

Markovian Weighted Ensemble Milestoning (M-WEM): Long-Time Kinetics from Short Trajectories

Dhiman Ray,* Sharon Emily Stone, and Ioan Andricioaei*

Cite This: *J. Chem. Theory Comput.* 2022, 18, 79–95

Read Online

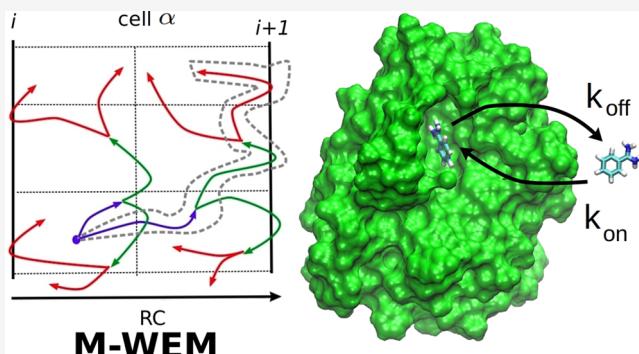
ACCESS |

Metrics & More

Article Recommendations

Supporting Information

ABSTRACT: We introduce a rare-event sampling scheme, named Markovian Weighted Ensemble Milestoning (M-WEM), which inlays a weighted ensemble framework within a Markovian milestoning theory to efficiently calculate thermodynamic and kinetic properties of long-time-scale biomolecular processes from short atomistic molecular dynamics simulations. M-WEM is tested on the Müller–Brown potential model, the conformational switching in alanine dipeptide, and the millisecond time-scale protein–ligand unbinding in a trypsin–benzamidine complex. Not only can M-WEM predict the kinetics of these processes with quantitative accuracy but it also allows for a scheme to reconstruct a multidimensional free-energy landscape along additional degrees of freedom, which are not part of the milestoning progress coordinate. For the ligand–receptor system, the experimental residence time, association and dissociation kinetics, and binding free energy could be reproduced using M-WEM within a simulation time of a few hundreds of nanoseconds, which is a fraction of the computational cost of other currently available methods, and close to 4 orders of magnitude less than the experimental residence time. Due to the high accuracy and low computational cost, the M-WEM approach can find potential applications in kinetics and free-energy-based computational drug design.



INTRODUCTION

It is a challenge to quantify with accuracy the kinetics of rare events in molecular biophysics via computational means. Molecular dynamics (MD) simulations provide atomistically detailed movies of the structural and functional dynamics of biological macromolecules. However, the majority of important dynamic processes in the cell involve broad length and time scales. A large fraction of such processes are rare over the time scale of the simulation. Energy barriers higher than thermal energy trap the simulated system in conformational basins of attraction, impeding proper sampling of all relevant states. Examples of rare processes include protein folding, conformational transitions, ligand binding, and unbinding, which in most cases involve $\sim 10^4$ – 10^6 atoms including the natural solution environment. Despite phenomenal advances in computing hardware, atomistic MD simulations of such large systems still go, typically, up to multiple microseconds only. This is many orders of magnitude smaller than the time scale relevant to biological function, which is often in the range of seconds to hours.

The current study focuses on protein–ligand interactions; its adequate sampling is pivotal to computer-aided drug design. Molecular dynamics (MD) simulations provide mechanistic insights into such interactions at atomistic detail and are one of the essential tools in the repertoire of the pharmaceutical research community. A wide range of alchemical free-energy

calculation methods^{1–3} and enhanced sampling methods (involving external biasing force)^{4–7} have been developed over the past few decades to calculate the binding free energy of a protein–ligand complex. Although the virtual screening of potential inhibitors is currently based on the binding free energy, the efficacy of a drug molecule is often dependent on the binding and unbinding kinetics or the residence time.^{8,9} It is difficult to compute the kinetic properties from traditional enhanced sampling simulations, as the dynamics become nonphysical due to the application of artificial biases (although there are methods to recover kinetics from simple constant force or constant velocity steered molecular dynamics (SMD)^{10–13}). On the other hand, using brute force MD simulation, one needs to sample multiple binding and unbinding events to obtain converged results for kinetic properties. This requires a simulation time many times higher than the time scale of one event, which itself is beyond the reach of even the most powerful supercomputers. This results in a dire need to develop

Received: August 9, 2021

Published: December 15, 2021



theoretical methods and computational algorithms to make quantitative predictions about the kinetics of long-time-scale processes such as rare events from short-time-scale trajectories.

A category of methods involves transition path sampling (TPS), a concept introduced by Pratt¹⁴ and later developed by Chandler and co-workers,^{15,16} to simulate transitions across energy barriers. Instead of applying external bias, path sampling methods utilize the statistical properties of the unbiased trajectory ensemble to compute experimental observables such as the kinetics of conformational transition or ligand unbinding, as well as molecular-scale properties like ligand release pathways and mechanisms.¹⁷ A different path sampling approach is the weighted ensemble (WE) method of Huber and Kim,¹⁸ which belongs to a broader category of variance-reduction algorithms that use “splitting” in the framework of Monte Carlo sampling (see, e.g., Kahn¹⁹). The WE method was further developed by Zuckerman, Chong, and collaborators (see, e.g., ref 20); it was also established that the weighted ensemble is statistically exact.²¹ In this approach, the conformational space between the initial and final states is discretized into multiple bins and a number of short trajectories are propagated from the starting bin. Trajectory segments are split or merged when they enter a new bin to keep an equal number of trajectories in each bin. Appropriate weights are assigned to the new set of trajectories to conserve the total probability. It allows for the sampling of fast moving but low-weight trajectories that reach the final state well before the mean first passage time (MFPT); this facilitates the calculation of converged kinetics, free energy, and pathways at a relatively low computational cost. With the implementation in the open source software WESTPA,²² the weighted ensemble method has seen a wide range of applications including folding and conformational transitions in proteins,^{23–25} formation of host–guest complexes,²⁶ protein–peptide²⁷ and protein–protein binding,²⁸ ion permeation through protein channels,²⁹ viral capsid assembly,³⁰ and many others. Many new variants, as well as new analysis schemes for the traditional WE approach, have emerged in recent years, including WExplore,³¹ resampling of ensembles by variation optimization (REVO),³² history augmented Markov state modeling (haMSM),³³ the RED scheme,³⁴ minimal adaptive binning (MAB),³⁵ and microbin analysis.³⁶ Particularly, the WExplore and REVO algorithms have been successfully applied to study the pathways and kinetics of protein–ligand dissociation,^{32,37,38} even for systems with residence times as high as seconds to minutes.^{39,40}

Another popular approach to study the kinetics of biophysical rare events is milestoning,^{41–43} which belongs to the larger category of trajectory stratification.^{44–47} In milestoning, multiple interfaces are placed along a reaction coordinate (RC), and short MD trajectories are propagated in between the interfaces, which thus serve as milestones for the progress of the transition of interest. Analyzing the milestone-to-milestone transition statistics via a statistical framework,⁴³ the kinetics and free-energy profile are estimated. This method has also been implemented in the software tool miles,⁴⁸ ScMile⁴⁹ and simulation enabled estimation of kinetic rates (SEEKR),⁵⁰ and has been used to study a variety of complex biological problems including protein allosteric transitions,⁵¹ membrane permeation by small molecules,^{52–55} protein small molecule interaction,^{56–58} simple ligand–receptor binding,⁵⁹ peptide transport through protein channels,^{60,61} DNA protein interaction,⁶² protein conformational dynamics,⁶³ etc. Apart from the necessity of having a predefined reaction coordinate, the milestones need to be placed far apart to preserve the

assumption of Markovianity.⁴² This itself increases the computational cost significantly, leaving aside the fact that two independent studies have shown that the majority of the total computational effort in milestoning simulation is spent on sampling along the milestone interfaces to generate starting structures in accordance with the equilibrium distribution.^{60,64}

A different variant of the milestoning approach has been developed: Markovian Milestoning with Voronoi Tesselation (MMVT),^{55,65} which removes the necessity of performing additional sampling along the milestone interface, reducing the overall computational cost to a large extent. The application of MMVT remained rather limited, being used primarily for studying small molecule transport through transmembrane proteins,^{66–69} substrate translocation through ATPase motor,⁷⁰ and the CO entry in myoglobin.⁷¹ Only recently, the Markovian milestoning approach has been tested on ligand–receptor binding for crown-ether host–guest complexes and for the dissociation of a benzamidine ligand from the trypsin protein.⁶⁴ Despite cutting down the computational cost in sampling at the milestone interface, this approach still suffers from the Markovian assumption and can be significantly expensive for complex systems.⁶⁴

In our previous work, we attempted to improve the milestoning scheme by accelerating transitions between distant milestones via the application of directed wind forces.⁷² This technique did increase the number of energetically uphill transitions, but the statistical properties of the computed observables were not significantly better.⁷³ More recently, we proposed the combined weighted ensemble milestoning (WEM) scheme, where we performed WE simulations in between milestones to accelerate the convergence of the transition between adequately spaced milestones.⁷⁴ The WEM method not only produced an accurate prediction of kinetics, free energy, and time correlation function for small molecular systems like alanine dipeptide,⁷⁴ but we could also reproduce protein–ligand binding and unbinding rate constants and binding affinity, previously obtained from a 30 μ s equilibrium simulation,⁷⁵ in less than 100 ns of WEM simulation.⁷⁶

Yet, the current methodology and the implementation of WEM have a few drawbacks. First, the sampling of the degrees of freedom perpendicular to the reaction coordinate (RC) is significantly poor, particularly in situations where slow conformational changes of the protein are coupled to the ligand unbinding.⁷⁶ This can potentially be rectified using multiple starting states on the milestone interface sampled from long umbrella-sampling simulations at the expense of a manifold increase in the computational cost similar to traditional milestoning. Second, the choice of the milestoning reaction coordinate (RC) is arbitrary and can possibly impact the quality of the results, depending on the complexity of the underlying free-energy landscape. Moreover, a major hindrance of the large-scale application of the WEM technique is the complexity of the simulation protocol, which requires propagating many short trajectories and stopping them upon reaching a nearby milestone.⁷⁶ It requires frequent monitoring of the trajectory as well as frequent communication to the dynamics engine to stop the propagation if the progress coordinate reaches a particular value; this makes the WEM algorithm particularly inefficient to implement in graphical processing unit (GPU) hardware.

We, thereby, present a novel Markovian Weighted Ensemble Milestoning (M-WEM) approach, in which we combine weighted ensemble with soft-wall⁶⁵-based Markovian Mile-

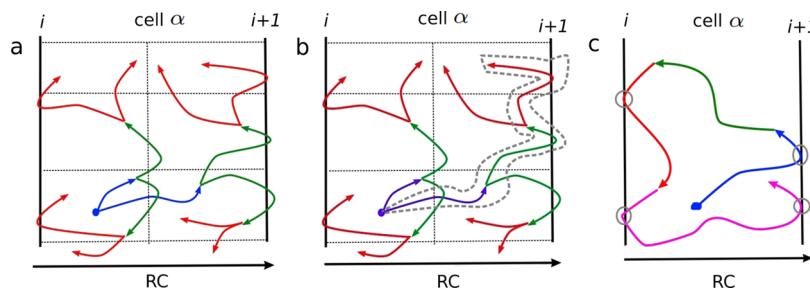


Figure 1. Schematic representation of the M-WEM simulation protocol. The thick lines indicate milestones (labeled as milestone index i and $i + 1$). The dotted lines indicate WE bin boundaries (which are adapted during the simulation, but in this figure we show fixed bins for clarity). Trajectories for different WE iterations are shown in different color schemes: iteration 1: blue, iteration 2: green, iteration 3: red, and iteration 4: pink. (a) First, WE simulation is performed, with harmonic walls placed at the milestone interfaces allowing for the trajectory to bounce back and forth. (b) The propagation history of individual trajectories is traced back from the last iteration (an example trace is highlighted with the gray dashed line). (c) The milestone crossing events (gray circles) are recorded from each trace and are used in subsequent analysis.

stoning, in an attempt to mitigate the deficiencies and improve the performance of the weighted ensemble milestone technique. We first provide a detailed description of the theory of Markovian milestone and the M-WEM approach. We then show the application of this method to the two-dimensional (2D) Müller–Brown potential, the conformational transition of alanine dipeptide, and the dissociation and association of the trypsin–benzamidine complex, a protein–ligand system with a residence time beyond millisecond. The choice of the trypsin–benzamidine complex is inspired by the fact that many existing path sampling and enhanced sampling methods have been applied to this system, including Markov state modeling (MSM),^{77,78} metadynamics,⁷⁹ adaptive multilevel splitting (AMS),⁸⁰ milestone,⁵⁰ MMVT,⁶⁴ WExplore,³⁸ and REVO.³² So, we compare the accuracy of the results and the performance of M-WEM with these existing techniques, as well as with the experimental rate constants and free-energy values obtained by Guillain and Thusius.⁸¹ We also discuss a new approach to construct multidimensional free-energy landscapes via post-analysis of MMVT and M-WEM trajectories obtained using a one-dimensional reaction coordinate, with a potential application in systems where orthogonal degrees of freedom are strongly coupled with the reaction coordinate.

THEORY

Markovian Milestoning. The theoretical details of the Markovian milestone with Voronoi tessellation (MMVT) approach are described elsewhere.^{55,64,65} Here, we provide only a brief description relevant to the current work.

In MMVT, the configurational space is discretized into Voronoi cells. A flat bottom potential is applied to each cell with half-harmonic walls placed at each milestone interface, preventing the trajectories from escaping out of the Voronoi cells. For a one-dimensional reaction coordinate, used here, the flat bottom potential has the expression

$$\tilde{V}_\alpha = \begin{cases} \frac{1}{2}\tilde{k}(x - x_{i+1}^\alpha)^2, & \text{if } x > x_{i+1}^\alpha \\ 0, & \text{if } x_i^\alpha < x < x_{i+1}^\alpha \\ \frac{1}{2}\tilde{k}(x - x_i^\alpha)^2, & \text{if } x < x_i^\alpha \end{cases} \quad (1)$$

where α is the cell index, x_i^α is the value of the reaction coordinate at the milestone i at the boundary of the cell α , and \tilde{k} is the force constant; the total number of cells is Λ and the total

number of milestones is M . One or more unbiased trajectories are propagated in each cell. The trajectories that cross the milestone interface are reflected back into the cell by the half-harmonic restraint. As a result, the trajectories remain confined into one cell and perform many transitions between the milestone interfaces constituting the boundaries of the cell. The portions of the trajectory outside the cell are to be discarded before performing further analysis. This protocol is referred to as the soft-wall restraint,^{65,66} which we adopt in the current work. Alternatively, a hard wall restraint^{55,64} can also be used where the direction of velocity is switched when a trajectory crosses a milestone.

From these confined trajectories, the transition counts between milestones are recorded. A flux matrix $\mathbf{k} \in \mathbb{R}^{\Lambda}$ is constructed whose elements are given by

$$k_{\alpha,\beta} = \frac{N_{\alpha,\beta}}{T_\alpha} \quad (2)$$

where $N_{\alpha,\beta}$ is the number of transitions from cell α to cell β recorded from a trajectory propagated for time T_α in the cell α . The equilibrium probability for each cell (π_α) is obtained by iteratively solving the linear equation (eq 3) in a self-consistent manner under the constraint of a constant total probability of one

$$\sum_{\beta=1, \beta \neq \alpha}^{\Lambda} \pi_\beta k_{\beta,\alpha} = \sum_{\beta=1, \beta \neq \alpha}^{\Lambda} \pi_\alpha k_{\alpha,\beta}; \quad \sum_{\alpha=1}^{\Lambda} \pi_\alpha = 1 \quad (3)$$

The free-energy profile at each cell is then computed as

$$G_\alpha = -k_B T \ln(\pi_\alpha) \quad (4)$$

For calculating kinetics, the transition matrix $\mathbf{N} \in \mathbb{R}^{M \times M}$ and the lifetime vector $\mathbf{R} \in \mathbb{R}^M$ are constructed, whose elements are computed as follows

$$\begin{aligned} N_{ij} &= T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \frac{N_{ij}^\alpha}{T_\alpha} \\ R_i &= T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \frac{R_i^\alpha}{T_\alpha} \end{aligned} \quad (5)$$

where N_{ij}^α is the number of times a trajectory in cell α collides with milestone j after having last visited milestone i ; R_i^α is the cumulative time the trajectory spends in cell α visiting milestone i and before reaching any other milestone. T is a constant for

dimensional consistency, which is not necessary to compute because it cancels out at a later stage. A rate matrix $\mathbf{Q} \in \mathbb{R}^{M \times M}$ is then defined as

$$Q_{ij} = \begin{cases} \frac{N_{ij}}{R_i}, & \text{if } R_i \neq 0; i \neq j \\ 0, & \text{if } R_i = 0; i \neq j \end{cases}$$

$$Q_{ii} = -\sum_{i \neq j} Q_{ij} \quad (6)$$

Considering milestone M is the target milestone, the mean first passage time of the process can be computed as

$$\hat{\mathbf{Q}}\tau^M = -\mathbf{1} \quad (7)$$

$\hat{\mathbf{Q}} \in \mathbb{R}^{M-1 \times M-1}$ is the matrix obtained by deleting the last row and column of \mathbf{Q} . $\mathbf{1}$ is a unit vector with $M-1$ elements, and $\tau^M \in \mathbb{R}^{M-1}$ is the vector with entries τ_i^M that are the MFPTs from milestone i to milestone M .

Markovian Weighted Ensemble Milestoning (M-WEM). In the current work, we introduce the Markovian Weighted Ensemble Milestoning (M-WEM) approach, where the conventional MD trajectories in the Markovian milestoning framework are replaced by weighted ensemble simulation. A schematic representation of the M-WEM protocol is depicted in Figure 1. WE bins are placed along the reaction coordinate in between the milestone interfaces, as well as along a different coordinate to accelerate sampling along the milestone interface. The additional non-RC coordinate should ideally be locally orthogonal to the RC, but this is not a necessary condition. WE simulation is performed in this 2D progress coordinate space using the recently developed minimal adaptive binning (MAB) scheme.³⁵ As opposed to the traditional fixed-binning scheme, the MAB approach adaptively changes the bin boundaries during the course of simulation, avoiding the requirement of an arbitrarily chosen predefined set of bins. It also provides an increased sampling of the conformational space. As the total number of occupied bins remains virtually unchanged throughout the simulation, the maximum amount of computational resources needed for the simulation can be easily estimated beforehand.³⁵ A stochastic integrator, e.g., for the Langevin equation, is needed to propagate the dynamics to ensure that the new set of trajectories, generated after a splitting event, follow different paths despite emerging from a single parent trajectory.

Unlike the MMVT approach with conventional MD, the WE trajectories hitting the milestones will have different weights. To properly take into account this effect, we take all of the trajectory segments at the last iteration and trace them back to the first iteration to obtain separate trajectory traces. The weight of each trajectory trace is set equal to the weight of the corresponding trajectory segment in the last iteration.

The total number of trajectory traces in cell α (M_α) is equal to the number of occupied bins \times the number of trajectories per bin in the final iteration. The elements of the flux matrix \mathbf{k} in this formalism are given by

$$k_{\alpha,\beta} = \sum_{j=1}^{M_\alpha} w_j \frac{N_{\alpha,\beta}^J}{T_\alpha^J} = \sum_{j=1}^{M_\alpha} w_j k_{\alpha,\beta}^J \quad (8)$$

where w_j is the weight of the J th trajectory trace. $N_{\alpha,\beta}^J$, T_α^J and $k_{\alpha,\beta}^J$ have similar definitions as in eq 2 except that they are computed

just from the J th trace. The equilibrium probability distribution and the free-energy profile are computed from the elements of the flux matrix obtained from eq 8 using eqs 3 and 4, respectively.

For calculating kinetics, the N_{ij} and the R_i matrix elements are to be constructed taking into account the different weights of the trajectory traces. The new transition matrix element becomes

$$N_{ij} = T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \sum_{j=1}^{M_\alpha} w_j \frac{N_{ij}^{\alpha,J}}{T_\alpha^J} = T \sum_{\alpha=1}^{\Lambda} \sum_{j=1}^{M_\alpha} w_j \pi_\alpha \frac{N_{ij}^{\alpha,J}}{T_\alpha^J} \quad (9)$$

where $N_{ij}^{\alpha,J}$ has the same definition as N_{ij}^α in eq 5 except it is for the J th trajectory trace. Now, we define a pseudo-transition matrix $\tilde{\mathbf{N}}^{\alpha,J}$, which is identical to the transition matrix \mathbf{N} but computed only from one (J th) trajectory trace in the cell α . Its elements are given by

$$\tilde{N}_{ij}^{\alpha,J} = \pi_\alpha \frac{N_{ij}^{\alpha,J}}{T_\alpha^J} \quad (10)$$

Now, the total transition matrix \mathbf{N} can be obtained as

$$\mathbf{N} = T \sum_{\alpha=1}^{\Lambda} \sum_{j=1}^{M_\alpha} w_j \tilde{\mathbf{N}}^{\alpha,J} \quad (11)$$

Similarly, the lifetime vector \mathbf{R} is given by

$$\mathbf{R} = T \sum_{\alpha=1}^{\Lambda} \sum_{j=1}^{M_\alpha} w_j \tilde{\mathbf{R}}^{\alpha,J}$$

where $\tilde{R}_i^{\alpha,J} = \pi_\alpha \frac{R_i^{\alpha,J}}{T_\alpha^J}$

$$(12)$$

and $R_i^{\alpha,J}$ is equivalent to R_i^α but computed only for the J th trajectory segment. Using the \mathbf{N} and \mathbf{R} , the rate matrix \mathbf{Q} is computed using eqs 6 and 7.

Error Analysis. Error analysis of milestoning-based simulations can be performed in a few different ways, primarily by generating an ensemble of rate matrices (\mathbf{Q}). Then, the desired properties (such as the MFPT) are calculated from many sample matrices and the uncertainty is estimated. When working with transition matrices (as in traditional milestoning), the kernels can be sampled from a β distribution.⁶⁰ Similar to our previous work,^{74,76} we generated the ensemble of rate matrices by sampling from a Bayesian-type conditional probability,^{56,82,83} given by

$$p(\mathbf{Q} | \{N_{ij}, R_i\}) = \prod_i \prod_{j \neq i} Q_{ij}^{N_{ij}} \exp(-Q_{ij} N_i R_i) P(\mathbf{Q}) \quad (13)$$

where $p(\mathbf{Q})$ is a uniform prior, N_{ij} is the number of trajectories transiting from milestone i to j , and $N_i = \sum_{j \in \mathbb{M}} N_{ij}$, where \mathbb{M} is the set of all milestones. We sampled the \mathbf{Q} matrices from the distribution in eq 13 using a nonreversible element exchange Monte Carlo scheme.⁸⁴ One randomly chosen off-diagonal element and the diagonal element of the corresponding row of \mathbf{Q} are updated to generate a new rate matrix \mathbf{Q}'

$$Q'_{ij} = Q_{ij} + \Delta \quad (i \neq j)$$

$$Q'_{ii} = Q_{ii} - \Delta \quad (14)$$

where Δ is a random number sampled from an exponential distribution of range $[-Q_{ij}, \infty]$ with mean zero. The new matrix \mathbf{Q}' is accepted with a probability equal to $\min(1, p_{\text{accept}})$, where

$$p_{\text{accept}} = \frac{p(\mathbf{Q}' | \{N_{ij}, R_i\})}{p(\mathbf{Q} | \{N_{ij}, R_i\})} = \left(\frac{Q_{ij} + \Delta}{Q_{ij}} \right)^{N_{ij}} \exp(-\Delta N_i R_i) \quad (15)$$

As each step modifies only one element, the sampled matrices are highly correlated.⁸⁴ So, we only considered the matrices sampled every 50 steps for uncertainty estimation. We recomputed the MFPT obtained from them using eq 7 and calculated the mean and 95% confidence intervals for all MMVT simulations.

For M-WEM simulations, estimating errors using non-reversible element exchange Monte Carlo can be difficult. The expression for p_{accept} (eq 15) has an exponential dependence on the number of transitions N_{ij} and number of trajectories reflected from milestone i (N_i). In the case of MMVT simulations, all such transitions have equal weight, unlike in M-WEM where the weights of transiting trajectories can vary to a large extent, often over many orders of magnitude. So, using the expression in eq 15, which counts all transitions with equal importance, one is unable to sample the ensemble of transition matrices accurately in our M-WEM approach. For each observable (MFPT, free energy, etc.), we computed its estimate over an M-WEM simulation of an increasing number of total iterations for each cell. We then compute the mean and 95% confidence interval of the sampled observables after the simulation is converged. The uncertainties obtained using this procedure capture the effect of the fluctuations of an observable around the mean after the simulation has converged. Such an approach is commonly used in the literature to compute the error estimates of free-energy profiles, as discussed, for example, in our previous work.⁸⁵ A more elaborate version of this technique is called block averaging,⁸⁶ which is a statistically rigorous way to compute error estimates in molecular dynamics simulations.⁸⁷

Specific details of error analysis for each test system are mentioned in the Computational Methods section. For the M-WEM technique, the derivation of a more rigorous approach for error analysis, similar to that described in eqs 13–15, will be addressed in future work.

Efficiency Analysis. Because of the difference in the uncertainties of the measured values from different methods, a direct comparison between the simulation time or convergence time may not always correctly represent the relative efficiency between different approaches. To address this issue, we used the algorithmic efficiency metric, η , which takes into account the variance of the data apart from the total simulation time.⁷⁴ This metric is evaluated as

$$\eta = \left(\frac{A}{\Delta A} \right)^2 N_s^{-1} \quad (16)$$

where A is the measured quantity and ΔA is the uncertainty of the measurement. N_s is the number of force evaluations performed to calculate A (although it can be thought of as the simulation time if expressed in units of time instead of time steps). In brief, η^{-1} is the number of force evaluations required to measure our observable A with an uncertainty equal to its mean.¹⁸ The lower the value of η^{-1} , the better the efficiency. For the first two systems, the Müller–Brown potential and the

alanine dipeptide, we reported the value of η^{-1} where we calculated the error bars from multiple runs. We did not report η^{-1} for each individual run because the error of MMVT calculations, computed using Bayesian analysis, and the error bars of the M-WEM approach computed from averaging over iterations are not directly comparable. For the same reason, we do not report an algorithmic efficiency for the trypsin–benzamidine system, for which η will also depend on the quantity we are measuring.

Calculation of k_{off} and k_{on} . For the protein–ligand binding problem considered in this paper, we computed the unbinding rate constant (k_{off}) as

$$k_{\text{off}} = \frac{1}{\tau} \quad (17)$$

where τ is the ligand residence time, which is equivalent to the MFPT of transitioning from the bound state milestone to the unbound state milestone and which, in turn, can be computed from the milestones or M-WEM framework described above.

We noted in our previous work that the ligand binding kinetics is often diffusion dominant.⁷⁶ So, the rate-determining step can be the arrival of the ligand on the outermost milestone surface, rather than going from the outermost milestone to the binding pocket. (Whether this is true, for a specific system, should be carefully evaluated on a case-by-case basis.) In our previous work, we delineated an analytical method to combine diffusion theory with milestones transition kernel integration to compute the binding rate constant k_{on} .^{76,77} We begin with the expression of the diffusion-dependent arrival rate of a small molecule on the surface of a sphere of radius r ⁸⁸

$$k(r) = 4\pi D r \quad (18)$$

where D is the diffusion constant of the ligand in water. We make two modifications to this expression. First, we assume that the only conformations that can lead to binding are in the space explored by the ligand in the outermost milestone. So, we scale the rate by the surface coverage factor α . The concept of coverage factor has been described more elaborately in our previous work.⁷⁶ To mention it only briefly here, α is the fraction of the surface area of the outermost milestone interface that is accessible to the ligand. Details of the calculation of α in this work are slightly different from our previous work. Instead of performing a Monte Carlo simulation, here we create a 2D grid on the interface of the outermost milestone based on the solid angles. Then, we calculate how many of those bins are occupied by the center of mass of the ligand in the outermost milestones cell. The coverage factor α is then computed as

$$\alpha = \frac{\sum \Delta\theta \Delta\phi \sin \theta \delta_{\text{occ}}(\theta, \phi)}{4\pi} \quad (19)$$

where $\delta_{\text{occ}}(\theta, \phi) = 1$ if the surface element $\Delta\theta \Delta\phi$ contains the angular coordinate of the ligand, and zero otherwise.

Second, we consider that a ligand can only reach the bound state if it moves toward the protein from the outermost milestone surface. So, we further scale the arrival rate by $K_{M,M-1}$, the transition probability of going from the last (M th) milestone to the previous milestone. However, as the transitional kernel is not directly available in the current Markovian milestones framework, we compute this inward transition probability from the N matrix

$$K_{M,M-1} = \frac{N_{M,M-1}}{N_{M,M-1} + N_{M,M+1}} \quad (20)$$

As $N_{M,M+1}$ becomes undefined when M is the last milestone, we perform this analysis with the last but one milestone. Including these factors, the flux of binding trajectories through that milestone becomes

$$k(r) = 4\pi D r \alpha K_{M,M-1} \quad (21)$$

from which k_{on} is calculated in $\text{M}^{-1} \text{s}^{-1}$ units by multiplying the flux with Avogadro's number N_{av}

$$\begin{aligned} k_{\text{on}} &= 4\pi D r \alpha K_{M,M-1} N_{\text{av}} \\ &= 7.569 \times 10^{13} D r \alpha K_{M,M-1} \end{aligned} \quad (22)$$

In the last step, we used the diffusion constant of small molecule in $\text{cm}^2 \text{s}^{-1}$ unit and the r is provided in Å.

Reconstruction of Free-Energy Landscape. The trajectories confined in the different Voronoi cells can be used to construct a higher-dimensional free-energy landscape for both Markovian milestones and M-WEM simulations. The trajectory data is first histogrammed in appropriate collective variables to obtain different independent histograms for each individual cell Λ . A weighted sum of these histograms are then performed to obtain the equilibrium high-dimensional probability distribution

$$p(\mathbf{R}) = \sum_{\alpha=1}^{\Lambda} \pi_{\alpha} p^{\alpha}(\mathbf{R}) \quad (23)$$

where $p(\mathbf{R})$ is the probability density along the collective variable space \mathbf{R} , and $p^{\alpha}(\mathbf{R})$ is the same obtained from the histogram only in the cell α . In the case of M-WEM, the $p^{\alpha}(\mathbf{R})$ is obtained from multiple (M_{α}) trajectories of different weights as

$$\begin{aligned} p^{\alpha}(\mathbf{R}) &= \sum_{j=1}^{M_{\alpha}} w_j p_j^{\alpha}(\mathbf{R}) \\ p(\mathbf{R}) &= \sum_{\alpha=1}^{\Lambda} \pi_{\alpha} \sum_{j=1}^{M_{\alpha}} w_j p_j^{\alpha}(\mathbf{R}) \end{aligned} \quad (24)$$

where $p_j^{\alpha}(\mathbf{R})$ is the histogram of the j th trajectory in cell α in the CV space \mathbf{R} . The free-energy landscape is then reconstructed using

$$G(\mathbf{R}) = -k_B T \ln(p(\mathbf{R})) \quad (25)$$

We demonstrate this approach for alanine dipeptide in the Results section.

Committor. The committer of a point in a conformational space is the probability of a trajectory starting from that point to reach the final state before visiting the initial state.⁸⁹ Recent work by Elber et al. established that it is possible to calculate the committer at the milestone interfaces.⁴⁸ However, to the best of our knowledge, such an approach has not been applied to Markovian milestones so far. We performed the committer calculation in the following way. First, we constructed a transition kernel (\mathbf{K}), equivalent to conventional milestones, from the \mathbf{N} matrix

$$K_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \quad (26)$$

The transition kernel \mathbf{K} is then modified with a boundary condition, which ensures that the flux through the final state will remain "absorbed" there and will not return to the previous milestones. This is achieved by replacing the last row of the \mathbf{K} matrix with zeros, except for the element corresponding to the last milestone for which the value is 1.^{48,90} For a three-milestone model, this can be illustrated as

$$\mathbf{K} = \begin{bmatrix} 0 & K_{12} & 0 \\ K_{21} & 0 & K_{23} \\ 0 & K_{32} & 0 \end{bmatrix} \rightarrow \mathbf{K}_C = \begin{bmatrix} 0 & K_{12} & 0 \\ K_{21} & 0 & K_{23} \\ 0 & 0 & 1 \end{bmatrix} \quad (27)$$

The vector containing the committer values of each milestone, \mathbf{C} , is then calculated as

$$\mathbf{C} = \lim_{n \rightarrow \infty} (\mathbf{K}_C)^n \mathbf{e}_p \quad (28)$$

where \mathbf{e}_p is a unit vector, all elements of which are zero except for the one corresponding to the final milestone. Multiple powers of \mathbf{K}_C are computed numerically until the committer converges. The results are considered to be converged when the change in the norm of the \mathbf{C} vector is less than 10^{-3} .

COMPUTATIONAL METHODS

We tested the Markovian Weighted Ensemble Milestoning approach on a toy model system of 2D Müller–Brown potential, conformational transition in alanine dipeptide, and the millisecond time-scale protein–ligand unbinding in the trypsin–benzamidine system. In the first two systems, we performed long equilibrium simulation and Markovian milestones simulation to compare with our M-WEM results.

Müller–Brown Potential. The two-dimensional Müller–Brown potential⁹¹ is defined as

$$U(x, y) = h \sum_{i=1}^4 A_i \exp[a_i(x - x_{0,i})^2 + b_i(x - x_{0,i}) \\ (y - y_{0,i}) + c_i(y - y_{0,i})^2] \quad (29)$$

where $A \in \{-200, -100, -170, 15\}$, $a \in \{-1, -1, -6.5, 0.7\}$, $b \in \{0, 0, 11, 0.6\}$, $c \in \{-10, -10, -6.5, 0.7\}$, $x_0 \in \{1, 0, -0.5, -1\}$, $y_0 \in \{0, 0.5, 1.5, 1\}$, and $h = 0.04$. This system has a nonlinear transition path with a barrier height about $4.5 k_B T$. For the purpose of this model, we set $k_B T = 1$.

To obtain a benchmark of the kinetics, a long overdamped Langevin dynamics simulation was propagated starting from $(x, y) = (-0.5, 1.5)$, which corresponds to the minimum A in Figure 2. The minimum B is chosen to be the target state. The simulation was propagated for 10^7 time steps, capturing 347 back and forth transitions. The free-energy landscape obtained from this equilibrium trajectory is depicted in Figure 2. The mean first passage time (MFPT) to go from minima A to B is depicted in Table 1.

For milestones simulations, eight milestones are placed at $y = \{-0.3, 0.0, 0.3, 0.6, 0.9, 1.2, 1.5, 1.8\}$ at equal intervals. The reaction coordinate is chosen to be parallel to the y axis. This poor choice of RC was made intentionally to represent realistic situations where the arbitrarily chosen empirical RCs are used to study complex biomolecular processes with many coupled degrees of freedom. MFPTs were computed for a transition from $y = 1.5$ to 0.0 for both the milestones-based methods.

For Markovian milestones (MMVT) simulations, an overdamped Langevin dynamics simulation is propagated in 2D in all

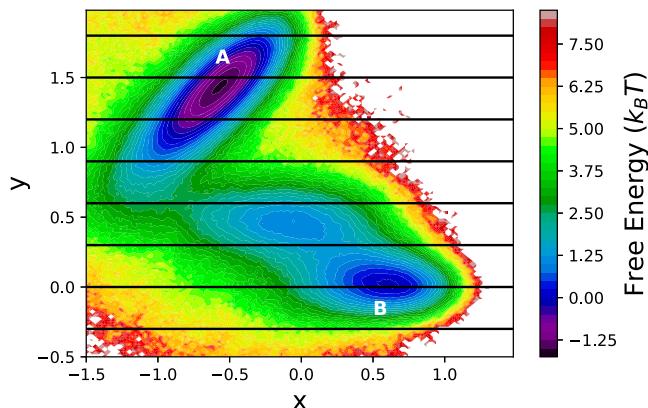


Figure 2. Free-energy landscape of the Müller–Brown potential explored using 10^7 steps of overdamped Langevin dynamics simulation. The positions of the milestones, used in MMVT and M-WEM calculations, are shown in black lines. The two minima relevant to this study are marked as A and B.

Table 1. Results of Conventional Langevin Dynamics, MMVT, and M-WEM Simulations for the Müller–Brown Potential

method	MFPT ($\times 10^3$ steps)	simulation time ($\times 10^6$ steps)	η^{-1} ($\times 10^6$ steps)
regular overdamped LD	25.2 ± 2.9	10	0.13
MMVT (trial 1)	12.3 ± 0.4	10.5	
MMVT (trial 2)	13.1 ± 0.3	10.5	
MMVT (trial 3)	13.2 ± 0.5	10.5	
MMVT (average) ^a	12.9 ± 1.2		0.09
M-WEM (trial 1)	22.2 ± 3.2	2.6	
M-WEM (trial 2)	17.3 ± 1.6	2.6	
M-WEM (trial 3)	28.2 ± 4.6	2.6	
M-WEM (average) ^a	22.6 ± 13.6		0.94

^aComputed from the three independent trials.

of the seven cells in the spacing between eight milestones. A half-harmonic wall is applied at both ends of the cell (milestones) with a force constant $\tilde{k} = 1000$ to confine the trajectories within the cell. Each simulation is propagated for 1.5×10^6 steps in each cell. Transition statistics between each milestone pair is computed using the method described in the *Theory* section, and the MFPT has been computed.

For the M-WEM method, the procedure remained identical to the MMVT approach, except weighted ensemble simulation was performed in between the milestone interfaces as opposed to conventional dynamics. The 2D adaptive binning (MAB)³⁵ was employed along the x and y directions with 5 bins per dimension with 4 trajectory segments per bin. This led to 33 bins in total, including the additional bins in each direction for the most forward, backward, and the bottleneck trajectories.³⁵ A total of 300 iterations of WE simulation are performed in each cell, with a recycling interval of 10 steps. The transition rate matrices and MFPTs were computed every 10 iterations of WE simulation in each cell between iterations 260 and 300. The mean and error estimates were performed on the five sampled data points for the MFPT values between iterations 260 and 300.

Alanine Dipeptide. The conformational transition of alanine dipeptide was simulated in the gas phase. The system was set up following the protocol described by Wei and Elber.⁴⁹ The 22-atom system is modeled using the CHARMM22 force

field⁹² with a 10 Å cutoff distance for the interatomic interactions. A time step of 0.5 fs was used, and all of the bonds between heavy atoms and hydrogen were constrained using the SHAKE algorithm.⁹³ All simulations were performed using the NAMD 2.13 package⁹⁴ with the colvars⁹⁵ module. The conformational change can be described adequately in the 2D coarse space of two backbone torsion angles Φ and Ψ . Half-harmonic walls with a mild force constant ($0.04 \text{ kcal mol}^{-1} \text{ deg}^{-2}$) are placed at a value of ± 175 for both the Φ and Ψ angles to avoid transitions along the edges of the free-energy surface (i.e., to remove periodicity).⁴⁹ This will ensure that we observe the transition along the center of the free-energy map.

The barrier height for the conformational transition of gas-phase alanine dipeptide is very high ($> 10 \text{ kcal mol}^{-1}$), and it is very difficult to observe direct transitions at room temperature. Following the earlier work,⁴⁹ we performed the simulations at 600 K temperature, which allowed us to sample 285 transitions between the two free-energy minima in a 500 ns equilibrium simulation.

For both MMVT and M-WEM simulations, the reaction coordinate was chosen to be the Φ dihedral angle and milestones were placed at $\Phi = -80, -60, -40, -20, 0, 20, 40, 60$, and 80° . The initial and final states were chosen to be the milestones at $\Phi = -80$ and 60° . In the case of MMVT simulation, 5 ns of conventional MD simulation was propagated in each cell confined between the two consecutive milestones, leading to a total computational effort of 40 ns. (The trajectories were extended to 10 ns with no difference in results. So, the result of 5 ns simulation is reported.) A force constant of $4 \text{ kcal mol}^{-1} \text{ deg}^{-2}$ was applied in the harmonic walls placed at the milestones to confine the trajectories in between the milestones. The portion of the trajectories outside the cell has been removed prior to further analysis.

In the case of M-WEM simulations, 5 WE bins were placed for each cell in Φ and Ψ coordinates leading up to 33 bins in total including separate bins for the forward, backward, and the bottleneck trajectories. Four trajectory segments were propagated in each occupied bin. The progress coordinate values were recorded at a very frequent interval (10 fs) to record the time of milestone crossings as accurately as possible. A total of 100 iterations of WE simulation are performed in each cell, with a recycling time of 1 ps. The transition rate matrices and MFPTs were computed every 2 iterations of WE simulation in each cell between iterations 2 and 10 and every 10 iterations between iterations 10 and 100 to monitor the convergence of the results. The convergence plots and related discussion are provided in the *Supporting Information*. The mean and error estimates were performed on the five sampled data points for the MFPT values between iterations 60 and 100.

Trypsin–Benzamidine Complex. The system setup for the trypsin–benzamidine complex is identical to the work by Votapka et al.⁵⁰ The structure, parameter, and topology files were obtained from the authors of ref 50. We point the reader to their original publication⁵⁰ for more details. To mention briefly, the atomic coordinates of the protein–ligand complex were obtained from Protein Data Bank (PDB) ID: 3PTB.⁹⁶ The protonation states of ASP, GLU, and HIS residues were determined at pH 7.7, which was used in this study to replicate the experimental condition.⁸¹ The protein was modeled using the AMBER ff14SB force field,⁹⁷ and generalized Amber force field (GAFF)⁹⁸ parameters were used for the ligand. The structure was solvated in a truncated octahedron box of TIP4Pew⁹⁹ water molecules, and eight Cl⁻ ions were added to

neutralize the system. Overall, the system contains $\sim 23\,000$ atoms. All MD simulations were performed using the NAMD 2.14b2 package⁹⁴ with a time step of 2 fs. A Langevin integrator with a damping coefficient of 5 ps⁻¹ was used to keep the temperature constant at 298 K. A Langevin piston was used to maintain the pressure at 1 atm.

The bound state structure was equilibrated for 10 ns in NPT ensemble. From the end point of this simulation, the ligand was pulled out of the binding pocket using a 10 ns steered molecular dynamics (SMD) simulation. The reaction coordinate (RC) description is identical to the previous work,⁵⁰ i.e., the center of mass distance between the benzamidine ligand and the C_α atoms of the following residues near the binding pocket: 190, 191, 192, 195, 213, 215, 216, 219, 220, 224, and 228 (numbered according to PDB: 3PTB). During the SMD simulation, a moving harmonic restraint of 1 kcal mol⁻¹ Å⁻² was applied on the RC with a pulling velocity of ~ 1.5 Å ns⁻¹. The collective variables were biased and monitored using the colvars module.⁹⁵ Representative structures for seeding the milestones simulations were sampled from the SMD trajectory.

Concentric spherical milestones were placed at the following values of the RC: 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 5.0, 6.0, 8.0, 10.0, 12.0, and 13.0 Å. These values are similar to previous studies^{50,64} except for a few additional milestones, as we were unable to observe energetically uphill transitions otherwise. The separation between milestones should be such that the transition time scales between one milestone to the other should be larger than the decay time of the velocity autocorrelation function of the RC. We checked this condition in our system, as discussed in detail in the *Supporting Information*.

The following distinction is worth noting at this point. As in Markov state modeling (which assumes the formalism of continuous-time Markov chains), the transitions between conformational macrostates¹⁰⁰ are to be Markovian. However, the dynamics inside the macrostates delimited by the milestones may not necessarily be Markovian, and it is for that dynamics that we check for the decay of velocity autocorrelation. Milestoning theory is built upon two key assumptions: the reaction channels are localized and the committor can be represented as a function of the reaction coordinate only. A sufficient condition to make these assumptions valid is that the dynamics be overdamped. To ensure this condition, milestones need to be placed sufficiently far from each other such that the time scale of transitions between milestones is higher than the decorrelation time of the velocity along the reaction coordinate.⁸³

A total of 12 cells were constructed in the spacing between 13 milestones. For each cell, a flat bottom potential (eq 1) is applied with a force constant of 100 kcal mol⁻¹ Å⁻² for the harmonic walls present at the milestones. First, the representative structure (sampled from SMD) is equilibrated at the center of the cell for 1 ns by restraining the RC via a harmonic potential. The force constant was gradually increased to 500 kcal mol⁻¹ Å⁻² over the first 500 ps and kept constant over the last 500 ps. From the end point of the 1 ns equilibration, weighted ensemble (WE) simulations were propagated for 300 iterations with a recycle time δt of 2 ps. A two-dimensional MAB scheme was used for the binning. The two progress coordinates were the RC and the root-mean-square deviation (RMSD) of the ligand with respect to the representative structure (sampled from SMD) corresponding to the specific cell. The progress coordinates were recorded using the colvars module.⁹⁵ The total computational cost of the M-WEM simulation was approximately 734 ns. The

simulation was stopped at multiple points at an interval of 10 WE iterations between 30 and 300 iterations for each cell. For each set, the trajectory traces were computed using which equilibrium probabilities, free-energy profiles, and MFPTs between the first milestone (at 1 Å) and the last milestone (at 13 Å) (residence time) were computed. This allowed us to monitor the convergence of residence time over the course of the simulation. The unbinding rate constant k_{off} was calculated as the inverse of residence time. Free-energy profile, binding rate constant k_{on} , and committors were calculated following the procedure described in the *Theory* section. The error bars for all quantities were calculated from the last five iterations sampled (i.e., iterations 160–200 for values reported at iteration 200 and iterations 260–300 for values reported at iteration 300). Before any calculation, the probabilities of the Voronoi cells are modified ($\pi_\alpha \rightarrow \tilde{\pi}_\alpha$) to take into account the Jacobian factor appearing due to the different surface area of milestones with different radius: $\tilde{\pi}_\alpha = \pi_\alpha r_\alpha^2 / \sum_\alpha^\Delta \pi_\alpha r_\alpha^2$ where r_α is the radius of the cell α , which we choose to be the radius of a sphere equidistant from the two milestones surrounding the cell.

RESULTS

Müller–Brown Potential. For the two-dimensional toy model of Müller–Brown potential, we performed three independent trials for both MMVT and M-WEM simulations and the results are presented in Table 1. The MFPT of the transition from milestone at $y = 1.5$ to 0.0, computed using the M-WEM approach, shows a quantitative agreement with the MFPT of the transition from minimum A to minimum B in the regular overdamped Langevin simulation. The results of MMVT simulation are off by a factor of ~ 2 . Although the simulation time for MMVT and regular MD was comparable, the M-WEM simulations produced converged results with ~ 4 times less computational expense. The algorithmic efficiency of M-WEM, on the contrary, is poorer compared to that of the MMVT approach because of the larger variance of the MFPTs obtained from the M-WEM method. Although the computational gain is not significant in the case of this low-dimensional model system, these results serve as a proof of concept of our method in the rare-event sampling problem. It also indicates that despite the choice of a poor and simplistic RC, accurate MFPTs can be calculated using the M-WEM method.

Alanine Dipeptide. Next, we tested the performance of MMVT and M-WEM methods on the conformational transition of alanine dipeptide. The results were compared to a 500 ns conventional MD simulation. The free-energy landscape along the Φ and Ψ torsion angles for the gas-phase alanine dipeptide (obtained from equilibrium MD simulation) is shown in Figure 3. The mean first passage time (MFPT) of transition from milestone $\Phi = -80^\circ$ to milestone $\Phi = 60^\circ$ is in agreement with the MFPT of transition from free-energy minima A to B obtained from long equilibrium MD simulation (Table 2). The M-WEM results show slightly better agreement, but the difference is not very significant; in fact, the error bars of the MMVT and M-WEM simulations obtained from independent runs overlap with each other. Both these methods produced accurate results within 1 order of magnitude less computational cost in comparison to the equilibrium MD. Although the M-WEM simulations took about twice as much computational effort as the MMVT simulation for full 100 iterations, the MFPT results converged as early as in 20–30 iterations (see the *Supporting Information*). Moreover, going from the 2D model of the Müller potential to the molecular model of alanine

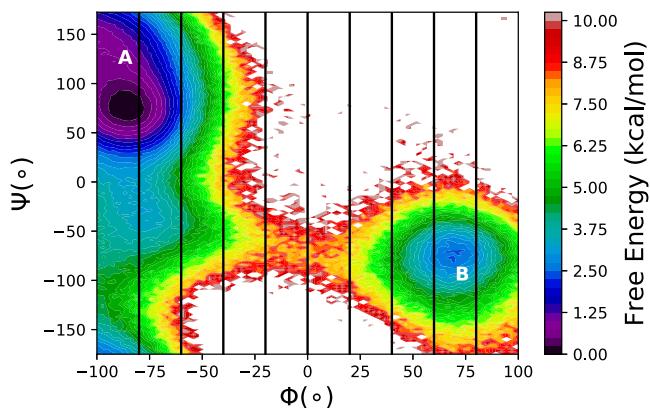


Figure 3. Free-energy landscape of the gas-phase alanine dipeptide along the Φ and Ψ torsion angles from a 500 ns equilibrium MD simulation. The positions of the milestones, used in MMVT and M-WEM calculations, are shown in black lines. The two conformations relevant to this study are marked as A and B.

Table 2. Results of Conventional MD, MMVT, and M-WEM Simulations for the Alanine Dipeptide

method	MFPT (ps)	simulation time (ns)	η^{-1} (ns)
regular MD ^a	1583 ± 188	500	7.05
MMVT trial 1	1465 ± 167	40	
MMVT trial 2	951 ± 13	40	
MMVT trial 3	984 ± 8	40	
MMVT (average) ^b	1133 ± 715		15.93
M-WEM trial 1	1690 ± 112	86.4	
M-WEM trial 2	1290 ± 69	84.4	
M-WEM trial 3	1286 ± 123	83.2	
M-WEM (average) ^b	1422 ± 576		13.78

^aFrom state A to B in Figure 3, not from milestone to milestone.
^bComputed from the three independent trials.

dipeptide, we see a significant improvement of the algorithmic efficiency as the η^{-1} of M-WEM is now slightly lower than that of MMVT. Nonetheless, both have poorer efficiency than regular; the latter yields a tighter confidence interval from ~ 300 transitions, whereas MMVT or M-WEM uncertainties are estimated from three independent runs.

The committer values at milestone interfaces for all three trials of M-WEM calculation are depicted in Figure 4. The

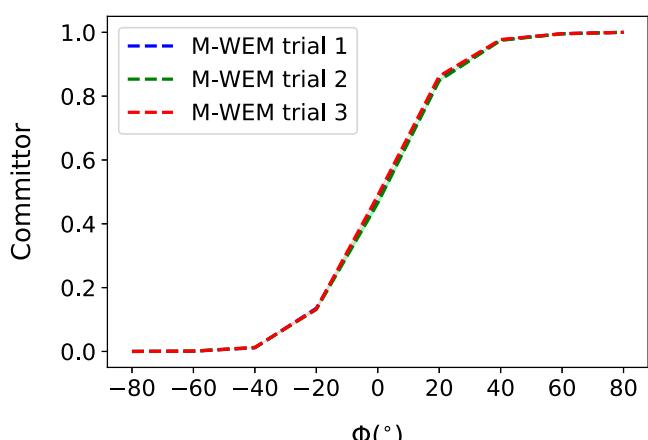


Figure 4. Committor values as a function of the milestone coordinate Φ for the alanine dipeptide system.

results from different trials are in excellent agreement with each other, and all of them show a committer value of ~ 0.5 for the milestone at $\Phi = 0^\circ$. A committer value of 0.5 indicates the transition state (TS). The milestone at $\Phi = 0^\circ$ is indeed present on top of the free-energy barrier aka TS as evident from the free-energy landscape in Figure 3.

We applied our free-energy reconstruction protocol to recover the free-energy landscape along the Φ and Ψ degrees of freedom. The crude probability distribution for each cell is obtained by histogramming the M-WEM and MMVT simulation data projected on those two degrees of freedom. This unscaled distribution for individual cells ($p^\alpha(\Phi, \Psi)$) as obtained from the M-WEM calculation is shown in Figure 5a. Then, the true probability distribution ($p(\Phi, \Psi)$) is computed by rescaling the distributions corresponding to each cell with the weight of their probabilities obtained using eq 3 (Figure 5b)

$$p(\Phi, \Psi) = \sum_{\alpha=1}^{\Lambda} \pi_\alpha p^\alpha(\Phi, \Psi) \quad (30)$$

The summation is over all Λ cells, and π_α is the equilibrium probability of each cell.

This rescaled probability distribution is used to reconstruct the free-energy landscape ($G(\Phi, \Psi)$) for the conformational transition of alanine dipeptide

$$G(\Phi, \Psi) = -k_B T \ln \left(\sum_{\alpha=1}^{\Lambda} \pi_\alpha p^\alpha(\Phi, \Psi) \right) \quad (31)$$

The reconstructed free-energy surfaces for both MMVT and M-WEM simulations are in excellent agreement with one obtained from a 500 ns conventional MD simulation but confined only between the initial and final milestones, i.e., $-80^\circ < \Phi < 80^\circ$ (Figure 6). This provides a way to study the free-energy landscape of orthogonal degrees of freedom, which are coupled with the RC but are not taken into consideration while devising the milestone progress coordinate.

Trypsin–Benzamidine Complex. Finally, we applied the M-WEM approach to calculate the kinetics and free energy for a protein–ligand binding and unbinding problems. We chose the system of the trypsin–benzamidine complex because of primarily two reasons. First, this system is studied extensively using MD simulations with various enhanced sampling and path sampling methods. Moreover, the residence time of the ligand is in the millisecond regime, which is beyond the reach of the currently available computational power. Benzamidine is also a very potent ligand, with an experimental binding affinity (K_d) of $(1.2 \pm 0.1) \times 10^{-5}$ M.⁸¹ This is a challenging enough test system for the M-WEM method and can also determine the utility of our approach in computer-aided drug design.

The ligand residence time, unbinding rate constant (k_{off}), binding rate constant (k_{on}), and the binding free energy (ΔG_b) have been computed from the M-WEM simulation, and the results are compared with the SEEKR⁵⁰ and MMVT SEEKR⁶⁴ results (which used identical simulation conditions as our work) and also with the experimental data⁸¹ (Table 3). All values obtained from M-WEM are in quantitative agreement with the experimental data. (We reported two sets of results for M-WEM, one after 200 iterations and another after 300 iterations.) The k_{off} value, predicted from the M-WEM simulation, is within the error bars of the experiment and within 1 order of magnitude of the SEEKR and MMVT SEEKR results. The same holds for residence time, which is the inverse of k_{off} . Our k_{on} results are

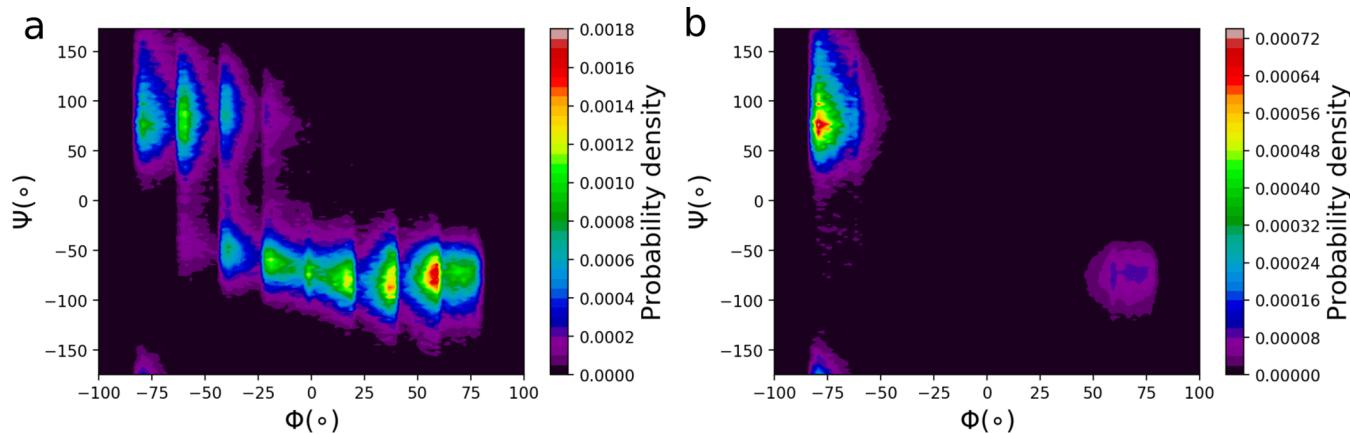


Figure 5. Reconstruction of equilibrium probability distribution (b) from raw unscaled probability distribution from M-WEM trajectories in each cell (a).

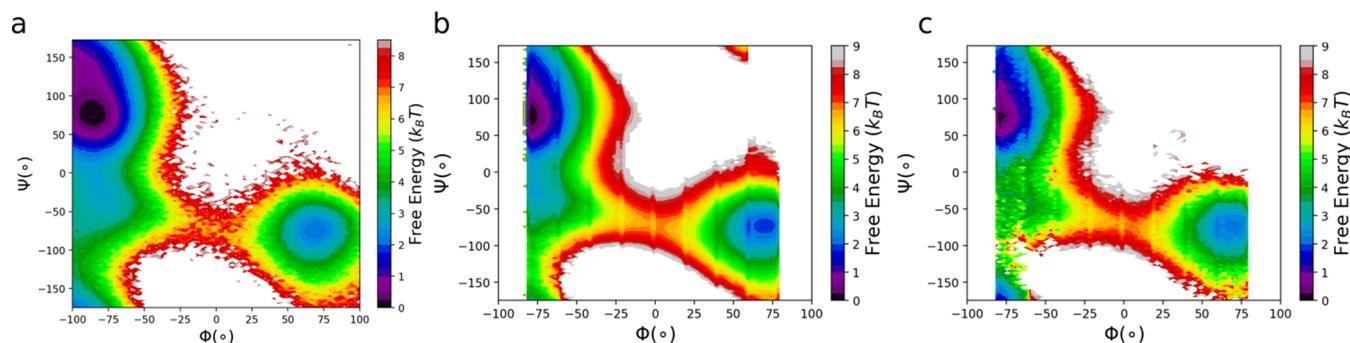


Figure 6. (a) Free-energy landscape of gas-phase alanine dipeptide obtained from equilibrium MD simulation. (b) Reconstructed free-energy landscape from MMVT simulation. (c) Reconstructed free-energy landscape from M-WEM simulation (trial 1). (For a better comparison, the free-energy landscape is constructed from M-WEM iteration 40, with approximately equal amount of the total computational cost in comparison to the MMVT calculation.).

Table 3. Comparison of the Results of the Different Milestoning-Based Methods for the Trypsin–Benzamidine Complex^a

method	residence time (ms)	k_{off} (s^{-1})	$k_{\text{on}} (\times 10^7 \text{ M}^{-1} \text{ s}^{-1})$	ΔG_b (kcal mol $^{-1}$)	simulation time (μs)
experiment ⁸¹	1.7	600 ± 300	2.9	-6.7 ± 0.05	
SEEKR ⁵⁰	12	83 ± 14	2.1 ± 0.3	-7.4 ± 0.10	~ 19
MMVT SEEKR ⁶⁴	5.6	174 ± 9	12 ± 0.5	-7.9 ± 0.04	~ 4.4
MMVT SEEKR ⁶⁴	16	62 ± 6	17 ± 1.0	-8.8 ± 0.07	~ 2.9
M-WEM (200) ^b	1.26 ± 0.32	791 ± 197	0.53 ± 0.08	-5.2 ± 0.16	~ 0.48
M-WEM (300) ^b	1.30 ± 0.44	769 ± 261	0.76 ± 0.38	-5.5 ± 0.37	~ 0.73

^aNumber of iterations of the M-WEM simulation are shown in parentheses. ^bError bars are computed from the last five iterations sampled.

different from the experimental value by a factor of $\sim 4\text{--}5$, while the results of MMVT SEEKR are approximately 1 order of magnitude higher. The ΔG_b value computed from M-WEM, as $k_B T \ln(k_{\text{off}}/k_{\text{on}})$, is also in excellent agreement with the experimental number (within 1.5 kcal mol $^{-1}$). The error bars of the M-WEM results and the SEEKR and MMVT SEEKR are not directly comparable because they are computed differently, as described in the **Theory** section.

The convergence patterns of the residence time and k_{off} are depicted in Figure 7. Both these values converged after about 150 iterations (~ 360 ns of total simulation time) except for small fluctuations. The k_{off} is computed indirectly as the inverse of residence time, which is directly obtained from M-WEM. So, small fluctuations in the residence time get amplified in the k_{off} results in Figure 7a. The computational cost of the M-WEM simulation is ~ 1 order of magnitude less than the other milestoning-based approaches,^{50,64} and the results are in better

agreement with the experiment. Our k_{off} results are also closer to the experimental numbers in comparison to other methods used by Buch et al. ($(9.5 \pm 3.3) \times 10^4 \text{ s}^{-1}$),⁷⁷ Plattner et al. ($(1.31 \pm 1.09) \times 10^4 \text{ s}^{-1}$),⁷⁸ Tiwary et al. ($9.1 \pm 2.5 \text{ s}^{-1}$),⁷⁹ Brotzakis et al. ($4176 \pm 324 \text{ s}^{-1}$),¹⁰¹ and Teo et al. ($260 \pm 240 \text{ s}^{-1}$);⁸⁰ all of these studies required multiple microseconds of simulation, with some in the range of 50–100 μs .^{77,78} A weighted ensemble-based approach has also been used to calculate the kinetics of this system by Dickson and Lotz ($k_{\text{off}} = 5555 \text{ s}^{-1}$)³⁸ and Donyapour et al. ($k_{\text{off}} = 266$ and 840 s^{-1}).³² However, unlike M-WEM, that method could only calculate the unbinding rate constant and dissociation pathways due to the use of a nonequilibrium steady state. To their credit, the authors could distinguish multiple ligand release pathways,³⁸ which is difficult to achieve using milestoning-based simulations with discontinuous trajectories. Nevertheless, we tried to identify some key

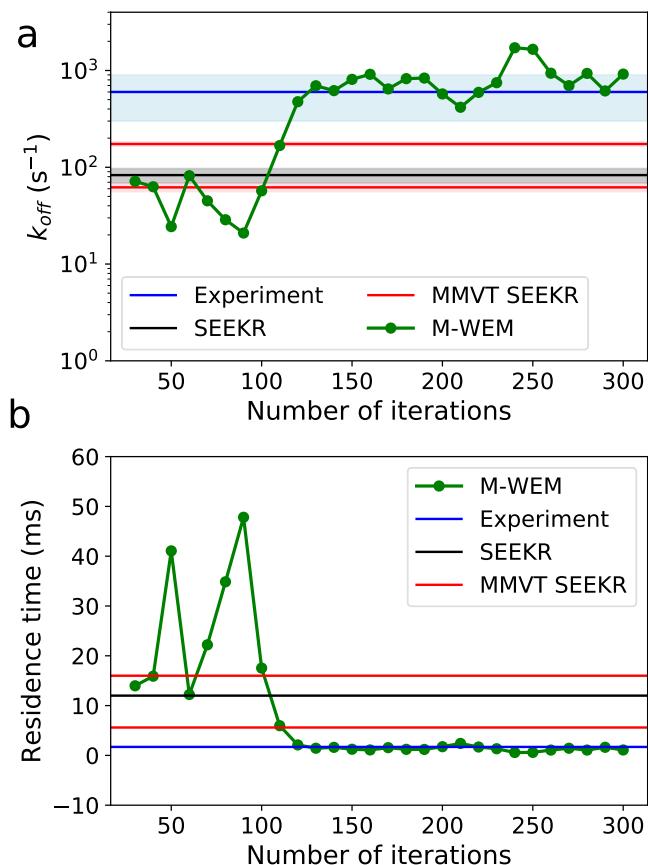


Figure 7. Convergence of (a) k_{off} and (b) ligand residence time for the trypsin–benzamidine complex, as a function of M-WEM iterations. In panel (b), a linear scale is used for a better idea of the quality of the convergence.

intermediates in the unbinding mechanism; we discuss them later in this paper.

A one-dimensional free-energy profile as a function of the milestone reaction coordinate is constructed from the equilibrium probabilities (π_α) obtained from the M-WEM simulation using eq 4. Alongside, a one-dimensional free-energy profile is reconstructed from the M-WEM trajectories following, as described in the Theory section. Error bars in the free-energy landscape are computed as the 95% confidence interval of the free-energy profiles obtained between iterations 160 and 200 with an interval of 10 iterations. The two free-energy profiles obtained from M-WEM using the two different techniques agree with each other, and both are in reasonable agreement with the free-energy surface obtained using well-tempered meta-eABF (WTM-eABF) simulation^{102,103} (see the Supporting Information for details) (Figure 8).

The committor values as a function of the milestone reaction coordinate were computed and are indicated in Figure 9. The results do not show much variation between 200 iterations and 300 iterations, both of which indicate that the transition state (committor = 0.5) is located between the milestones at 6 and 8 Å.

The distribution of the ligand around the protein for three cells (bound state, unbound state and the cell containing the TS) is depicted in Figure 10. It shows the amount of three-dimensional space explored by the ligand during the unbinding process. A two-dimensional projection of the ligand distribution for all cells is shown in Figure 11. The fraction of the spherical

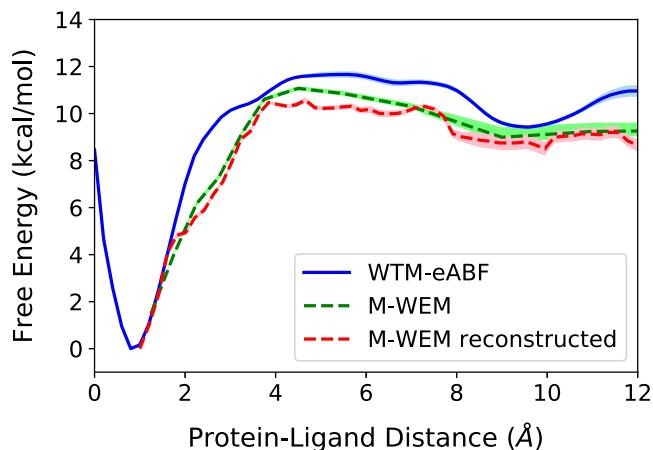


Figure 8. Free-energy profile of the dissociation of the trypsin–benzamidine complex as a function of the milestone reaction coordinate (the center of mass distance between the binding pocket residues and the benzamidine ligand. See the Computational Methods section for details). The results are compared between WTM-eABF simulation and the M-WEM calculation after iteration 200.

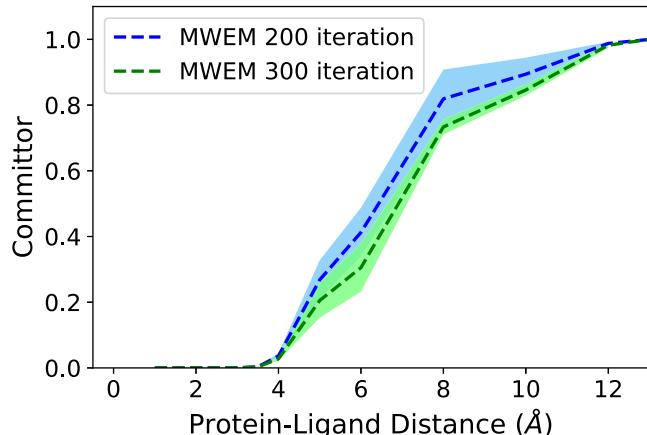


Figure 9. Committor values computed as a function of the milestone reaction coordinate for the M-WEM simulation of the trypsin–benzamidine complex.

surface covered by the ligand in the outermost cell (α) is used for the calculation of k_{on} , as described in the Theory section. The increase in the exploration of the configuration space after 300 iterations in comparison to that after 200 iterations is small.

To get an idea of the intermediate states involved in the protein–ligand interaction, we clustered all of the trajectory frames corresponding to each cell based on heavy-atom RMSD. The number of frames in each cell ranged between ~26 000 and 30 000, with one frame every 2 ps (the length of each WE segment). The clustering of the structures was performed using the GROMOS clustering algorithm¹⁰⁴ implemented in GROMACS v2018.1¹⁰⁵ with an RMSD cutoff of 0.9 Å. The cutoff was chosen such that the total number of clusters is between 10 and 40. All of the cluster centers obtained from different cells were combined together, and a second round of clustering is performed with an RMSD cutoff of 1.1 Å. This resulted in 14 clusters, some of which are depicted in Figure 12. The structures are in qualitative agreement with the metastable states observed by Tiwary et al.⁷⁹ and Brotzakis et al.,¹⁰¹ despite their use of a different enhanced sampling method and of a different version of the AMBER force field in the former study.

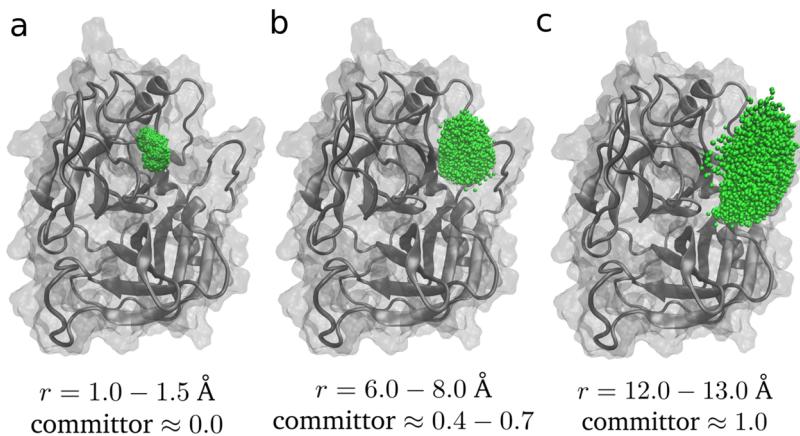


Figure 10. Distribution of the benzamidine ligand around the trypsin protein for three cells of the M-WEM simulations, which, respectively, include (a) the bound state, (b) an apparent transition state with a committor value of ~ 0.5 , and (c) the unbound state.

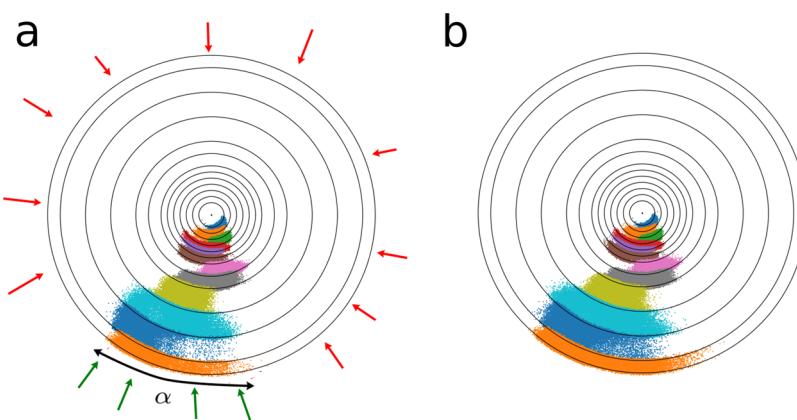


Figure 11. Two-dimensional projection of the distribution of the ligands around the trypsin protein (a) after M-WEM iteration 200 and (b) after iteration 300. The different colors represent structures from M-WEM simulations confined in different cells. The surface coverage, α , used in the k_{on} calculation is also depicted in panel (a). For k_{on} calculation, we assumed that the green trajectories can lead to binding events but the red trajectories cannot.

Particularly, both our study and the work of Tiwary et al. show the presence of a metastable state in which the benzamidine is aligned in a reverse direction (the charged groups facing the aqueous environment and the hydrophobic ring facing the protein). A PDB file with all of the clusters is provided in the Supporting Information.

■ DISCUSSION AND CONCLUSIONS

We developed a new path sampling approach, which combines Markovian milestoning with a weighted ensemble scheme to efficiently calculate the kinetics and free energy of rare events using atomistic MD simulations. This method, which we call Markovian Weighted Ensemble Milestoning (M-WEM), has been applied to study the barrier crossing in a 2D toy system using the Müller–Brown potential, a conformational transition in alanine dipeptide, and, most importantly, to the dissociation and association of the trypsin–benzamidine complex, which has a millisecond scale residence time. For the Müller–Brown potential and the alanine dipeptide systems, the mean first passage time (MFPT) of conformational transition obtained from long equilibrium simulation was quantitatively reproduced by the M-WEM method at a significantly lower computational cost. In the case of alanine dipeptide, we showed how one can also reproduce the two-dimensional free-energy landscape as a function of two backbone torsion angles from one-dimensional

M-WEM and Markovian milestoning simulation, using a free-energy rescaling strategy based on the equilibrium probabilities of each milestone. This approach can be generalized to any other collective variables other than the milestoning coordinate and can potentially elucidate the role of coupled orthogonal degrees of freedom in complex biophysical systems.

For the trypsin–benzamidine complex, the ligand residence time, k_{off} , k_{on} , and the binding free energy could be computed using the M-WEM method in about 1 order of magnitude less computational cost than the Markovian Milestoning-based MMVT simulation and 1–3 orders of magnitude less computational effort compared to other approaches previously used to study this system such as Markov state modeling, metadynamics, adaptive multilevel splitting, weighted ensemble, and traditional milestoning. Our results are in good agreement with the experimental data available for this system.

A key advantage of the M-WEM method is its simple workflow, which essentially requires the user to perform weighted ensemble simulation under flat bottom restraints. This is easy to implement in any simulation engine using an open-access weighted ensemble code such as WESTPA. We implemented M-WEM using the NAMD simulation engine and the WESTPA toolkit. Our implementation uses a minimal adaptive binning (MAB) scheme, which allows for the adaptation of the WE bins throughout the simulation to

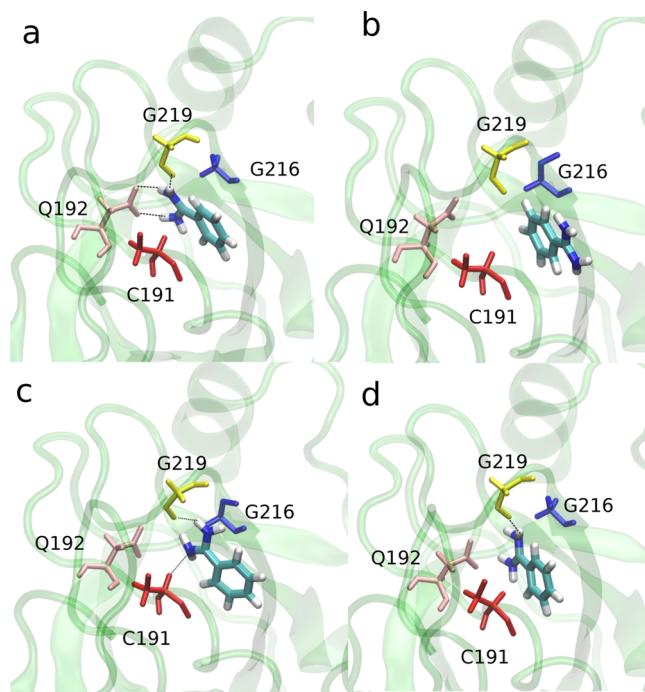


Figure 12. Representative structures sampled from clustering of the M-WEM trajectories of the binding/unbinding of the trypsin–benzamidine complex. The ligand and the residues interacting with the ligand are shown in licorice. Hydrogen bonds between protein and ligand are shown in dashed line.

increase sampling in high-energy regions. Consequently, it does not require preexisting knowledge of the energy landscape and can efficiently sample all possible transitions between milestone interfaces. Moreover, in contrast to traditional milestones approaches, M-WEM (or Markovian milestones in general), it does not require additional simulation (e.g., umbrella sampling) along the milestone interface to sample starting structures, a process that accounts for the majority of the total computational effort. In our previous work, we attempted to replace this expensive additional step using a weighted ensemble restraint-and-release scheme.⁷⁶ The Markovian milestones technique completely removes this step as the trajectory, confined between two milestones, explores by itself the configurational space orthogonal to the milestones coordinate. In the M-WEM approach, we accelerated this “orthogonal sampling” using 2D WE bins along two progress coordinates: the milestones reaction coordinate (to accelerate milestone-to-milestone transitions) and also in another coordinate along the milestone interface. Due to dimensionality scaling, the advantage of the M-WEM over traditional Markovian milestones is more pronounced in the case of trypsin–benzamidine complex, where results, in better agreement with the experiment, could be obtained using M-WEM simulation within a fraction of the computational cost of MMVT SEEKR calculations on the same system. Also, the M-WEM protocol does not need to stop the trajectory at milestone interfaces, avoiding frequent intervention to the dynamics engine and therefore making it more efficient to implement in GPU-based hardware.

Apart from these unique achievements, M-WEM also shares some common advantages with our previously developed WEM methodology. They include the possibility of massively parallelizing the simulations over each milestone, which will be even more pronounced in the current implementation, as

MAB binning has been shown to utilize GPU-based hardware more efficiently than the traditional fixed-binning scheme we used in our earlier work. The convergence of the transition statistics in between milestones is also quicker in M-WEM in comparison to that in MMVT, as evident from the results for the trypsin–benzamidine complex. We also show that a relatively crude reaction coordinate is capable of producing accurate kinetics, particularly in the cases of the Müller–Brown potential and the alanine dipeptide model.

The M-WEM approach, being a combination of two fairly complex path sampling methods, inherits all of the assumptions from each of the individual techniques. Similar to MMVT, M-WEM also requires the milestones to be sufficiently far apart so that the transitions between them are independent of the transitions from other milestones.^{42,83} It also assumes a complete exploration of the configurational space in the milestone hypersurface. Although the weighted ensemble approach is invoked to satisfy these assumptions at a lower computational cost, it comes with additional assumptions inherited from the WE scheme. For example, each individual trajectory trace contributes independently toward the transition statistics despite sharing a significant portion of its propagation history.²⁰ Additionally, due to the use of a stochastic integrator, the time scales obtained from the M-WEM simulations can be dependent on the damping coefficient (γ) of the Langevin dynamics.^{24,106} However, Hall et al. have shown that γ has little effect on ligand residence times obtained from explicit solvent simulation as long as $\gamma > 0.1 \text{ ps}^{-1}$.¹⁰⁶ Our choice of $\gamma = 5 \text{ ps}^{-1}$ was motivated by previous milestones and MMVT work on trypsin–benzamidine^{50,64} and also by the fact that a γ of 5 ps^{-1} reproduces the experimental diffusion constant of water most accurately.¹⁰⁷

A recent study using the MMVT SEEKR approach has predicted the k_{off} of the trypsin–benzamidine complex to be $990 \pm 70 \text{ s}^{-1}$,¹⁰⁸ a value that is relatively closer to the experimental number compared to the previous result of $62 \pm 6 \text{ s}^{-1}$ (Table 3). However, the fact that a virtually identical simulation scheme can produce results that are more than 1 order of magnitude apart clearly shows the high level of uncertainty involved in the milestones approach. So, despite having a better agreement with the experimental value compared to some previous studies, the order-of-magnitude agreement of k_{off} remains the key achievement of the M-WEM approach.

One of the limitations of the current implementation of M-WEM is the use of an analytical approach to compute the binding rate constant. The alternative is to use a multiscale Brownian dynamics (BD) approach,^{50,56,64} which is more rigorous but more computationally expensive. However, BD methods allow us to include the effect of position-dependent variation of the diffusion constant, as well as of the ionic strength of the solution, both of which are absent in our current implementation.

Our M-WEM method can find applications in studying the kinetics and free energy of biomolecular rare events not only for the purpose of fundamental understanding of biological processes but also for kinetics-driven computer-aided drug design. Evidence has emerged over the past decade showing that the efficacy of a small molecule therapeutic drug is more correlated with the residence time than with the binding affinity.^{8,9} Yet, the majority of the drug design effort in the pharmaceutical industry is based on binding free energy; among other things, this is because it is easier to compute than kinetics. The M-WEM approach is a cheap alternative to computationally

expensive traditional enhanced sampling and to path sampling methods and can be included in a computational drug design pipeline using both binding free energy and kinetics. In the future, we plan to test this method on protein–ligand systems with longer residence time, e.g., in the range of minutes to hours, a time frame more characteristic of the drug molecules used in practical applications. The increased sampling of orthogonal coordinates in M-WEM can also facilitate the study of systems where a protein conformational change is coupled to a ligand binding coordinate. Overall, our novel Markovian Weighted Ensemble Milestoning approach is expected to be successful in predicting the free energy and kinetics of biophysical rare events with quantitative accuracy, and it holds the potential of becoming a useful tool in the large-scale computational screening of therapeutic drugs.

■ ASSOCIATED CONTENT

§ Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/acs.jctc.1c00803>.

Discussion on the convergence of M-WEM simulations for alanine dipeptide system, the computational details for the meta-eABF simulation of the trypsin–benzamidine complex, details of the calculation of diffusion coefficient of benzamidine in water, and discussion on the appropriate placement of milestones ([PDF](#))

PDB file containing 14 clusters for the trypsin–benzamidine complex ([PDB](#))

Compressed folder containing the codes for M-WEM implementation ([ZIP](#))

■ AUTHOR INFORMATION

Corresponding Authors

Ioan Andricioaei – Department of Chemistry and Department of Physics and Astronomy, University of California Irvine, Irvine, California 92697, United States; Email: andricio@uci.edu

Dhiman Ray – Department of Chemistry, University of California Irvine, Irvine, California 92697, United States; [ORCID iD: 0000-0002-7103-0886](https://orcid.org/0000-0002-7103-0886); Email: dray1@uci.edu

Author

Sharon Emily Stone – Department of Chemistry, University of California Irvine, Irvine, California 92697, United States

Complete contact information is available at:

<https://pubs.acs.org/10.1021/acs.jctc.1c00803>

Notes

The authors declare no competing financial interest.

The implementation of the M-WEM method using WESTPA software and NAMD simulation package is available from the github repository: <https://github.com/dhimanray/MWEM>. Additional scripts for input file preparation, data analysis, and visualization are provided in the same repository. A compressed zip file of the codes is provided in the [Supporting Information](#). All of the rest of the data are provided in the manuscript and in the [Supporting Information](#). Additional data are available from the corresponding author upon request.

■ ACKNOWLEDGMENTS

The authors thank Rommie Amaro and Lane Votapka for kindly sharing the structure and topology files of the trypsin–

benzamidine complex and Trevor Gokey for providing scripts for protein structure visualization and analysis. This work was supported partially by the National Science Foundation (NSF) via grant MCB 2028443. S.E.S. acknowledges support through Pfizer-La Jolla Academic Industrial Relations (AIR) Diversity Research Fellowship in Chemistry. The authors thank the University of California Irvine High Performance Computing (HPC) facility for providing the computational resources. The authors declare no competing financial interest.

■ REFERENCES

- (1) Wang, L.; et al. Accurate and reliable prediction of relative ligand binding potency in prospective drug discovery by way of a modern free-energy calculation protocol and force field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
- (2) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annu. Rev. Biophys.* **2017**, *46*, 531–558.
- (3) Wereszczynski, J.; McCammon, J. A. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. *Q. Rev. Biophys.* **2012**, *45*, 1–25.
- (4) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *J. Comput. Phys.* **1977**, *23*, 187–199.
- (5) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562–12566.
- (6) Darve, E.; Pohorille, A. Calculating free energies using average force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
- (7) Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *J. Phys. Chem. Lett.* **2018**, *9*, 4738–4745.
- (8) Guo, D.; Mulder-Krieger, T.; IJzerman, A. P.; Heitman, L. H. Functional efficacy of adenosine A 2A receptor agonists is positively correlated to their receptor residence time. *Br. J. Pharmacol.* **2012**, *166*, 1846–1859.
- (9) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730–739.
- (10) Xing, C.; Andricioaei, I. On the calculation of time correlation functions by potential scaling. *J. Chem. Phys.* **2006**, *124*, No. 034110.
- (11) Nummela, J.; Andricioaei, I. Exact Low-Force Kinetics from High-Force Single-Molecule Unfolding Events. *Biophys. J.* **2007**, *93*, 3373–3381.
- (12) Wolf, S.; Stock, G. Targeted Molecular Dynamics Calculations of Free Energy Profiles Using a Nonequilibrium Friction Correction. *J. Chem. Theory Comput.* **2018**, *14*, 6175–6182.
- (13) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nat. Commun.* **2020**, *11*, No. 2918.
- (14) Pratt, L. R. A statistical method for identifying transition states in high dimensional problems. *J. Chem. Phys.* **1986**, *85*, 5045.
- (15) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition path sampling and the calculation of rate constants. *J. Chem. Phys.* **1998**, *108*, 1964–1977.
- (16) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling: Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annu. Rev. Phys. Chem.* **2002**, *53*, 291–318.
- (17) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-sampling strategies for simulating rare events in biomolecular systems. *Curr. Opin. Struct. Biol.* **2017**, *43*, 88–94.
- (18) Huber, G.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97–110.
- (19) Kahn, H. *Use of Different Monte Carlo Sampling Techniques*; RAND Corporation: Santa Monica, CA, 1955.
- (20) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.

- (21) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *J. Chem. Phys.* **2010**, *132*, No. 054107.
- (22) Zwier, M. C.; Adelman, J. L.; Kaus, J. W.; Pratt, A. J.; Wong, K. F.; Rego, N. B.; Suárez, E.; Lettieri, S.; Wang, D. W.; Grabe, M.; Zuckerman, D. M.; Chong, L. T. WESTPA: An Interoperable, Highly Scalable Software Package for Weighted Ensemble Simulation and Analysis. *J. Chem. Theory Comput.* **2015**, *11*, 800–809.
- (23) Abdul-Wahid, B.; Feng, H.; Rajan, D.; Costauoc, R.; Darve, E.; Thain, D.; Izaguirre, J. A. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. *J. Chem. Inf. Model.* **2014**, *54*, 3033–3043.
- (24) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *J. Am. Chem. Soc.* **2019**, *141*, 6519–6526.
- (25) Suárez, E.; Lettieri, S.; Zwier, M. C.; Stringer, C. A.; Subramanian, S. R.; Chong, L. T.; Zuckerman, D. M. Simultaneous computation of dynamical and equilibrium information using a weighted ensemble of trajectories. *J. Chem. Theory Comput.* **2014**, *10*, 2658–2667.
- (26) Zwier, M. C.; Kaus, J. W.; Chong, L. T. Efficient Explicit-Solvent Molecular Dynamics Simulations of Molecular Association Kinetics: Methane/Methane, Na⁺/Cl⁻, Methane/Benzene, and K⁺/18-Crown-6 Ether. *J. Chem. Theory Comput.* **2011**, *7*, 1189–1197.
- (27) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein–Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *J. Phys. Chem. Lett.* **2016**, *7*, 3440–3445.
- (28) Saglam, A. S.; Chong, L. T. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci.* **2019**, *10*, 2360–2372.
- (29) Adelman, J. L.; Grabe, M. Simulating Current–Voltage Relationships for a Narrow Ion Channel Using the Weighted Ensemble Method. *J. Chem. Theory Comput.* **2015**, *11*, 1907–1918.
- (30) Spiriti, J.; Zuckerman, D. M. Tabulation as a high-resolution alternative to coarse-graining protein interactions: Initial application to virus capsid subunits. *J. Chem. Phys.* **2015**, *143*, No. 243159.
- (31) Dickson, A.; Brooks, C. L. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *J. Phys. Chem. B* **2014**, *118*, 3532–3542.
- (32) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. *J. Chem. Phys.* **2019**, *150*, No. 244112.
- (33) Suárez, E.; Adelman, J. L.; Zuckerman, D. M. Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. *J. Chem. Theory Comput.* **2016**, *12*, 3473–3481.
- (34) Degraeve, A. J.; Bogetti, A. T.; Chong, L. T. The RED scheme: Rate-constant estimation from pre-steady state weighted ensemble simulations. *J. Chem. Phys.* **2021**, *154*, No. 114111.
- (35) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *J. Phys. Chem. A* **2021**, *125*, 1642.
- (36) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian “microbin” Analysis. *J. Chem. Theory Comput.* **2020**, *16*, 6763–6775.
- (37) Dickson, A.; Lotz, S. D. Ligand Release Pathways Obtained with WExplore: Residence Times and Mechanisms. *J. Phys. Chem. B* **2016**, *120*, 5377–5385.
- (38) Dickson, A.; Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophys. J.* **2017**, *112*, 620–629.
- (39) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-Mediated Ligand Unbinding of the PK-11195 Ligand from TSPO. *Biophys. J.* **2021**, *120*, 158–167.
- (40) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *J. Am. Chem. Soc.* **2018**, *140*, 618–628.
- (41) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestone. *J. Chem. Phys.* **2004**, *120*, 10880–10889.
- (42) West, A. M. A.; Elber, R.; Shalloway, D. Extending molecular dynamics time scales with milestone: Example of complex kinetics in a solvated peptide. *J. Chem. Phys.* **2007**, *126*, No. 145104.
- (43) Bello-Rivas, J. M.; Elber, R. Exact milestone. *J. Chem. Phys.* **2015**, *142*, No. 094102.
- (44) van Erp, T. S.; Moroni, D.; Bolhuis, P. G. A novel path sampling method for the calculation of rate constants. *J. Chem. Phys.* **2003**, *118*, 7762.
- (45) Dinner, A. R.; Mattingly, J. C.; Tempkin, J. O. B.; Koten, B. V.; Weare, J. Trajectory Stratification of Stochastic Dynamics. *SIAM Rev.* **2018**, *60*, 909–938.
- (46) Dinner, A. R.; Thiede, E. H.; Koten, B. V.; Weare, J. Stratification as a General Variance Reduction Method for Markov Chain Monte Carlo. *SIAM/ASA J. Uncertainty Quantif.* **2020**, *8*, 1139–1188.
- (47) Guttenberg, N.; Dinner, A. R.; Weare, J. Steered transition path sampling. *J. Chem. Phys.* **2012**, *136*, No. 234103.
- (48) Elber, R.; Bello-Rivas, J.; Ma, P.; Cardenas, A.; Fathizadeh, A. Calculating Iso-Committor Surfaces as Optimal Reaction Coordinates with Milestone. *Entropy* **2017**, *19*, No. 219.
- (49) Wei, W.; Elber, R. ScMile: A Script to Investigate Kinetics with Short Time Molecular Dynamics Trajectories and the Milestone Theory. *J. Chem. Theory Comput.* **2020**, *16*, 860–874.
- (50) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin–Benzamidine Binding. *J. Phys. Chem. B* **2017**, *121*, 3597–3606.
- (51) Elber, R. A Milestoning Study of the Kinetics of an Allosteric Transition: Atomically Detailed Simulations of Deoxy Scapharca Hemoglobin. *Biophys. J.* **2007**, *92*, L85–L87.
- (52) Votapka, L. W.; Lee, C. T.; Amaro, R. E. Two Relations to Estimate Membrane Permeability Using Milestoning. *J. Phys. Chem. B* **2016**, *120*, 8606–8616.
- (53) Fathizadeh, A.; Elber, R. Ion Permeation through a Phospholipid Membrane: Transition State, Path Splitting, and Calculation of Permeability. *J. Chem. Theory Comput.* **2019**, *15*, 720–730.
- (54) Cardenas, A. E.; Elber, R. Markovian and Non-Markovian Modeling of Membrane Dynamics with Milestoning. *J. Phys. Chem. B* **2016**, *120*, 8208–8216.
- (55) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *J. Chem. Phys.* **2009**, *130*, No. 194101.
- (56) Votapka, L. W.; Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLoS Comput. Biol.* **2015**, *11*, No. e1004381.
- (57) Tang, Z.; Chen, S. H.; Chang, C. E. A. Transient States and Barriers from Molecular Simulations and the Milestoning Theory: Kinetics in Ligand-Protein Recognition and Compound Design. *J. Chem. Theory Comput.* **2020**, *16*, 1882–1895.
- (58) Narayan, B.; Buchete, N.-V.; Elber, R. Computer Simulations of the Dissociation Mechanism of Gleevec from Abl Kinase with Milestoning. *J. Phys. Chem. B* **2021**, *125*, 5706–5715.
- (59) Jagger, B. R.; Lee, C. T.; Amaro, R. E. Quantitative Ranking of Ligand Binding Kinetics with a Multiscale Milestoning Simulation Approach. *J. Phys. Chem. Lett.* **2018**, *9*, 4941–4948.
- (60) Ma, P.; Cardenas, A. E.; Chaudhari, M. I.; Elber, R.; Rempe, S. B. The Impact of Protonation on Early Translocation of Anthrax Lethal Factor: Kinetics from Molecular Dynamics Simulations and Milestoning Theory. *J. Am. Chem. Soc.* **2017**, *139*, 14837–14840.
- (61) Ma, P.; Cardenas, A. E.; Chaudhari, M. I.; Elber, R.; Rempe, S. B. Probing Translocation in Mutants of the Anthrax Channel: Atomically Detailed Simulations with Milestoning. *J. Phys. Chem. B* **2018**, *122*, 10296–10305.
- (62) Wang, H.; Elber, R. Catalytic Magnesium as a Door Stop for DNA Sliding. *J. Phys. Chem. B* **2021**, *125*, 3494–3500.

- (63) Narayan, B.; Fathizadeh, A.; Templeton, C.; He, P.; Arasteh, S.; Elber, R.; Buchete, N. V.; Levy, R. M. The transition between active and inactive conformations of Abl kinase studied by rock climbing and Milestoning. *Biochim. Biophys. Acta, Gen. Subj.* **2020**, *1864*, No. 129508.
- (64) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *J. Chem. Theory Comput.* **2020**, *16*, 5348–5357.
- (65) Maragliano, L.; Vanden-Eijnden, E.; Roux, B. Free energy and kinetics of conformational transitions from voronoi tessellated milestoning with restraining potentials. *J. Chem. Theory Comput.* **2009**, *5*, 2589–2594.
- (66) Alberini, G.; Benfenati, F.; Maragliano, L. Molecular Dynamics Simulations of Ion Selectivity in a Claudin-15 Paracellular Channel. *J. Phys. Chem. B* **2018**, *122*, 10783–10792.
- (67) Cottone, G.; Chiodo, L.; Maragliano, L. Thermodynamics and kinetics of ion permeation in wild-type and mutated open active conformation of the human $\alpha 7$ nicotinic receptor. *J. Chem. Inf. Model.* **2020**, *60*, 5045–5056.
- (68) Alberini, G.; Benfenati, F.; Maragliano, L. Structural Mechanism of ω -Currents in a Mutated Kv7.2 Voltage Sensor Domain from Molecular Dynamics Simulations. *J. Chem. Inf. Model.* **2021**, *61*, 1354–1367.
- (69) Jiang, W.; Lin, Y.-C.; Botello-Smith, W.; Contreras, J. E.; Harris, A. L.; Maragliano, L.; Luo, Y. L. Free energy and kinetics of cAMP permeation through connexin26 via applied voltage and milestoning. *Biophys. J.* **2021**, *120*, 2969–2983.
- (70) Ma, W.; Schulten, K. Mechanism of substrate translocation by a ring-shaped ATPase motor at millisecond resolution. *J. Am. Chem. Soc.* **2015**, *137*, 3031–3040.
- (71) Yu, T. Q.; Lapelosa, M.; Vanden-Eijnden, E.; Abrams, C. F. Full kinetics of CO entry, internal diffusion, and exit in myoglobin from transition-path theory simulations. *J. Am. Chem. Soc.* **2015**, *137*, 3041–3050.
- (72) Grazioli, G.; Andricioaei, I. Advances in milestoning. I. Enhanced sampling via wind-assisted reweighted milestoning (WARM). *J. Chem. Phys.* **2018**, *149*, No. 084103.
- (73) Wang, H.; Elber, R. Milestoning with wind: Exploring the impact of a biasing potential in exact calculation of kinetics. *J. Chem. Phys.* **2020**, *152*, No. 224105.
- (74) Ray, D.; Andricioaei, I. Weighted ensemble milestoning (WEM): A combined approach for rare event simulations. *J. Chem. Phys.* **2020**, *152*, No. 234114.
- (75) Pan, A. C.; Xu, H.; Palpant, T.; Shaw, D. E. Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2017**, *13*, 3372–3377.
- (76) Ray, D.; Gokey, T.; Mobley, D. L.; Andricioaei, I. Kinetics and free energy of ligand dissociation using weighted ensemble milestoning. *J. Chem. Phys.* **2020**, *153*, No. 154117.
- (77) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184–10189.
- (78) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, No. 7653.
- (79) Tiwary, P.; Limongelli, V.; Salvaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, E386–E391.
- (80) Teo, I.; Mayne, C. G.; Schulten, K.; Lelièvre, T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *J. Chem. Theory Comput.* **2016**, *12*, 2983–2989.
- (81) Guillain, F.; Thusius, D. The use of Proflavin as an Indicator in Temperature-Jump Studies of the Binding of a Competitive Inhibitor to Trypsin. *J. Am. Chem. Soc.* **1970**, *92*, 5534–5536.
- (82) Májek, P.; Elber, R. Milestoning without a Reaction Coordinate. *J. Chem. Theory Comput.* **2010**, *6*, 1805–1817.
- (83) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *J. Chem. Phys.* **2008**, *129*, No. 174102.
- (84) Noé, F. Probability distributions of molecular observables computed from Markov models. *J. Chem. Phys.* **2008**, *128*, No. 244103.
- (85) Ray, D.; Andricioaei, I. Free Energy Landscape and Conformational Kinetics of Hoogsteen Base Pairing in DNA vs. RNA. *Biophys. J.* **2020**, *119*, 1568–1579.
- (86) Flyvbjerg, H.; Petersen, H. G. Error estimates on averages of correlated data. *J. Chem. Phys.* **1989**, *91*, 461.
- (87) Bussi, G.; Tribello, G. A. Analyzing and Biasing Simulations with PLUMED. *Biomolecular Simulations: Methods in Molecular Biology*; Humana Press, 2019; Vol. 2022, pp 529–578.
- (88) McCammon, J. A.; Northrup, S. H.; Allison, S. A. Diffusional dynamics of ligand-receptor association. *J. Phys. Chem. A.* **1986**, *90*, 3901–3905.
- (89) Berezhkovskii, A. M.; Szabo, A. Committors, first-passage times, fluxes, Markov states, milestones, and all that. *J. Chem. Phys.* **2019**, *150*, No. 054106.
- (90) Ma, N.; Van Der Vaart, A. Free Energy Coupling between DNA Bending and Base Flipping. *J. Chem. Inf. Model.* **2017**, *57*, 2020–2026.
- (91) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theor. Chim. Acta* **1979**, *53*, 75–93.
- (92) Huang, J.; Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135–2145.
- (93) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (94) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kalé, L.; Schulten, K. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (95) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Mol. Phys.* **2013**, *111*, 3345–3362.
- (96) Marquart, M.; Walter, J.; Deisenhofer, J.; Bode, W.; Huber, R. The geometry of the reactive site and of the peptide groups in trypsin, trypsinogen and its complexes with inhibitors. *Acta Crystallogr., Sect. D: Struct. Biol.* **1983**, *39*, 480–490.
- (97) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.
- (98) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
- (99) Horn, H. W.; Swope, W. C.; Pitera, J. W.; Madura, J. D.; Dick, T. J.; Hura, G. L.; Head-Gordon, T. Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. *J. Chem. Phys.* **2004**, *120*, 9665–9678.
- (100) Shalloway, D. Macrostates of classical stochastic systems. *J. Chem. Phys.* **1996**, *105*, 9986.
- (101) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the Calculation of Protein–Ligand Binding Free Energy and Residence Times Using Dynamically Optimized Collective Variables. *J. Chem. Theory Comput.* **2019**, *15*, 743–750.
- (102) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **2019**, *52*, 3254–3264.
- (103) Fu, H.; Chen, H.; Wang, X.; Chai, H.; Shao, X.; Cai, W.; Chipot, C. Finding an optimal pathway on a multidimensional free-energy landscape. *J. Chem. Inf. Model.* **2020**, *60*, 5366–5374.
- (104) Daura, X.; et al. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236–240.
- (105) Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindah, E. Gromacs: High performance molecular simulations

through multi-level parallelism from laptops to supercomputers.
SoftwareX **2015**, *1–2*, 19–25.

(106) Hall, R.; Dixon, T.; Dickson, A. On Calculating Free Energy Differences Using Ensembles of Transition Paths. *Front. Mol. Biosci.* **2020**, No. 106.

(107) Wu, Y.; Lee, J.; Wang, Y. A Comparative Study of GROMACS and NAMD, 2007, https://courses.physics.illinois.edu/phys466/sp2013/projects/2007/team_h2o/NAMDvsGromacs.pdf.

(108) Votapka, L.; Stokely, A.; Ojha, A.; Amaro, R. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM 7.5 Molecular Dynamics Engine. *ChemRxiv* **2021**, DOI: 10.33774/chemrxiv-2021-pplfs.