the simulation terminates. This reconnection is not efficient, so once again, we always recommend using the transition-based rule for adaptive domain restriction when seeking to connect basins.

**3.2. Actin Flattening.** Actin is a key component of the cytoskeleton,<sup>57</sup> but the structural bases and mechanisms of its regulation are only partially understood, even at the monomer level. The monomer is also called globular actin or G-actin, and aspects of it have been studied using metadynamics<sup>58,59</sup> and umbrella sampling.<sup>60</sup> However, G-actin is a dynamically sensitive protein with a high degree of allostery that can be difficult to design good CVs to investigate and can resist enhanced sampling analysis.<sup>61</sup>

A previous paper<sup>60</sup> used umbrella sampling to investigate nucleotide-dependent G-actin conformational distributions measured in terms of an interdomain distance and torsion (Figure 5) but was unable to reach convergence, presumably due to the presence of HSVs. Attempting to use metadynamics with the same torsion and distance CVs showed hysteresis, confirming the presence of HSVs, but this hysteresis typically involved unexpected and apparently irreversible behavior. Hypothesizing that the CVs were adequate in low-energy regions but inadequate in higher energy configurations, we applied MBMetaD to the problem in order to sample the lowest-lying metastable basins well and learn enough to refine the CVs before investigating further.

To do this, we performed 8 simulations of WTMetaD with and without adaptive domain restriction for 50 ns each. The simulations were set up as in previous work,<sup>60</sup> with proteins simulated in NAMD<sup>62</sup> patched with a customized version of PLUMED2<sup>49</sup> using the CHARMM22/27 force field with CMAP.<sup>63</sup> Our CVs were similar to those in the previous work, consisting of the 2-1-3-4 torsion between the cores of the four

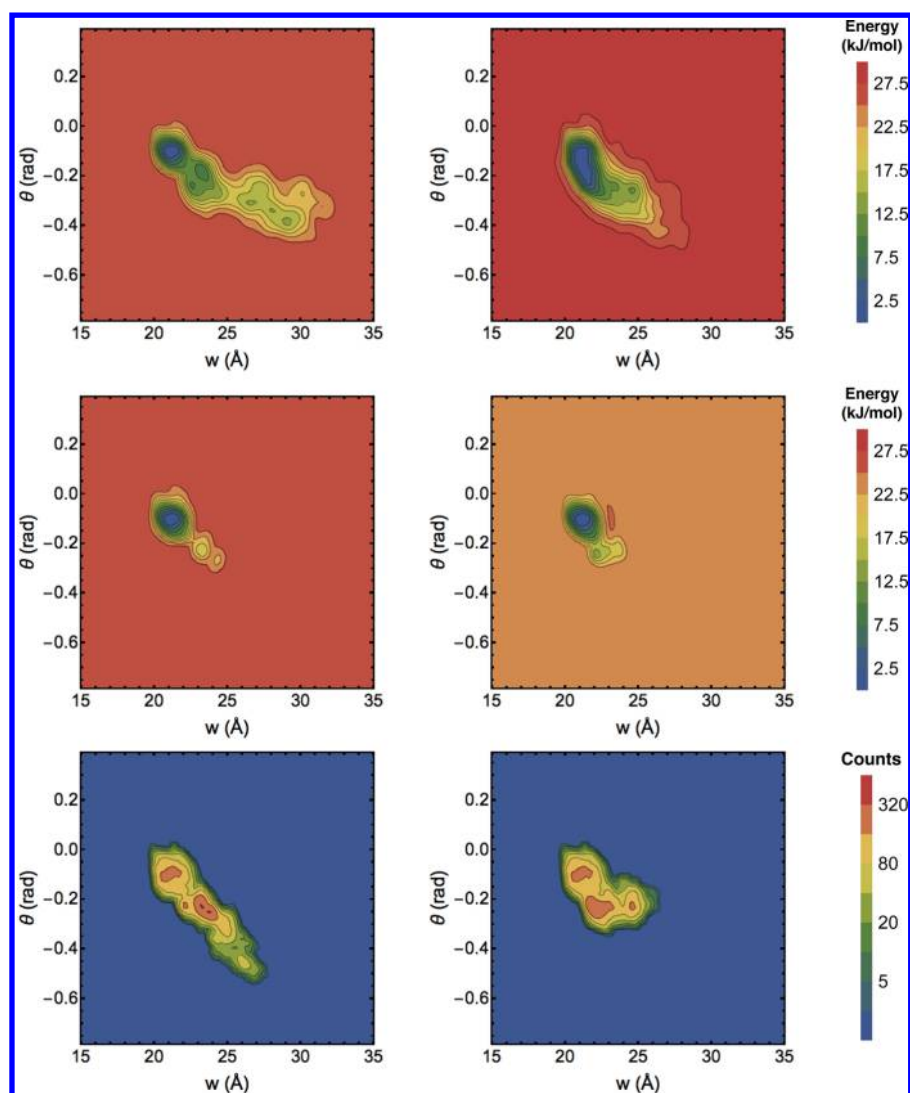


**Figure 5.** Schematic illustration of the CVs used to investigate nucleotide-dependent actin dynamics, consisting of the distance between subdomains 2 and 4 and the torsion angle between subdomains 2, 1, 3, and 4.

actin subdomains—the flatness of the monomer—and the distance between subdomains 2 and 4—the width of the nucleotide binding cleft. However, for the sake of computational efficiency the positions of these subdomains were calculated as the centers of mass of only the backbone and  $C_{\beta}$  carbons in the subdomain cores rather than the centers of mass of all of the subdomain core atoms. All simulations were initialized from a flattened conformation like that of filamentous actin (PDB ID 2ZWH).<sup>64</sup> Metadynamics added hills of initial height 0.004184 kJ/mol and widths 0.3 radians and 0.02 nm every 200 fs, with heights sequentially adjusted according to the WTMetaD rule with a bias factor of 10; the MBMetaD domain was defined as every point less than 25.104 kJ/mol above the current estimate of the free energy minimum and was updated every 200 ps if the exterior bias had increased by 2.092 kJ/mol since the last update. A minimum reference is appropriate because the simulations do not study a particular transition but rather the fluctuations about a single basin.

Representative biases and sampling histograms at the end of 50 ns are shown in Figure 6. It is evident that the MBMetaD rule successfully limits the bias to a smaller metastable basin of CV space than the unrestricted WTMetaD does without artificially restricting the sampling to the same basin. The regions selected by MBMetaD correspond to the biased regions. They consist of single primary domains around the starting states together with smaller disconnected domains that spring up wherever the simulation occupies another state long enough. However, the primary purpose of this test is to demonstrate greater simulation repeatability for MBMetaD compared to the WTMetaD reference, investigated in Figure 7.

This figure demonstrates that overlaps of the biases of different randomly initialized runs are greater in MBMetaD than WTMetaD, i.e., that the sampling is more similar across runs of MBMetaD than WTMetaD, just as desired. The disagreement between predictions of which regions of CV space are below a threshold energy shown in this figure correspond to one minus the ratio of the area of the intersection of the regions to the area of their union; expressed as an integral, this is  $1 - \int \min(I(F_i(s) < F_{th}), I(F_j(s) < F_{th})) ds / \int \max(I(F_i(s) < F_{th}), I(F_j(s) < F_{th})) ds$ , with  $F_i$  and  $F_j$  being the PMF estimates from runs  $i$  and  $j$ ,  $F_{th}$  being the threshold energy, and  $I(f(s) < g)$  being the indicator of the function  $f$  being less than scalar  $g$  at point  $s$ . This indicates greater stability and more physically meaningful—or at least more easily



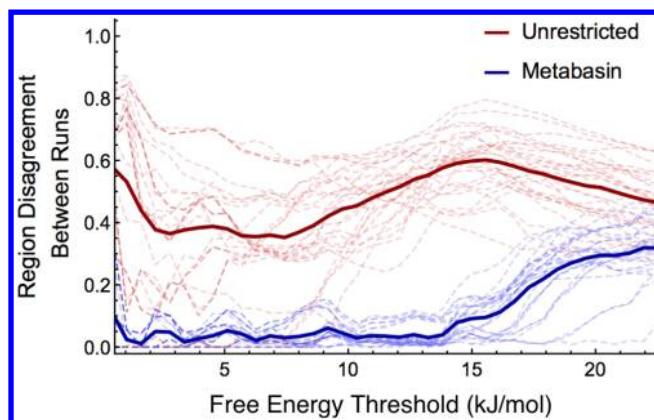
**Figure 6.** Representative sampling from unrestricted WTMetaD (top row) and well tempered MBMetaD (lower two rows) with the same initial conditions (columns) as a function of cleft width ( $w$ ) and twist angle ( $\theta$ ). In the unrestricted case only bias PMF estimates are shown because sampling and bias are directly related; for MBMetaD both bias PMF estimates (middle row) and sampling histograms (bottom row) are shown.

interpreted—exploration. The nonrepeatable free energy basin predictions seen with unrestricted WTMetaD do not correspond to states that appear biologically relevant or well-populated at equilibrium but rather appear to correspond to metastable traps found only because of overzealous non-equilibrium kicks. The MBMetaD data proved more useful for refining the choice of CVs and designing new simulations, but because refining CVs is a system specific process rather than a matter of MBMetaD methodology, describing that process will be left for future publication focused on the structure of G-actin. For our purposes here, it is sufficient to demonstrate that the method enhances sampling out of well-understood basins repeatably and without applying risky nonequilibrium driving forces in less well understood regions of CV space.

**3.3. CIC Antiporter Chloride Transport.** Our final example concerns the CIC-ec1  $\text{Cl}^-/\text{H}^+$  exchanger,<sup>65</sup> a paradigmatic protein of the ubiquitous CIC family.<sup>66–69</sup> In this example we will focus on investigating one particular part of the molecular mechanism of chloride transfer: motion of chloride anions in the channel when a key central acidic residue, E148, is protonated. The previous examples showed that MBMetaD is an accurate and effective self-limiting

mechanism when the CVs are appropriate in a metastable basin. However, it is often the case that CVs will not be ideal even in the domain of interest, and with this final example we intend to show that MBMetaD continues to have both desirable self-limiting behavior and dynamics with desirable hysteresis in that case. Therefore, in this case our CVs are simply the displacements along the membrane normal direction,  $z$ , from each of two chlorides, one closer to the cytosolic side of the membrane,  $z_{\text{cyt}}$  and the other closer to the extracellular side,  $z_{\text{ext}}$  to the center of mass of a central group of protein atoms that is chosen only as a stable reference for the protein frame of reference and has no other intended physical significance. This leaves motion of E148, water fluctuations, helix motion, and other protein environment fluctuations as potential HSVs.

We aim here to compare the function of MBMetaD with unrestricted WTMetaD in estimating PMFs of simultaneous translation of these chlorides throughout the protein, with the mechanism presumed to involve five key chloride states identified in previous structural studies.<sup>67,70,71</sup> First, when  $z$  is greater than 1 nm, the chloride is said to be extracellular. Second, when  $z$  is near 0.6 nm, the chloride is expected to be in



**Figure 7.** Figure showing the disagreement between the free energy estimates of different runs of like type. Disagreement ( $y$  axis) is measured as one minus the ratio of the area of the intersection of the two regions predicted by two different runs to be below a given free energy threshold ( $x$  axis) to the area of their union. All PMF estimates are standardized so that their minima are zero. Dashed lines represent the disagreement as a function of free energy threshold for individual pairs, while solid lines indicate their averages. The unrestricted WMetaD estimates disagree with one another more than the MBMetaD estimates at all free energy levels, showing much improved repeatability for the latter.

its external binding site. Third, when  $z$  is approximately  $-0.5$  nm, it is expected to be in its central binding site. Finally, when  $z$  is less than  $-1$ , this coordinate does not distinguish well between whether the chloride is in its internal binding site or unbound on the cytosolic side of the membrane.

The CIC model system was set up as a dimer of the wild-type CIC-ec1 structure (PDB: 1OTS) embedded into a lipid bilayer (163 POPE) and solvated with 11,000 TIP3P waters<sup>72</sup> in a  $92 \times 92 \times 79 \text{ \AA}^3$  box under periodic boundary conditions. Protein and lipids were modeled with the CHARMM27 force field;<sup>73,74</sup> long-ranged electrostatic interactions were treated with the Particle Mesh Ewald (PME) method,<sup>75</sup> and the cutoff distance for the short-ranged interactions from both the Lennard-Jones and the real-space Coulomb interaction was set to be  $12 \text{ \AA}$ . The classical MD simulation was performed using the Gromacs MD package<sup>54</sup> patched with a customized version of PLUMED<sup>49</sup> and a 2 fs time step. The initial configuration of the system was taken from our previous work,<sup>76</sup> where the system was first equilibrated for 7 ns in the NPT ensemble with a temperature of 310 K and a pressure of 1 atm and then further equilibrated for 10 ns in the NVT ensemble with a temperature of 300 K. Residues E113 in monomers A and B and D417 in monomer A were protonated as suggested by previous calculations,<sup>77</sup> and E148 in monomer A was also protonated to allow chloride translations from the external to the central site. All other residues were set to their default protonation states.

Metadynamics was applied using the CVs described above for the chloride ions in monomer A only, as previous experiments<sup>78</sup> showed that each monomer carries out  $\text{Cl}^-/\text{H}^+$  exchange independently. Monomer B was still included in the system and simulated as normal, not subject to any direct bias. Hills of initial height 0.2092 kJ/mol and widths 0.035 nm in both dimensions were added each 1 ps. For WMetaD runs, the well-tempered bias factor was 13, whereas for transition-tempered MBMetaD runs the transition-tempered bias factor was 5 using transition wells (0.45, 1.17) and (1.75, 0.12). MBMetaD domains were targeted to the transition barrier

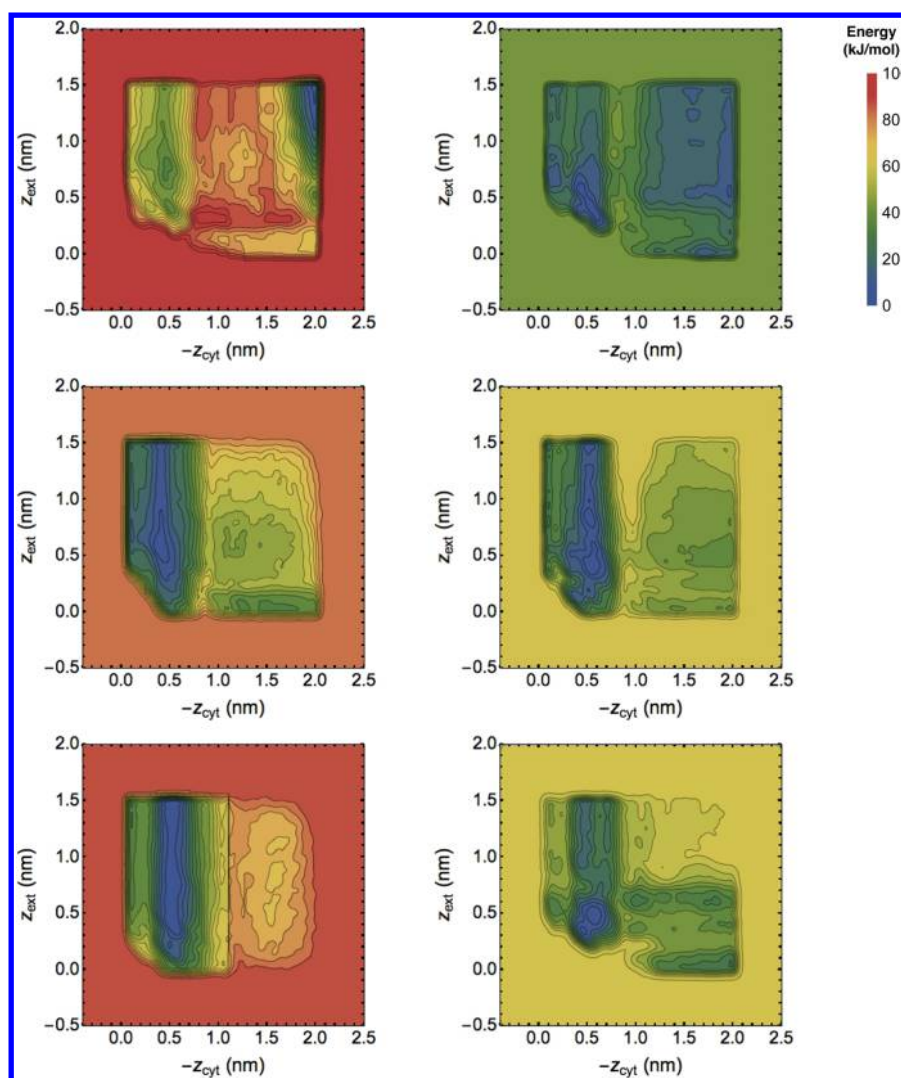
energy between those points plus 12 kJ/mol and were updated every 100 ps if the exterior bias level had changed by 1 kJ/mol. To restrict sampling to cases where the chlorides were in the protein in the appropriate locations in the protein interior, the coordinates of the extracellular and cytosolic-side chlorides were both restrained. The cytosolic-side chloride was restrained to the region  $-3.24 < x < -0.93$ ;  $-2.58 < y < -0.38$ ;  $0.09 < z < 2.00$ , while the extracellular-side chloride was restrained to the region  $-2.90 < x < 0.70$ ;  $-2.94 < y < 0.21$ ;  $-1.49 < z < 0.0$ . The restraints were implemented as half-harmonic walls with force constants of  $200 \text{ kJ}/(\text{mol \AA}^2)$ .

The free energy estimates given by three simulations using each method after various times, shown in Figure 8, do not correspond to converged potentials of mean force. They reflect the essential randomness of attempting enhanced sampling while neglecting key HSVs and show that MBMetaD prevents the nonequilibrium driving from spoiling the dynamics as it does so frequently in the unrestricted simulations. In the unrestricted case, the metastable band structure seen in the MBMetaD estimate of the starting basin is overshadowed by other features in two of the three WMetaD replicates.

The top left energy profile shows an unexpected and implausible prediction of a sharp well where both chlorides are outside of the protein that is much deeper than any of the expected binding sites, while the bottom left shows a sharp line near  $z_{\text{cyt}} = 1$  that is not seen in any other simulations and also does not correspond to any plausible physics: furthermore, the estimate beyond this line appears to simply be a blob without the expected basin structure for  $z_{\text{ext}}$ . In the simulation corresponding to the top unrestricted WMetaD figure, the trajectory shows that the cytosolic chloride becomes stuck inside the protein between helices that are normally firmly bound together. In the simulation corresponding to the bottom unrestricted WMetaD figure, one of the main helices unfolds and extends into the solvent: the chloride associates with the helix, and then the nonequilibrium bias ratchets the helix apart by pushing on the chloride, unfolding it turn by turn, before the chloride dissociates and goes into the solvent, with effectively permanent damage done.

Judging by the noticeably different energy scale, it is evident that MBMetaD has expected self-limiting behavior even in this highly challenging and poorly tuned test case, and judging by the fact that the MBMetaD runs do not exhibit the odd behavior seen in the others, it also appears that self-limiting behavior improves robustness to CV quality as predicted. However, it is also clear that the MBMetaD runs in the right column of Figure 8 do not fully agree among one another, showing that it does not cover up core deficiencies in the CVs that can be seen by comparing multiple independent replicates.

The disagreement occurs at the level of which regions of interest are selected on-the-fly, as well. The two topmost right panels substantially agree. However, in the bottom right panel, the final region of interest does not include the upper right quadrant of the allowed CV space. At the time of termination, the simulation was sampling the upper right region but not adding bias. Given time, the histogram correction would ensure that this region would properly register as part of the CV space to bias. Still, as seen in the alanine dipeptide example, it is better to use good CVs and a robust region definition than to rely on long-time convergence of the domains. This reflects the inefficiency of any CV-based enhanced sampling in the presence of HSVs.



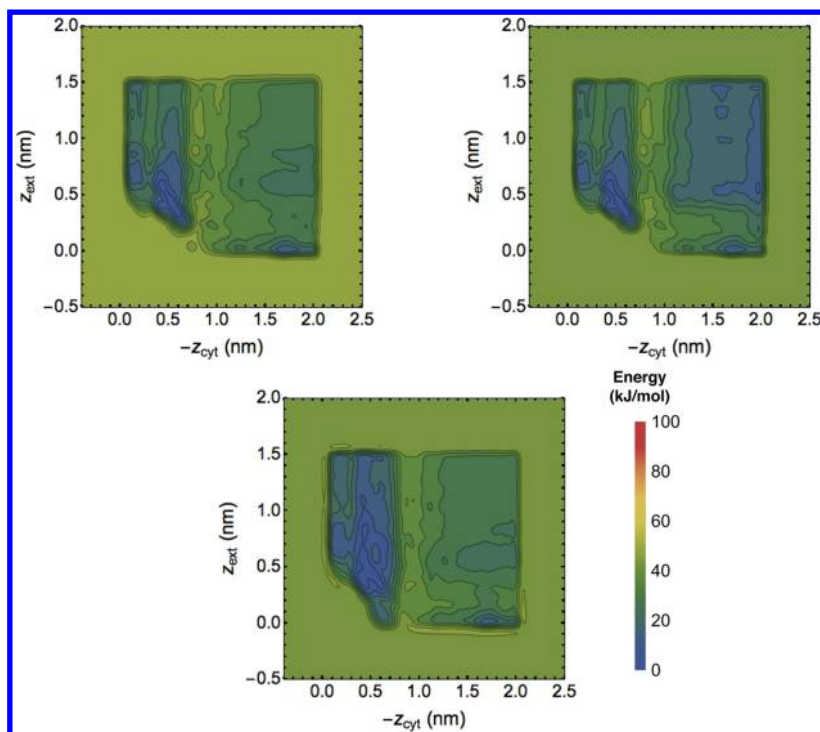
**Figure 8.** Bias-based free energy estimates from unrestricted WTMetaD (left column) and transition-tempered MBMetaD (right column) over three runs (rows), each run until features stabilized qualitatively or appeared to be irreparably incorrect. Clockwise from top left, these PMF estimates correspond to 250, 300, 300, 300, 350, and 400 ns of simulation. None appear fully converged, yet the latter are more physically plausible and the former show distinct signs of dynamics gone irrevocably astray in the top and bottom runs (see text for descriptions of the atomistic details behind these unphysical features).

Next, mimicking the usual process of checking whether one run might have luckily converged to the correct result despite the gauntlet of possible issues that evidently affected the other runs, we initialized another MBMetaD replicate of length 100 ns beginning with the bias and stored histogram of the top run just before 200 ns to check if the PMF estimate of this replicate would agree with the original copy of the top run after 300 ns. As Figure 9 makes evident, that is not the case. The original run and its branched replicate differ; the former samples configurations in which both chlorides are far from the center of the protein, whereas the other samples configurations in which the cytosolic-side chloride resides in the central pocket. The new replicate primarily explores the initial basin, whereas the original primarily explored the other basins. Motion between the basins is not fully facilitated by the bias because key HSVs are involved in the true barriers between them, and the bias reports that faithfully. MBMetaD correctly reports the presence of these HSVs even as it prevents the less physically interesting and implausible dynamics seen in the unrestricted

WTMetaD simulations. The domain restriction rule provides increased safety without covering up essential problems.

#### 4. CONCLUSIONS

The new method MBMetaD fundamentally eliminates one of the primary limitations of metadynamics, that adding energy indefinitely causes undesired and irreversible change in many sensitive dynamical systems, by providing an effective and convenient self-limiting mechanism that causes it to fill up to a flexibly defined free energy level and no farther—requiring no *a priori* estimate of barrier heights but rather estimating them on-the-fly as necessary. This allows for more focused study that should be especially practical when using many CVs at once and incidentally demonstrates a solution to the problem of boundary artifacts in metadynamics on domains of any shape in any number of dimensions. Most importantly, it means that using MBMetaD makes designing good CVs for metadynamics simpler. If energy is no longer added in regions of CV space that are irrelevant to transition mechanisms, the CVs no longer must be carefully tuned there.



**Figure 9.** Biased-based free energy estimates from a branched MBMetaD trajectory. The bias-based PMF estimate of the primary trajectory at 200 ns (left) is compared with estimates at 300 ns from direct continuation of that trajectory (middle) and continuation from a new random equilibrated initial conformation (right). The prominent differences between the two show that sampling remains highly autocorrelated on a 100 ns time scale even when biased, a hysteresis that properly indicates the presence of hidden barriers.

Furthermore, having a controlled energy level makes setting the hill height and the tempering rates for WMetaD and TMetaD simpler. Current practice requires users to choose their desired energy level by balancing the level of tempering with the initial rate of biasing;<sup>9,26</sup> but the tempering must also be matched to only partially known CV hysteresis time scales, and the biasing rate must also be matched to the only partially known rate of dissipation in the system. Thus, the parameters are overdetermined in a manner depending on unknowns with opaque trade-offs; using MBMetaD to disentangle the choice of final energy level from the choice of tempering rate and hill height can make the choices more straightforward. Additional technical complexity of the method for the implementer is therefore counterbalanced by decreased intuitive complexity for actual application.

By restricting our changes to the hill function alone, we ensure compatibility with many other advances in metadynamics methodology that cannot be discussed in depth here. For instance, the new procedure is compatible in principle with untempered,<sup>7</sup> well-tempered,<sup>26</sup> and transition-tempered<sup>27</sup> metadynamics, multiple walkers<sup>40</sup> and bias exchange,<sup>29</sup> driven metadynamics,<sup>42</sup> experimentally directed (ensemble-biased) metadynamics,<sup>39,79</sup> concurrent metadynamics,<sup>30</sup> multiple time steps,<sup>80</sup> and any choice of CV. It is not compatible with adaptive Gaussian hills<sup>32</sup> but is compatible with field-coordinate metadynamics.<sup>18</sup> As noted earlier, it is implemented in a public fork of the PLUMED2 package<sup>49</sup> and is available with a short guide upon request.

Moreover, there is little reason to think that this strategy applies only to metadynamics. All other adaptive enhanced sampling methods such as the adaptive biasing force<sup>11,12</sup> and orthogonal space random walk<sup>81</sup> approaches can similarly suffer hysteresis effects related to adding reckless driving forces when

their parameters are set too aggressively. It may be that adding region-limiting mechanisms like the one presented here may allow for more of those aggressive parameter choices to be used safely, with substantial potential efficiency gains. In each case, one could use the bias and an auxiliary histogram together to determine regions of interest on the fly based on the simple definitions we propose here, such as ‘everything less than 2 kT above the barrier between these two states’.

Finally, we find the physical underpinnings of MBMetaD unexpectedly natural, and we hope they will inspire further thought in the field. The new hills can be understood as approximate Green’s functions of diffusion inside the domain of interest plus an approximate quasistationary distribution in the exterior of the domain. Mathematical work on the foundations of accelerated dynamics<sup>82,83</sup> is finding deep theoretical power in using quasistationary distributions in lifting the Green’s function of a coarse-grained Markov process onto a fine-grained configuration space. Therefore, the MBMetaD hills appear to emerge as approximate mixed-resolution Green’s functions, connecting the worlds of Markov state modeling, accelerated dynamics, multiscale modeling, and CV-based adaptive enhanced sampling in a surprising way.

## AUTHOR INFORMATION

### Corresponding Author

\*E-mail: gavoth@uchicago.edu.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

This research was supported by the National Science Foundation through the Center for Multiscale Theory and

Simulation (NSF Grant CHE-1136709) and NSF Grant CHE-1465248, as well as by the National Institutes of Health (NIH Grant R01-GM053148) and by the Department of Energy through the LANL/LDRD program. Simulations were performed in part using resources provided by the University of Chicago Research Computing Center (Midway) and the San Diego Supercomputing Center (Comet) through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053575. The authors thank Sangyun Lee for his aid in preparing simulations of CIC.

## REFERENCES

- (1) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646–52.
- (2) Adcock, S. A.; McCammon, J. A. *Chem. Rev.* **2006**, *106*, 1589–615.
- (3) 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.
- (4) Kollman, P. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (5) Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. *J. Comput. Chem.* **2009**, *31*, 1569–1582.
- (6) Abrams, C.; Bussi, G. *Entropy* **2014**, *16*, 163–199.
- (7) Laio, A.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 12562–6.
- (8) Laio, A.; Gervasio, F. L. *Rep. Prog. Phys.* **2008**, *71*, 126601.
- (9) Barducci, A.; Bonomi, M.; Parrinello, M. *WIREs Comput. Mol. Sci.* **2011**, *1*, 826–843.
- (10) Dama, J. F.; Parrinello, M.; Voth, G. A. *Phys. Rev. Lett.* **2014**, *112*, 240602.
- (11) Darve, E.; Pohorille, A. *J. Chem. Phys.* **2001**, *115*, 9169.
- (12) Darve, E.; Rodríguez-Gómez, D.; Pohorille, A. *J. Chem. Phys.* **2008**, *128*, 144120.
- (13) Wang, F.; Landau, D. *Phys. Rev. Lett.* **2001**, *86*, 2050–2053.
- (14) Huber, T.; Torda, A. E.; van Gunsteren, W. F. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.
- (15) Branduardi, D.; Gervasio, F. L.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 054103.
- (16) Tribello, G. A.; Ceriotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17509–14.
- (17) Sutto, L.; Marsili, S.; Gervasio, F. L. *WIREs Comput. Mol. Sci.* **2012**, *2*, 771–779.
- (18) Tribello, G. A.; Ceriotti, M.; Parrinello, M. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 5196–201.
- (19) Lu, J.; Vanden-Eijnden, E. *J. Chem. Phys.* **2014**, *141*, 044109.
- (20) Wales, D. J. *J. Chem. Phys.* **2015**, *142*, 130901.
- (21) Torrie, G.; Valleau, J. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (22) Kästner, J. *WIREs Comput. Mol. Sci.* **2011**, *1*, 932–942.
- (23) Stecher, T.; Bernstein, N.; Csányi, G. *J. Chem. Theory Comput.* **2014**, *10*, 4079–4097.
- (24) Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. *J. Comput. Chem.* **1992**, *13*, 1011–1021.
- (25) Shirts, M. R.; Chodera, J. D. *J. Chem. Phys.* **2008**, *129*, 124105.
- (26) Barducci, A.; Bussi, G.; Parrinello, M. *Phys. Rev. Lett.* **2008**, *100*, 020603.
- (27) Dama, J. F.; Rotskoff, G.; Parrinello, M.; Voth, G. A. *J. Chem. Theory Comput.* **2014**, *10*, 3626–3633.
- (28) Bussi, G.; Gervasio, F. L.; Laio, A.; Parrinello, M. *J. Am. Chem. Soc.* **2006**, *128*, 13435–41.
- (29) Piana, S.; Laio, A. *J. Phys. Chem. B* **2007**, *111*, 4553–9.
- (30) Gil-Ley, A.; Bussi, G. *J. Chem. Theory Comput.* **2015**, *11*, 1077–1085.
- (31) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615–1621.
- (32) Branduardi, D.; Bussi, G.; Parrinello, M. *J. Chem. Theory Comput.* **2012**, *8*, 2247–2254.
- (33) Tiwary, P.; Parrinello, M. *J. Phys. Chem. B* **2015**, *119*, 736–742.
- (34) Rosta, E.; Hummer, G. *J. Chem. Theory Comput.* **2015**, *11*, 276–285.
- (35) Thiede, E.; Van Koten, B.; Weare, J. *ArXiv.org, ID 1410.1431*, 2014; pp 1–24.
- (36) Wojtas-Niziurski, W.; Meng, Y.; Roux, B.; Bernèche, S. *J. Chem. Theory Comput.* **2013**, *9*, 1885–1895.
- (37) Marsili, S.; Barducci, A.; Chelli, R.; Procacci, P.; Schettino, V. *J. Phys. Chem. B* **2006**, *110*, 14011–3.
- (38) Min, D.; Liu, Y.; Carbone, L.; Yang, W. *J. Chem. Phys.* **2007**, *126*, 194104.
- (39) White, A. D.; Dama, J. F.; Voth, G. A. *J. Chem. Theory Comput.* **2015**, *11*, 2451–2460.
- (40) Raiteri, P.; Laio, A.; Gervasio, F. L.; Micheletti, C.; Parrinello, M. *J. Phys. Chem. B* **2006**, *110*, 3533–9.
- (41) Singh, S.; Chiu, C.-c.; de Pablo, J. J. *J. Stat. Phys.* **2011**, *145*, 932–945.
- (42) Moradi, M.; Tajkhorshid, E. *J. Phys. Chem. Lett.* **2013**, *4*, 1882–1887.
- (43) Crespo, Y.; Marinelli, F.; Pietrucci, F.; Laio, A. *Phys. Rev. E* **2010**, *81*, 055701.
- (44) Baftizadeh, F.; Cossio, P.; Pietrucci, F.; Laio, A. *Curr. Phys. Chem.* **2012**, *2*, 79–91.
- (45) McGovern, M.; de Pablo, J. J. *J. Chem. Phys.* **2013**, *139*, 084102.
- (46) Bussi, G.; Laio, A.; Parrinello, M. *Phys. Rev. Lett.* **2006**, *96*, 090601.
- (47) Whitmer, J. K.; Fluiitt, A. M.; Antony, L.; Qin, J.; McGovern, M.; de Pablo, J. J. *J. Chem. Phys.* **2015**, *143*, 044101.
- (48) Cormen, T. H.; Leiserson, C. E.; Rivest, R. L. *Introduction to Algorithms*, 1st ed.; The MIT Press: Cambridge, MA, 1990; p 485.
- (49) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. *Comput. Phys. Commun.* **2014**, *185*, 604–613.
- (50) Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M. C.; Xiong, G.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.; Wang, J.; Kollman, P. *J. Comput. Chem.* **2003**, *24*, 1999–2012.
- (51) Berendsen, H.; van der Spoel, D.; van Drunen, R. *Comput. Phys. Commun.* **1995**, *91*, 43–56.
- (52) Lindahl, E.; Hess, B.; van der Spoel, D. *J. Mol. Model.* **2001**, *7*, 306–317.
- (53) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–18.
- (54) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.
- (55) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
- (56) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (57) Pollard, T. D.; Cooper, J. A. *Science* **2009**, *326*, 1208–12.
- (58) Pfaendtner, J.; Branduardi, D.; Parrinello, M.; Pollard, T. D.; Voth, G. A. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12723–8.
- (59) McCullagh, M.; Saunders, M. G.; Voth, G. A. *J. Am. Chem. Soc.* **2014**, *136*, 13053–13058.
- (60) Saunders, M. G.; Tempkin, J.; Weare, J.; Dinner, A. R.; Roux, B.; Voth, G. A. *Biophys. J.* **2014**, *106*, 1710–20.
- (61) Leone, V.; Marinelli, F.; Carloni, P.; Parrinello, M. *Curr. Opin. Struct. Biol.* **2010**, *20*, 148–154.
- (62) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. *J. Comput. Chem.* **2005**, *26*, 1781–802.
- (63) MacKerell, A. D.; Feig, M.; Brooks, C. L. *J. Comput. Chem.* **2004**, *25*, 1400–15.
- (64) Oda, T.; Iwasa, M.; Aihara, T.; Maéda, Y.; Narita, A. *Nature* **2009**, *457*, 441–445.
- (65) Accardi, A.; Miller, C. *Nature* **2004**, *427*, 803–7.
- (66) Maduke, M.; Miller, C.; Mindell, J. A. *Annu. Rev. Biophys. Biomol. Struct.* **2000**, *29*, 411–438.
- (67) Dutzler, R.; Campbell, E. B.; Cadene, M.; Chait, B. T.; MacKinnon, R. *Nature* **2002**, *415*, 287–294.
- (68) Jentsch, T. J. *Crit. Rev. Biochem. Mol. Biol.* **2008**, *43*, 3–36.
- (69) Stölting, G.; Fischer, M.; Fahlke, C. *Front. Physiol.* **2014**, *5*, 378.

- (70) Dutzler, R.; Campbell, E. B.; MacKinnon, R. *Science* **2003**, *300*, 108–112.
- (71) Feng, L.; Campbell, E. B.; Hsiung, Y.; MacKinnon, R. *Science* **2010**, *330*, 635–641.
- (72) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
- (73) 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.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (74) Feller, S. E.; MacKerell, A. D., Jr. *J. Phys. Chem. B* **2000**, *104*, 7510–7515.
- (75) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
- (76) Wang, D.; Voth, G. A. *Biophys. J.* **2009**, *97*, 121–131.
- (77) Faraldo-Gómez, J. D.; Roux, B. *J. Mol. Biol.* **2004**, *339*, 981–1000.
- (78) Robertson, J. L.; Kolmakova-Partensky, L.; Miller, C. *Nature* **2010**, *468*, 844–847.
- (79) Marinelli, F.; Faraldo-Gómez, J. *Biophys. J.* **2015**, *108*, 2779–2782.
- (80) Ferrarotti, M. J.; Bottaro, S.; Pérez-Villa, A.; Bussi, G. *J. Chem. Theory Comput.* **2015**, *11*, 139–146.
- (81) Zheng, L.; Chen, M.; Yang, W. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 20227–32.
- (82) Le Bris, C.; Lelièvre, T.; Luskin, M.; Perez, D. *Monte Carlo Methods Appl.* **2012**, *18*, 119–146.
- (83) Lelièvre, T.; Nier, F. *Anal. PDE* **2015**, *8*, 561–628.