bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

Dhiman Ray,[†] Sharon E. Stone,[†] and Ioan Andricioaei[*,‡,‡]

†*Department of Chemistry, University of California Irvine, Irvine CA 92697*

‡*Department of Physics and Astronomy, University of California Irvine, Irvine CA 92697*

E-mail: andricio@uci.edu

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Abstract

We introduce a rare-event sampling scheme, named Markovian Weighted Ensemble Milestoning (M-WEM), which inlays the weighted ensemble framework within a Markovian milestoning theory to efficiently calculate thermodynamic and kinetic properties of long-timescale biomolecular processes from short atomistic molecular dynamics simulations. We then showcase the application of this method to the Müller Brown potential model, the conformational switching in alanine dipeptide, and the millisecond timescale protein-ligand unbinding in the trypsin-benzamidine complex. Not only can we predict the kinetics of these processes with quantitative accuracy, but we also present 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 approach within a few hundreds of nanoseconds simulation time, which is a fraction of the computational cost of other currently available methods, and close to four 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 application 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 macro-molecules. 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 force the simulated system in conformational basins of attraction, impeding it to sample all relevant states in a proper manner, thereby breaking

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ergodicity. Examples of such processes include protein folding, conformational transitions, ligand binding and unbinding etc., which in most cases involve $\sim 10^4 - 10^6$ atoms including the natural solution environment. Despite phenomenal advancements in computing hardware over the past few decades atomistic MD simulations of such large systems still go, typically, up to multiple microseconds only. This is many orders of magnitude smaller than the timescale relevant to biological function, which is often in the range of seconds to hours.

The current study focuses on protein-ligand interactions; adequate sampling of which is pivotal to computer aided drug design. Molecular dynamics (MD) simulations provide the mechanistic insights on such interactions in atomistic detail, and is 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 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 and the residence time.[8,9] It is difficult to compute the kinetic properties from traditional enhanced sampling simulations as the dynamics become non-physical due to the application of artificial bias. 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 timescale of one event which itself is beyond the reach of world's most powerful supercomputers. This results in a dire need to develop theoretical methods and computational algorithms to make quantitative predictions about the kinetics of long timescale processes such as rare events from short timescale trajectories.

In response to this situation, path sampling based methods were introduced to the field, with the pioneering contribution by Chandler and co-workers on transition path sampling (TPS).[10,11] These approaches have become increasingly popular in the last two decades. In stead of applying external bias, path sampling methods utilize the statistical properties of

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 mechanism.[12] One of the most popular path sampling method is the weighted ensemble (WE) approach,[13] pioneered by Huber and Kim in 1996,[14] and, much later, popularized by Zhang et al.[15] who also established that the WE method is statistically exact. In WE approach, the conformational space between the initial and final state 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 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 some of the fast moving but low weight trajectories to reach the final state well before the mean first passage time, facilitating the calculation of converged kinetics, free energy and pathways at a relatively low computational cost. With the implementation in the open source software WESTPA,[16] weighted ensemble method has seen a wide range of applications including protein folding and conformational transitions,[17–19] formation of host-guest complexes,[20] protein-peptide[21] and protein-protein binding,[22] ion permeation through protein channels,[23] viral capsid assembly[24] and many others. Many new variants as well as new analysis schemes for the traditional WE approach has emerged in recent years including WExplore,[25] resampling of ensembles by variation optimization (REVO),[26] history augmented Markov State Modeling (haMSM),[27] the RED scheme,[28] minimal adaptive binning (MAB),[29] and micro-bin analysis.[30] Particularly the WExplore and REVO algorithms have been successfully applied to study the pathways and kinetics of protein-ligand dissociation,[26,31,32] even for systems with residence times as high as seconds to minutes.[33,34]

Another popular approach to study the kinetics of biophysical rare events is milestoning.[35–37] In this method, multiple interfaces are placed along a reaction coordinate 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transition statistics via a mathematical framework,[37] the kinetics and free energy profile is estimated. This method has also been implemented in the software tools miles,[38] ScMile[39] and SEEKR,[40] and has been used to study a variety of complex biological problems including protein allosteric transitions,[41] membrane permeation by small molecules,[42–45] protein small molecule interaction,[40,46–48] simple ligand-receptor binding,[49] peptide transport through protein channels,[50,51] DNA protein interaction,[52] protein conformational dynamics[53] 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.[36] This increases the computational cost significantly leaving aside the fact that almost 80% of the total simulation time is spent on sampling along the milestone interface to generate starting structures in accordance with the equilibrium distribution. A different variant of the milestoning approach has been developed: Markovian Milestoning with Voronoi Tesselation (MMVT),[45,54] which removes the necessity of performing additional sampling along 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,[55–58] substrate translocation through ATPase motor,[59] and the CO entry in myoglobin.[60] Only recently, the Markovian milestoning approach has been tested on ligand-receptor binding for crown ether based host guest complexes and for the dissociation of a benzamidine ligand from the trypsin protein.[61] 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.[61]

In our previous work we attempted improve the milestoning scheme by accelerating transitions between distant milestones via the application of directed wind force.[62] This technique did increase the number of energetically uphill transitions, but the statistical properties of the computed observables were not significantly better.[63] 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

spaced milestones.[64] The WEM method not only produced very accurate prediction of kinetics, free energy and time correlation function for small molecular systems like Alanine dipeptide,[64] we could also reproduce protein-ligand binding and unbinding rate constants and binding affinity, obtained from 30 μs equilibrium simulation,[65] in less than 100 ns of WEM simulation.[66]

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.[66] This can potentially be rectified by using multiple starting states on the milestone interface sampled from long umbrella sampling simulation at the expense of manifold increase in the computational cost similar to traditional milestoning. Second, the choice of 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 WEM technique is the complexity of the simulation protocol which requires propagating many short trajectories and stopping them upon reaching a nearby milestone.[66] 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 inefficient to implement, particularly in Graphical Processing Unit (GPU) based hardware.

We, thereby, present the novel Markovian Weighted Ensemble Milestoning (M-WEM) approach where we combine weighted ensemble with soft-wall[54] based Markovian Milestoning scheme, in attempt to mitigate the deficiencies and improve the performance of the weighted ensemble milestoning technique. We first provide a detailed description of the theory of the Markovian milestoning and M-WEM approach. Then we show the application of this method on the two dimensional Müller Brown potential, the conformational transition of alanine dipeptide, and the dissociation and association of the trypsin-benzamidine com-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

plex, 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 on this system, including Markov State Modeling (MSM),[67,68] Metadynamics,[69] Adaptive Multilevel Splitting (AMS),[70] Milestoning,[40] MMVT,[61] WExplore,[32] and REVO.[26] 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 Guillian and Thusius.[71] We also discuss a new approach to construct multidimensional free energy landscape via post-analysis of MMVT and M-WEM trajectories obtained using 1D reaction coordinate, with a potential application in systems were orthogonal degrees of freedom are strongly coupled with the reaction coordinate.

Theory

Markovian Milestoning

The theoretical details of the Markovian milestoning with Voronoi tesselation (MMVT) approach is described elsewhere.[45,54,61] Here, we provide only a brief description relevant to the current work.

In MMVT, the configurational space or reaction coordinate is discretized into Voronoi cells A flat bottom potential is applied to each cell with half harmonic walls places at each milestone interface preventing the trajectories from escaping out of the Voronoi cells. For 1D reaction coordinate used here the flat bottom potential has the expression:

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

where α is the cell index, x_i^α is the value of the reaction coordinate at milestone i at the boundary of 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 which 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 milestones 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[54,55] which we adopt in the current work. Alternatively a *hard wall* restraint[45,61] 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:

$$k_{\alpha,\beta} = \frac{N_{\alpha,\beta}}{T_\alpha} \tag{2}$$

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

$$\sum_{\beta=1,\beta\neq\alpha}^{\Lambda} \pi_\beta k_{\beta,\alpha} = \sum_{\beta=1,\beta\neq\alpha}^{\Lambda} \pi_\alpha k_{\alpha,\beta}; \qquad \sum_{\alpha=1}^{\Lambda} \pi_\alpha = 1 \tag{3}$$

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

$$G_\alpha = -k_B T \ln(\pi_\alpha) \tag{4}$$

For calculating kinetics, the transition matrix $\mathbf{N} \in \mathbb{R}^{M\times M}$ and the lifetime vector $\mathbf{R} \in \mathbb{R}^M$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

are constructed, whose elements are computed as follows:

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

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:

$$
\begin{aligned}
Q_{ij} &= \begin{cases} \frac{N_{ij}}{R_i} & \text{if} \quad R_i \neq 0; \quad i \neq j \\ 0 & \text{if} \quad R_i = 0; \quad i \neq j \end{cases} \\
Q_{ii} &= -\sum_{i \neq j} Q_{ij}
\end{aligned}
\tag{6}
$$

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

$$
\widehat{\mathbf{Q}} \boldsymbol{\tau}^M = -\mathbf{1} \tag{7}
$$

$\widehat{\mathbf{Q}} \in \mathbb{R}^{M-1 \times M-1}$ is the matrix obtained by deleting the last row and column of $\mathbf{Q}$. $\mathbf{1}$ is an $M-1$ dimensional unit vector, and $\boldsymbol{\tau}^M \in \mathbb{R}^{M-1}$ 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

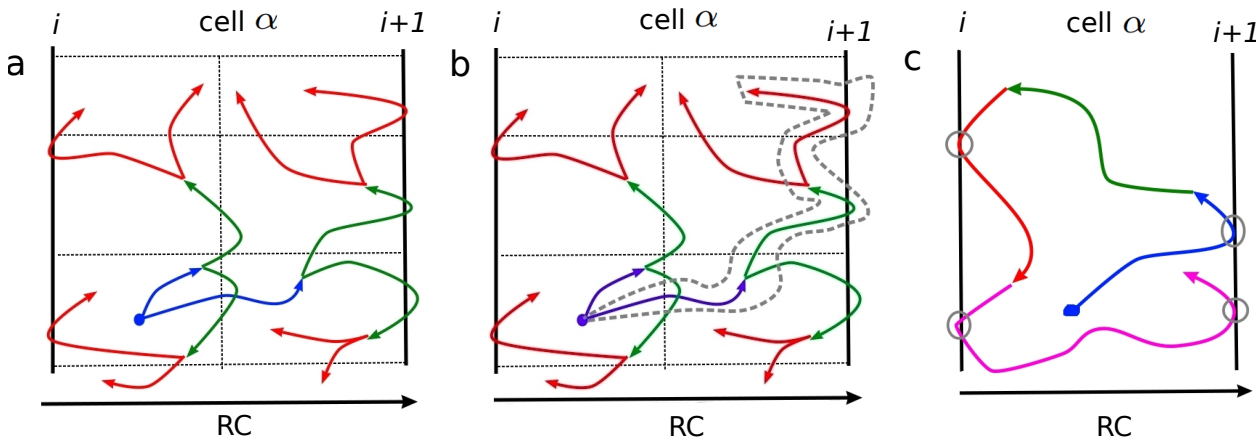

Figure 1: A schematic representation of the M-WEM simulation protocol. The thick lines indicate milestones (labeled as milestone index i and $i+1$). The dotted lines indicate a WE bin boundaries (which are adapted during the simulation but in this figure we show fixed bins for clarity). Trajectories for different WE iteration is shown in different color scheme: 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 are traced back from the last iteration (an example trace is highlighted with gray dashed line). (c) The milestone crossing events (gray circles) are recorded from each trace, and are used in subsequent analysis.

protocol is depicted in Fig. 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.[29] 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 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.[29]

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 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:

$$k_{\alpha,\beta} = \sum_{J=1}^{M_\alpha} w_J \frac{N^J_{\alpha,\beta}}{T^J_\alpha} = \sum_{J=1}^{M_\alpha} w_J k^J_{\alpha,\beta} \tag{8}$$

where w_J is the weight of the Jth trajectory trace. $N^J_{\alpha,\beta}$, T^J_α, and $k^J_{\alpha,\beta}$ have similar definitions as in Eq. 2 except that they are computed just from the Jth trace. The equilibrium probability distribution and the free energy profile are computed from the elements of the flux matrix obtained from Eq. 8 using Eq. 3 and 4, respectively.

For calculating kinetics, the $N_{i,j}$ 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:

$$N_{ij} = T \sum_{\alpha=1}^{\Lambda} \pi_\alpha \sum_{J=1}^{M_\alpha} w_J \frac{N^{\alpha,J}_{ij}}{T^J_\alpha} = T \sum_{\alpha=1}^{\Lambda} \sum_{J=1}^{M_\alpha} w_J \pi_\alpha \frac{N^{\alpha,J}_{ij}}{T^J_\alpha} \tag{9}$$

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

$$\widetilde{N}^{\alpha,J}_{ij} = \pi_\alpha \frac{N^{\alpha,J}_{ij}}{T^J_\alpha}. \tag{10}$$

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

$$\mathbf{N} = T \sum_{\alpha=1}^{\Lambda} \sum_{J=1}^{M_\alpha} w_J \widetilde{\mathbf{N}}^{\alpha,J} \tag{11}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

$$\begin{aligned} \mathbf{R} &= T \sum_{\alpha=1}^{\Lambda} \sum_{J=1}^{M_\alpha} w_J \widetilde{\mathbf{R}}^{\alpha,J} \\ \text{where} \quad \widetilde{R}_i^{\alpha,J} &= \pi_\alpha \frac{R_i^{\alpha,J}}{T_\alpha^J} \end{aligned} \tag{12}$$

and $R_i^{\alpha,J}$ is R_i^α are computed for the Jth 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 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 beta distribution.[50] Similar to our previous work,[64,66] we generated the ensemble of rate matrices by sampling from a Bayesian type conditional probability,[46,72,73] given by

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

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 non-reversible element exchange Monte-Carlo scheme.[74] 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 $\mathbf{Q}'$.

$$\begin{aligned} Q'_{ij} &= Q_{ij} + \Delta \quad (i \neq j) \\ Q'_{ii} &= Q_{ii} - \Delta. \end{aligned} \tag{14}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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:

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

As each step modifies only one element, the sampled matrices are highly correlated.[74] 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 mean and 95% confidence intervals. This protocol is followed throughout the manuscript unless mentioned otherwise.

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

$$k_{\text{off}} = \frac{1}{\tau}, \tag{16}$$

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 milestoning or M-WEM framework described above.

We noted in our previous work that the ligand-binding kinetics is often diffusion dominant. 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 the specific system in hand, should be carefully evaluated in a case-by-case basis.) In our previous work we delineated an analytical method to combine diffusion theory with milestoning transition kernel to compute the binding rate constant k_{on}.[66] We begin with the expression of diffusion-dependent arrival rate of small molecule on the surface of a sphere of radius r:[75]

$$k(r) = 4\pi D r \tag{17}$$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. Second, we consider that a ligand can only reach the bound state if it moves towards 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 (Mth) milestone to the previous milestone. But as the transitional kernel is not directly available in the current Markovian milestoning framework, we compute this inward transition probability from the **N** matrix.

$$K_{M,M-1} = \frac{N_{M,M-1}}{N_{M,M-1} + N_{M,M+1}} \tag{18}$$

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

$$k(r) = 4\pi D r \alpha K_{M,M-1}, \tag{19}$$

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

$$\begin{aligned} k_{\text{on}} &= 4\pi D r \alpha K_{M,M-1} N_{av} \\ &= 7.569 \times 10^8 r \alpha K_{M,M-1} \end{aligned} \tag{20}$$

In the last step we used the fact that the typical value of the diffusion constant of small molecules in water is 1×10^5 cm^2 s^{-1} and that 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, both for Markovian milestoning and M-WEM simulations. The trajectory data is first histogrammed in appropriate collective variables to obtain differ-

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ent independent histograms for each individual cell Λ. A weighted sum of these histograms are then performed to obtain the equilibrium high dimensional probability distribution

$$p(\mathbf{R}) = \sum_{\alpha=1}^{\Lambda} \pi_\alpha p^\alpha(\mathbf{R}), \tag{21}$$

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 case of M-WEM the $p^\alpha(\mathbf{R})$ is obtained from multiple (M_α) trajectories of different weights as:

$$\begin{aligned} p^\alpha(\mathbf{R}) &= \sum_{J=1}^{M_\alpha} w_J p_J^\alpha(\mathbf{R}) \\ p(\mathbf{R}) &= \sum_{\alpha=1}^{\Lambda} \pi_\alpha \sum_{J=1}^{M_\alpha} w_J p_J^\alpha(\mathbf{R}), \end{aligned} \tag{22}$$

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

$$G(\mathbf{R}) = -k_B T \ln(p(\mathbf{R}). \tag{23}$$

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

Committor

The committor 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.[76] Recent work by Elber et al. established that it is possible to calculate the committor at the milestone interfaces.[38] But, to the best of our knowledge, such approach has not been applied to Markovian milestoning techniques so far. We performed the committor calculation in the following way. First, we constructed a transition kernel ($\mathbf{K}$), equivalent to conventional

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

milestoning, from the $\mathbf{N}$ matrix:

$$K_{ij} = \frac{N_{ij}}{\sum_j N_{ij}} \tag{24}$$

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 back to the previous milestones. This is achieved by replacing the last row of $\mathbf{K}$ matrix with zeros except for the element corresponding to the last milestone for which the value is one.[38,77] For a three-milestone model, this can be illustrated as:

$$\mathbf{K} = \begin{bmatrix} 0 & K_{12} & 0 \\ K_{21} & 0 & K_{23} \\ 0 & K_{32} & 0 \end{bmatrix} \rightarrow \mathbf{K_C} = \begin{bmatrix} 0 & K_{12} & 0 \\ K_{21} & 0 & K_{23} \\ 0 & 0 & 1 \end{bmatrix} \tag{25}$$

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

$$\mathbf{C} = \lim_{n\rightarrow\infty} (\mathbf{K_C})^n \mathbf{e_p} \tag{26}$$

where $\mathbf{e_p}$ is a unit vector whose all elements are zero except for the one corresponding to the final milestone. Multiple powers of $\mathbf{K_C}$ are computed numerically until the committor converges. The results are considered to be converged when the change in the norm of the $\mathbf{C}$ vector to be 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 on the millisecond timescale protein-ligand unbinding in the trypsin-benzamidine system. In the first two systems we performed long equilibrium simulation and Markovian milestoning simulation to compare with our M-WEM results.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Müller-Brown Potential

The two dimensional Müller-Brown potential[78] is defined as

$$U(x,y) = h\sum_{i=1}^{4} A_i \exp[a_i(x-x_{0,i})^2 + b_i(x-x_{0,i})(y-y_{0,i}) + c_i(y-y_{0,i})^2] \tag{27}$$

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 non-linear transition path with barrier height about 4.5 k_BT. For the purpose of this model we set $k_BT = 1$.

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 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 milestoning simulations 8 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 transition from $y = 1.5$ to $y = 0.0$ for both the milestoning based methods.

For Markovian milestoning (MMVT) simulations, an overdamped Langevin dynamics simulation is propagated in 2D in all the seven cells in the spacing between 8 milestones. A half-harmonic wall is applied at both ends of the cell (milestones) with a force constant $\widetilde{k} = 1000$, 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 theory section and the MFPT has been computed.

For the M-WEM method, the procedure remained identical to the MMVT approach

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

except weighted ensemble simulation was performed in between the milestone interfaces as opposed to conventional dynamics. The 2D adaptive binning (MAB)[29] was employed along 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.[29] A total of 300 iterations of WE simulation are performed in each cell with a recycling interval of 10 steps.

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.[39] The 22-atom system is modelled using the CHARMM22 force field[79] with a 10 Å cut-off distance for the inter-atomic interactions. A time step of 0.5 fs was used and all the bonds between heavy atoms and hydrogen were constrained using the SHAKE algorithm.[80] All simulations were performed using the NAMD 2.13 package[81] with the colvars[82] 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 kcal mol^{-1} deg^{-2}) are placed at the 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).[39] 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 kcal/mol) and it is very difficult to observe direct transitions at room temperature. Following the earlier work[39] we performed the simulations at 600K temperature which allowed us to sample 285 transitions between the two free energy minima in 500 ns equilibrium simulation.

For both MMVT and M-WEM simulation, the reaction coordinate was chosen to be the Φ dihedral angle and milestones were placed at $\Phi = -80^o$, -60°, -40°, -20°, 0°, 20°, 40°, 60°, and 80°. The initial and final states were chosen to be the milestones at $\Phi = -80^o$ and $\Phi =$

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

60°. In 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 kcal mol^{-1} deg^{-2} were 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 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 very frequent interval (10 fs) to record the time of milestone crossings as accurately as possible. A total 100 iterations of WE simulation are performed in each cell with a recycling time of 1 ps.

Trypsin-Benzamidine Complex

The system setup for the trypsin-benzamidine complex is identical to the work by Votapka et al.[40] The structure, parameter and topology files were obtained from the authors of Ref. 40. We point the reader to their original publication[40] for more details. To mention briefly, the atomic coordinates of the protein-ligand complex were obtained from Protein Data Bank (PBD) PDB ID: 3PTB.[83] 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.[71] The protein was modelled using AMBER ff14SB force field[84] and Generalized Amber Force Field (GAFF)[85] parameters were used for the ligand. The structure was solvated in a truncated octahedron box of TIP4Pew[86] water molecules and 8 Cl^- ions were added to neutralize the system. Overall the system contains ∼23000 atoms. All MD simulations were performed using NAMD 2.14b2 package[81] 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 previous work,[40] 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 $\sim$ 1.5 Å/ns. The collective variables were biased and monitored using the colvars module.[82] Representative structures for seeding the milestoning 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[40,61] 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 timescales between one milestone to the other should be larger than the decay time of the velocity auto-correlation function of the RC. We checked this condition in our system as discussed in detail in the Supporting Information. A total of 12 cells were constructed in the spacing between 13 milestones. For each cell a flat bottom potential (Eqn. 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 the 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 RMSD of the ligand with respect to the representative structure (sampled from SMD) corresponding to the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

specific cell. The progress coordinates were recorded using the colvars module.[82] 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 iterations between 30 and 200 iterations, and at an interval of 20 iterations between 200 and 300 iterations. 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) was computed. This allowed us the 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. Before any calculation, the probabilities of the voronoi cells are modified ($\pi_\alpha \rightarrow \widetilde{\pi}_\alpha$) to take into account the Jacobian factor appearing due to the different surface area of milestones with different radius: $\widetilde{\pi}_\alpha = \pi_\alpha r_\alpha^2 / \sum_\alpha^\Lambda \pi_\alpha r_\alpha^2$ 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 simulation and the results are presented in Table 1. The MFPT of transition from milestone at $y = 1.5$ to $y = 0.0$, computed using M-WEM approach shows quantitative agreement with MFPT of the transition from minimum A to minimum B in 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 were comparable, the M-WEM simulations produced converged results with ~ 4 times less computational expense. Although the computational gain is not significant in case of this low dimensional model system, these results serve as a proof of concept of our method in rare

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

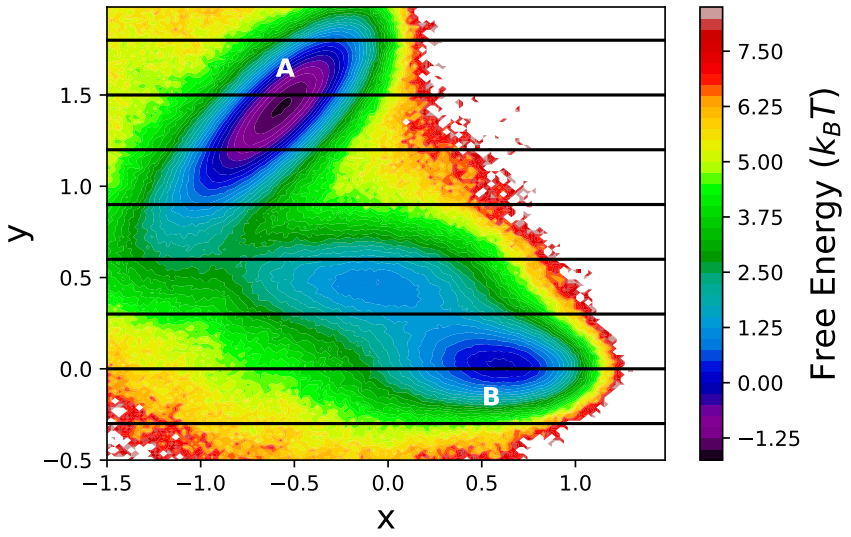

Figure 2: The free energy landscape of the Müller Brown potential explored using 10^7 steps of over-damped Langevin dynamics simulation. The position of the milestones, used in MMVT and M-WEM calculations, are shown in black lines. The two minima relevant to this study are marked as A and B.

event sampling problem. It also indicates that despite the choice of a poor and simplistic RC, accurate MFPTs can be calculated using M-WEM method.

Table 1: Results of conventional Langevin dynamics, MMVT and M-WEM simulations for the Müller Brown potential

Method	MFPT ($\times 10^3$)	Simulation time ($\times 10^6$)
Regular LD	25.2±2.9	10
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
M-WEM (trial 1)	23.6±2.3	2.6
M-WEM (trial 2)	20.4±0.4	2.6
M-WEM (trial 3)	28.1±1.6	2.6

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

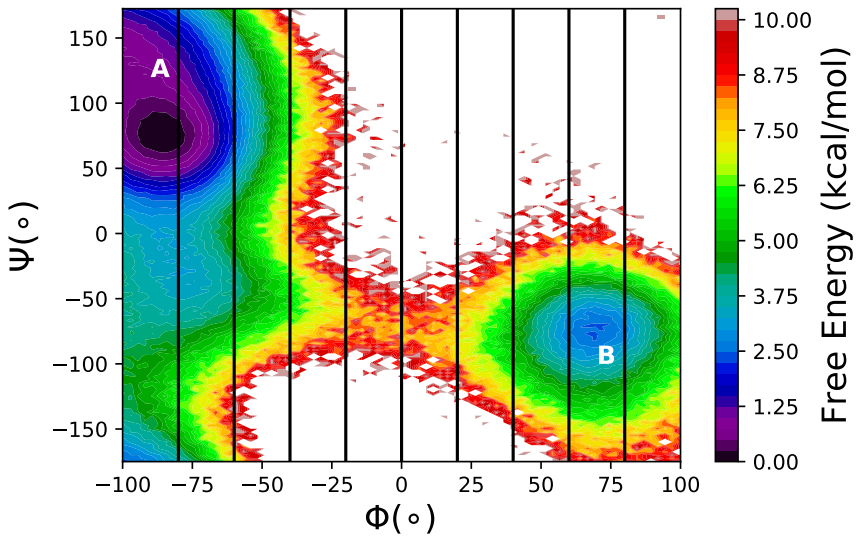

Figure 3: The free energy landscape of the gas phase Alanine dipeptide along the Φ and Ψ torsion angles, from 500ns equilibrium MD simulation. The position of the milestones, used in MMVT and M-WEM calculations, are shown in black lines. The two conformations relevant to this study are marked as A and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

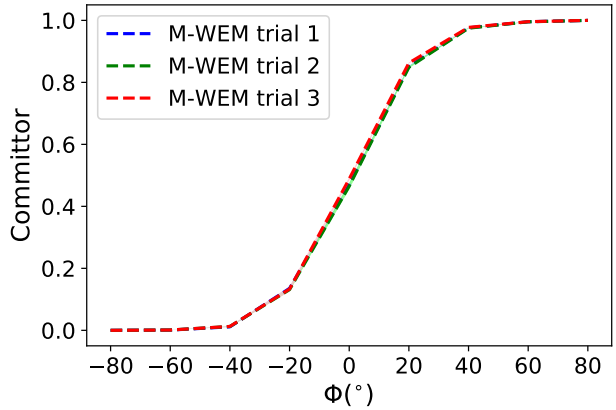

Figure 4: The committor values as a function of the milestoning coordinate Φ for the Alanine dipeptide system.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Alanine dipeptide (obtained from equilibrium MD simulation) is shown in Fig. 3. The mean first passage time (MFPT) of transition from milestone $\Phi = -80^o$ to milestone $\Phi = 60^o$ 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. Both these methods produced accurate results within one order of magnitude less computational cost in comparison the 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 Supporting Information).

The committor values at milestone interfaces for all three trials of M-WEM calculation are depicted in figure 4. The results from different trials are in excellent agreement with each other and all of them shows a committor value of ~ 0.5 for the milestone at $\Phi = 0°$. A committor value of 0.5 indicates the transition state (TS). The milestone at $\Phi = 0°$ is indeed present on top of the free energy barrier aka TS as evident from the free energy landscape in Fig. 3.

Table 2: Caption

Method	MFPT (ps)	Simulation time (ns)
Regular MD[a]	1583±188	500
MMVT trial 1	1465±167	40
MMVT trial 2	951±13	40
MMVT trial 3	984±8	40
M-WEM trial 1[b]	3031±757	86.4
M-WEM trial 2[b]	2540±556	84.4
M-WEM trial 3[b]	2379±446	83.2

[a]From state A to B in figure 3, not from milestone to milestone.
[b]At the end of 100 iterations

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

from the M-WEM calculation is shown in figure 5a. Then the true probability distribution ($p(\Phi, \Psi)$) is computed by re-scaling the distributions corresponding to each cell with the weight of their probabilities obtained using Eq. 3 (Fig. 5b).

$$p(\Phi, \Psi) = \sum_{\alpha=1}^{\Lambda} \pi_\alpha p^\alpha(\Phi, \Psi) \tag{28}$$

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

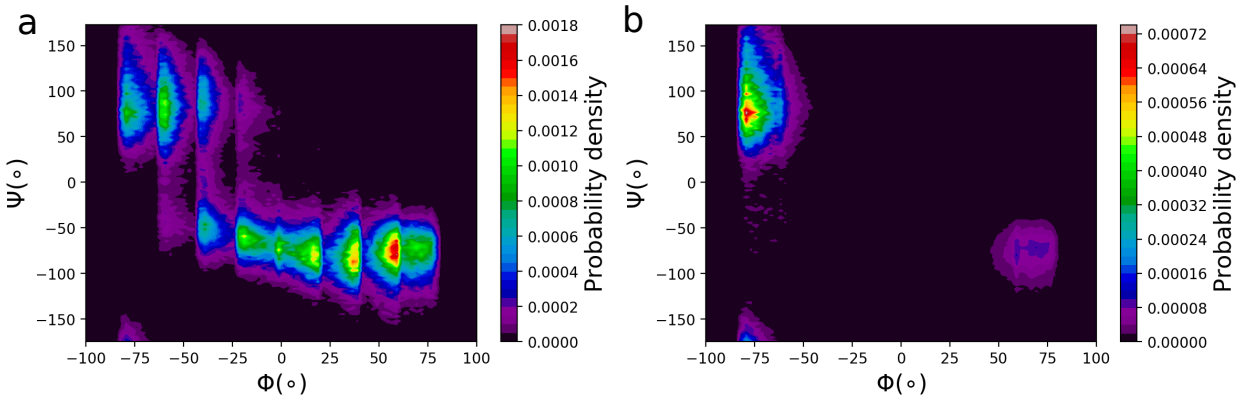

Figure 5: Reconstruction of equilibrium probability distribution (b) from raw unscaled probability distribution from M-WEM trajectories in each cell (a).

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

$$G(\Phi, \Psi) = -k_B T \ln \left(\sum_{\alpha=1}^{\Lambda} \pi_\alpha p^\alpha(\Phi, \Psi) \right) \tag{29}$$

The reconstructed free energy surface for both MMVT and M-WEM simulations are in excellent agreement with one obtained from 500 ns conventional MD simulation, but confined omly between the initial and final milestone i.e. $-80^o < \Phi < 80^o$. This provides a way to study the effect of orthogonal degrees of freedom which is coupled with the reaction coordinate but is not taken into account in devising the milestoning coordinate.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

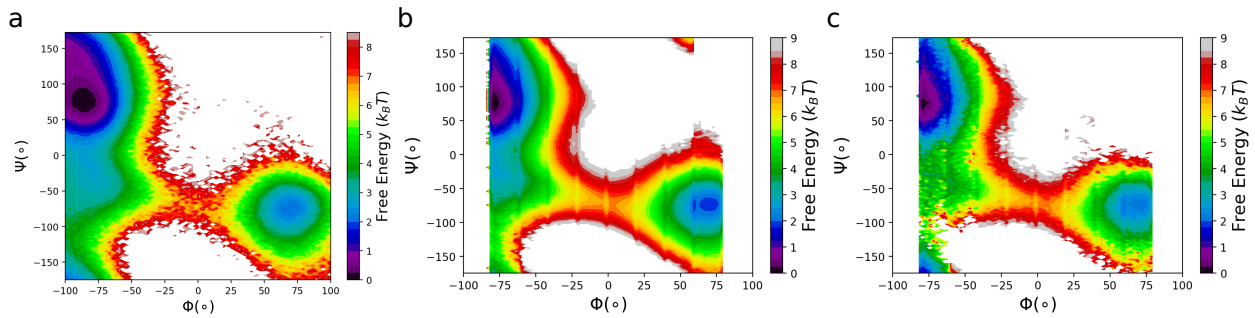

Figure 6: (a) Free energy landscape of gas phase alanine dipeptide obtained from equilibrium MD simulation. (b) Reconstructed free energy landscape from MMVT simulation. (c) Reconstructed free energy landscape from M-WEM simulation (trial 1). (For a better comparison the free energy landscape is constructed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

Trypsin-Benzamidine Complex

Finally, we applied the M-WEM approach to calculate the kinetics and free energy for a protein ligand binding and unbinding problem. We chose the system of 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 and is beyond the reach of 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.[71] This is a challenging enough test 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 unbinding free energy ΔG_u have been computed from M-WEM simulation and the results are compared with the SEEKR[40] and MMVT SEEKR[61] results (which used identical simulation condition as our work) and also with the experimental data[71] (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 M-WEM simulation is within the error bars of the experiment and within one order of magnitude of the SEEKR and MMVT SEEKR results. The same holds

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for residence time which is the inverse of k_{off}. Our k_{on} results are in agreement with SEEKR values but different from experimental value by a factor of $\sim$ 1.5-2. The k_{on} results of MMVT SEEKR are one order of magnitude higher. The ΔG_u 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 0.5 kcal/mol).

Table 3: Comparison of the results of the different milestoning based methods for the trypsin-benzamidine complex. (Number of iterations of M-WEM simulation are shown in parenthesis)

Method	Residence time (ms)	k_{off} (s^{-1})	k_{on} ($\times 10^7$ M^{-1} s^{-1})	ΔG_u (kcal/mol)	Simulation time (μs)
Experiment[71]	1.7	600±300	2.9	6.7±0.05	-
SEEKR[40]	12	83±14	2.1±0.3	7.4±0.10	∼19
MMVT SEEKR[61]	5.6	174±9	12±0.5	7.9±0.04	∼4.4
MMVT SEEKR[61]	16	62±6	17±1.0	8.8±0.07	∼2.9
M-WEM (200)	2.4±0.2	410±40	1.4±0.2[a]	6.3±0.10	∼0.48
M-WEM (300)	1.6±0.2	608±91	2.0±1.0[a]	6.2±0.32	∼0.73

[a]Error bars are computed from the last 5 iterations sampled.

The convergence paterns of the residence time and k_{off} are depicted in Fig. 7. Both these values converged after about 150 iterations ($\sim$ 360 ns of total simulation time) except for small fluctuations. (As k_{off} is computed indirectly as the inverse of residence time directly obtained from M-WEM, small fluctuations in the residence time get amplified in the k_{off} results in Fig. 7a.) The computational cost of the M-WEM simulation is $\sim$ 1 order of magnitude less than the other milestoning-based approaches[40,61] 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. ($(9.5 \pm 3.3) \times 10^4$ s^{-1}),[67] Plattner et al. ($(1.31 \pm 1.09) \times 10^4$ s^{-1}),[68] Tiwary et al. (9.1 ± 2.5 s^{-1}),[69] and Teo et al. (260 ± 240 s^{-1});[70] all these studies required multiple microseconds of simulation with some in the range of 50 μs - 100 μs.[67,68] A weighted ensemble-based approach has also been used to calculate the kinetics of this system by Dickson and Lotz ($k_{\text{off}} = 5555$ s^{-1})[32] and Donyapour et al. ($k_{\text{off}} = 266$ s^{-1} and 840 s^{-1}).[26] But, unlike M-WEM, that method could only calculate the unbinding rate constant and dissociation pathways due to the use of non-equilibrium steady state. To their

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

credit, the authors could distinguish multiple ligand release pathways,[32] which is difficult to achieve using milestoning-based simulations with discontinuous trajectories. Nevertheless, we tried to identify some of the key intermediates in the unbinding mechanism; we discuss them later in this paper.

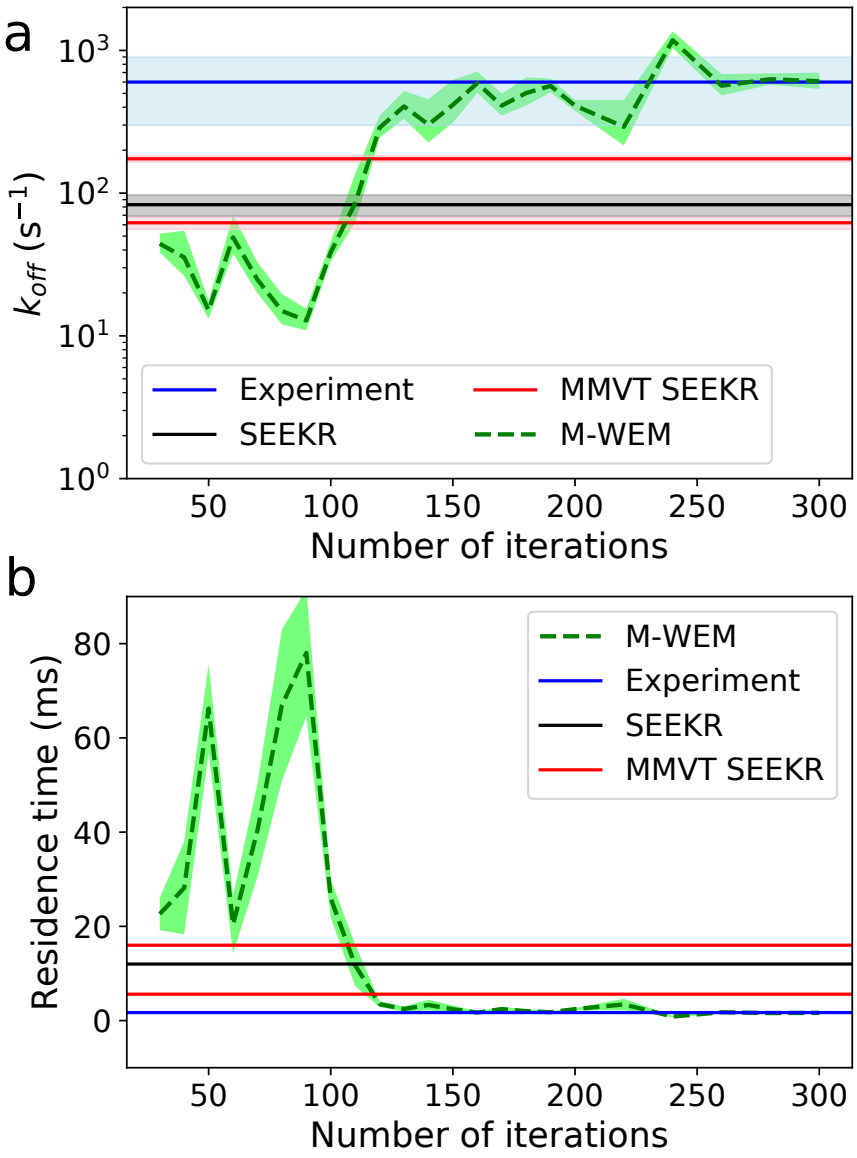

Figure 7: The convergence of (a) k_{off} and (b) ligand residence time for trypsin-benzamidine complex, as a function of M-WEM iterations. In figure (b) a linear scale is used for a better idea of the quality of the convergence.

A one-dimensional free energy profile as a function of the milestoning reaction coordinate is constructed from the equilibrium probabilities (π_α) obtained from the M-WEM simulation

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

using Eq. 4. Alongside a one dimensional free energy profile is reconstructed from the M-WEM trajectories following section Theory. Error bars in the free energy landscape are computed as the 95% confidence interval of 5 free energy profiles obtained between iteration 160 and 200 with an interval of 10 iterations. The two free energy profiles obtained from M-WEM using two different techniques agree with each other and both are in reasonable agreement with the free energy surface obtained using fast converging well tempered meta-eABF (WTM-eABF) simulation[87,88] (see Supporting Information for details).

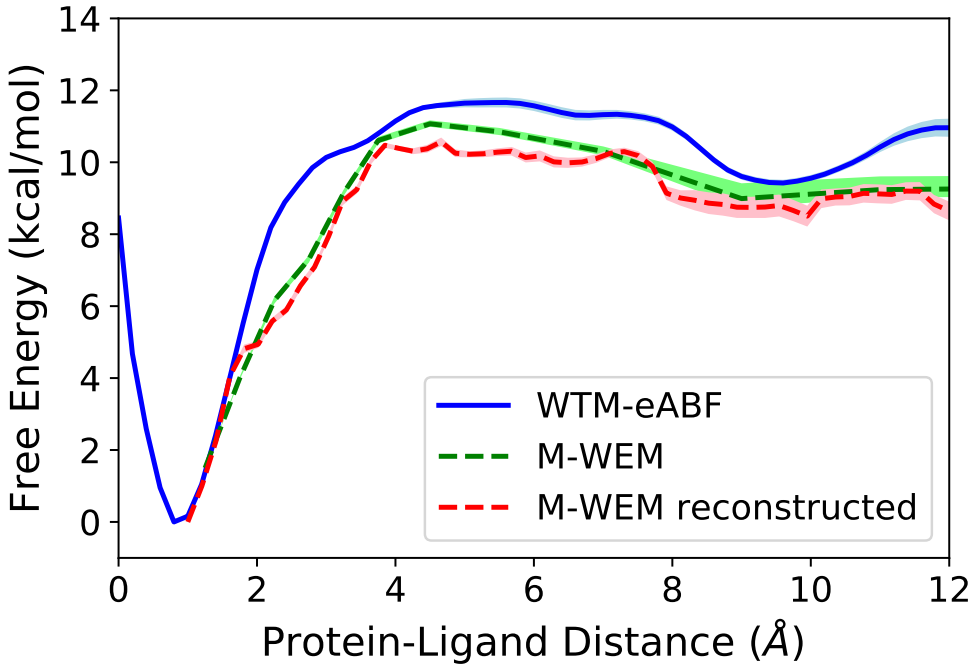

Figure 8: Free energy profile of the dissociation of the trypsin-benzamidine complex as a function of the milestoning reaction coordinate (the center of mass distance between the binding pocket residues and the benzamidine ligand.) The results are compared between WTM-eABF simulation and the M-WEM calculation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

The committor values as a function of the milestoning 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 milestone at 6 Å and 8 Å.

The distribution of the ligand around the protein for three cells (bound state, unbound state and the cell containing the TS) have been depicted in Fig. 10. It shows the amount

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

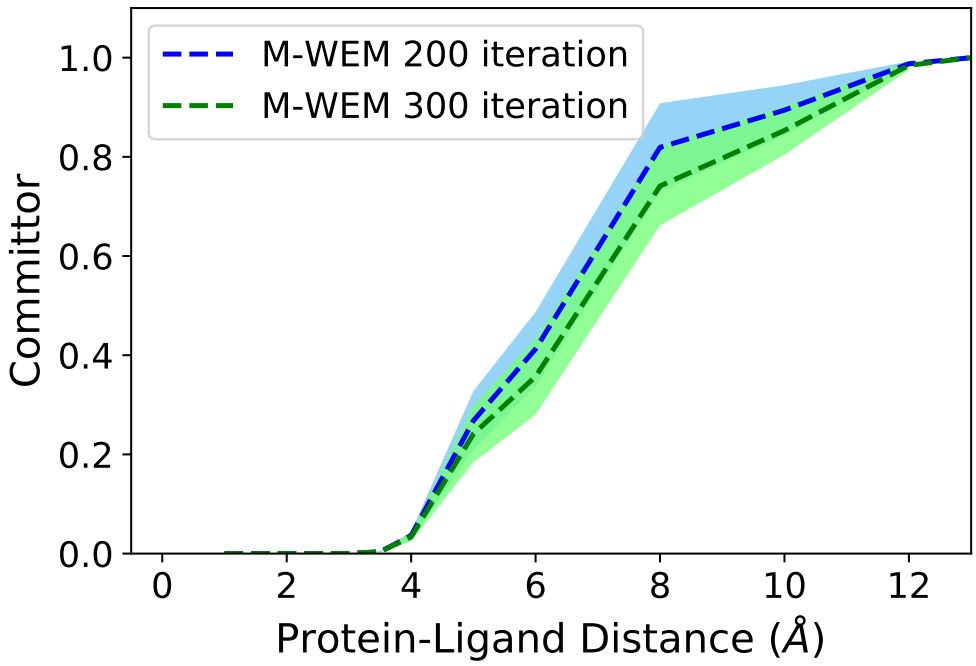

Figure 9: Committor values computed as a function of the milestoning reaction coordinate for the M-WEM simulation of the trypsin-benzamidine complex.

of three dimensional space explored by the ligand in the process of dissociation. A two dimensional projection of the ligand distribution for all cells is shown in Fig. 11. The fraction of the spherical surface covered by the ligand in the outermost cell (α) is used for the calculation of k_{on}, as described in the Theory section. There is a slight increase in the exploration of the configuration space after 300 iterations in comparison to 200 iterations, but the difference is not significant.

To get an idea of the intermediate states involved in ligand dissociation we clustered all the trajectory frames corresponding to each cell based on heavy atom RMSD. The number of frames in each cell ranged between $\sim 26,000 - 30,000$ with one frame every 2 ps (the length of each WE segment). The clustering of the structures was performed using the GROMOS clustering algorithm[89] implemented in GROMACS v2018.1[90] with an RMSD cutoff of 0.9 Å. The cut-off was chosen such that the total number of clusters are between 10 to 40. All the cluster centers obtained from different cells were combined together and a second round of clustering is performed with and RMSD cutoff of 1.1 Å. This resulted in 14 clusters, some

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

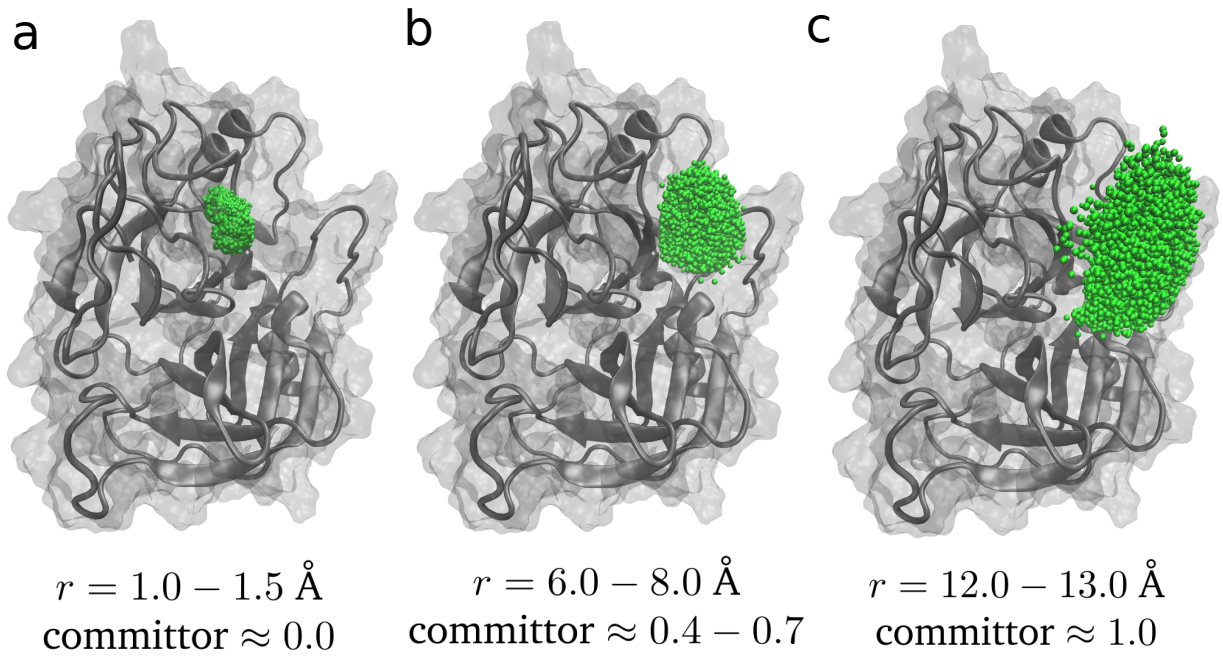

Figure 10: The distribution of the benzamidine ligand around the trypsin protein for three cells of the M-WEM simulations which, respectively, include : (a) the bound state, (b) an apparent transition state with committor value ~ 0.5, and (c) the unbound state.

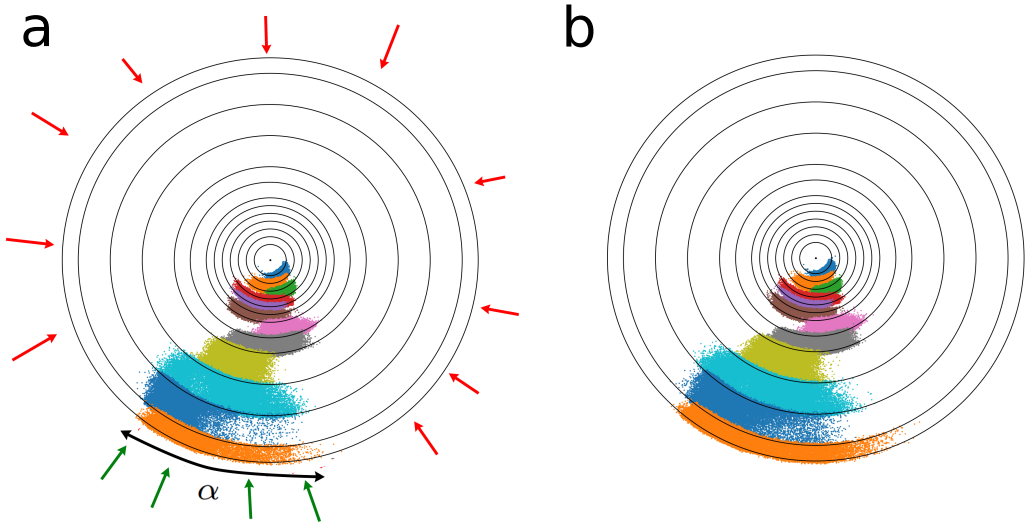

Figure 11: Two dimensional projection of the distribution of the ligands around the trypsin protein for (a) after M-WEM iteration 200 and (b) after iteration 300. The different colors represents structures from M-WEM simulations confined in different cells. The surface coverage α, used in the k_{on}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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

of which are depicted in Fig. 12. The structures are in qualitative agreement with the meta-stable states observed by Tiwary et al.,[69] despite their use of a different enhanced sampling method and a different version of the AMBER force field. Particularly, both our study and the work of Tiwary et al. show the presence of a meta-stable 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 the clusters is provided in the Supporting Information.

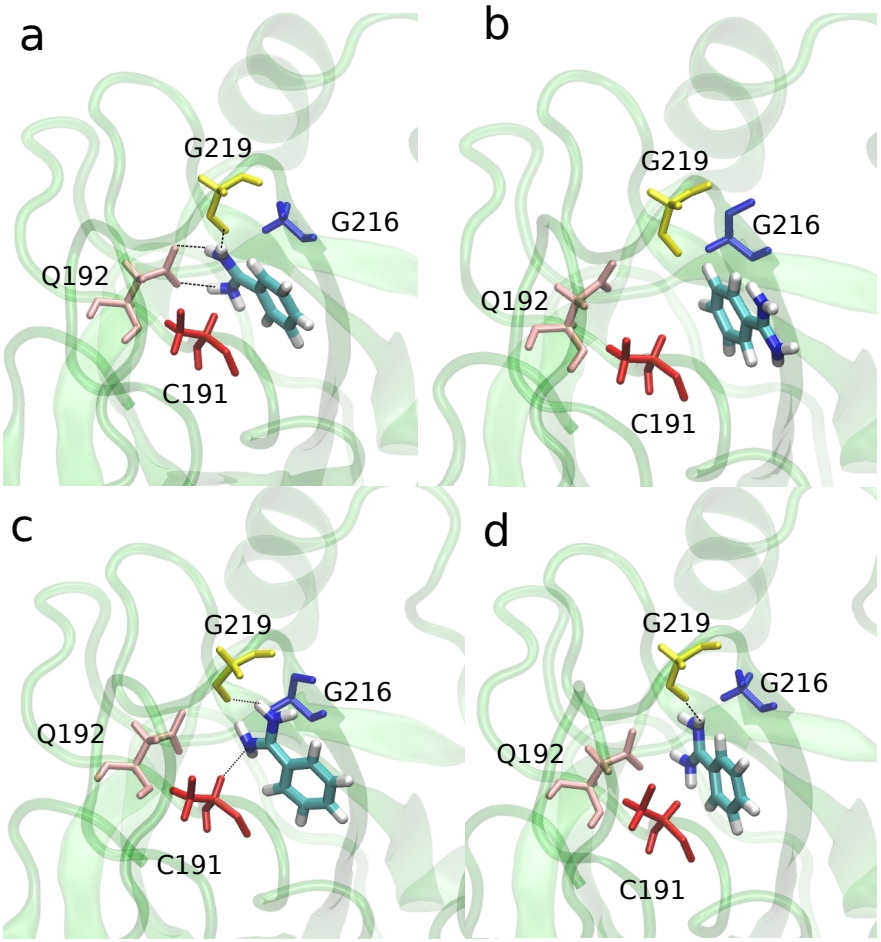

Figure 12: Representative structures sampled from clustering of the M-WEM trajectories of the binding/unbinding of trypsin-benzamidine complex. The ligand and the residues interaction with the ligand are shown in licorice. Hydrogen bonds between protein and ligand are shown in dashed line.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Discussions 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 of 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 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 significantly lower computational cost. In the case of alanine dipeptide we could 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 re-scaling strategy based on the equilibrium probabilities of each milestone. This approach can be generalized to any other collective variable other than the milestoning coordinate, and can potentially elucidate roles 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 M-WEM method in about one 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 excellent 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 makes it very easy to implement in any simulation engine using an open access weighted ensemble

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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 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 interface. Moreover, in contrast to traditional milestoning based approaches, M-WEM (or Markovian milestoning in general) does not require additional simulation (e.g. umbrella sampling) along the milestone interface to sample starting structures, a process which accounts for more than 80% of the total computational effort. In our previous work, we attempted to replace this expensive additional step using an weighted ensemble based restrain-and-release scheme.[66] The Markovian Milestoning based technique completely removes this step as the trajectory, confined between two milestones, by itself explores the configurational space orthogonal to the milestoning coordinate. In the M-WEM approach, we accelerated this "orthogonal sampling", by using 2D WE bins along two progress coordinates: the milestoning reaction coordinate (to accelerate milestone to milestone transitions) and also in another coordinate along the milestone interface. The advantage of the M-WEM over traditional Markovian milestoning is 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 interface, avoiding frequent intervention to the dynamics engine, making it more efficient to implement in GPU-based hardware.

Apart from these unique achievements, M-WEM also shares some of the common advantages of 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

The convergence of the transition statistics in between milestones is also quicker in M-WEM in comparison to 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 case of the Müller Brown potential and alanine dipeptide.

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,[40,46,61] which is more rigorous but more computationally expensive. However, the BD methods allow us to include the effect of position dependent variation of the diffusion constant, as well as the ionic strength of the solution which are absent in our current implementation.

Our M-WEM method can find application in studying the kinetics and free energy of biomolecular rare events, not only for the purpose of fundamental understanding biological processes, but also for kinetics-driven computer aided drug design. Evidences emerged over the past decade 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 because it is easier to compute than kinetics. The M-WEM approach is a cheap alternative to the computationally expensive traditional enhanced sampling and path sampling based methods, and can be included in a computational drug design pipeline using both binding free energy and kinetics. In future, we plan to test this method on protein-ligand systems with longer residence time such as in the range of minutes to hours which is more characteristic of the drug molecules used in practical application. The increased sampling of orthogonal coordinates in M-WEM can also facilitate the study of systems where a protein conformational change is coupled to the ligand binding coordinate, which we also hope to look at in future work. Overall, our novel Markovian Weighted Ensemble Milestoning approach has been successful in predicting the free energy and kinetics of biophysical rare events with quantitative accuracy and holds the potential of becoming an essential tool in large scale

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

computational screening of therapeutic drugs.

Acknowledgement

The authors thank Rommie Amaro and Lane Votapka for kindly sharing the structure and topology files of the trypsin-benzamidine complex. The authors thank 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. SES acknowledges support through Pfizer - La Jolla Academic Industrial Relations (AIR) Diversity Research Fellowship in Chemistry. The authors thank the UC Irvine High Performance Computing (HPC) facility for providing the computational resources. The authors declare no competing financial interest.

Data Availability Statement

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. All 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.

Supporting Information Available

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, and discussion on the appropriate placement of milestones are provided in the supporting information. Fig. S1-S3 are included in the supporting information.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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. *Journal of the American Chemical Society* **2015**, *137*, 2695–2703.

(2) Mobley, D. L.; Gilson, M. K. Predicting Binding Free Energies: Frontiers and Benchmarks. *Annual Review of Biophysics* **2017**, *46*, 531–558.

(3) Wereszczynski, J.; McCammon, J. A. Statistical mechanics and molecular dynamics in evaluating thermodynamic properties of biomolecular recognition. 2012; `https://doi.org/10.1017/S0033583511000096`.

(4) Torrie, G.; Valleau, J. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.

(5) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 12562–6.

(6) Darve, E.; Pohorille, A. Calculating free energies using average force. *The Journal of Chemical Physics* **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. *The Journal of Physical Chemistry Letters* **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. *British Journal of Pharmacology* **2012**, *166*, 1846–1859.

(9) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nature Reviews Drug Discovery* **2006**, *5*, 730–739.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(10) Dellago, C.; Bolhuis, P. G.; Csajka, F. S.; Chandler, D. Transition path sampling and the calculation of rate constants. *The Journal of Chemical Physics* **1998**, *108*, 1964–1977.

(11) Bolhuis, P. G.; Chandler, D.; Dellago, C.; Geissler, P. L. Transition Path Sampling : Throwing Ropes Over Rough Mountain Passes, in the Dark. *Annual Review of Physical Chemistry* **2002**, *53*, 291–318.

(12) Chong, L. T.; Saglam, A. S.; Zuckerman, D. M. Path-sampling strategies for simulating rare events in biomolecular systems. *Current Opinion in Structural Biology* **2017**, *43*, 88–94.

(13) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annual Review of Biophysics* **2017**, *46*, 43–57.

(14) Huber, G.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophysical Journal* **1996**, *70*, 97–110.

(15) 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. *The Journal of Chemical Physics* **2010**, *132*, 054107.

(16) 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. *Journal of Chemical Theory and Computation* **2015**, *11*, 800–809.

(17) Abdul-Wahid, B.; Feng, H.; Rajan, D.; Costaouec, R.; Darve, E.; Thain, D.; Izaguirre, J. A. AWE-WQ: Fast-Forwarding Molecular Dynamics Using the Accelerated Weighted Ensemble. *Journal of Chemical Information and Modeling* **2014**, *54*, 3033–3043.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(18) Adhikari, U.; Mostofian, B.; Copperman, J.; Subramanian, S. R.; Petersen, A. A.; Zuckerman, D. M. Computational Estimation of Microsecond to Second Atomistic Folding Times. *Journal of the American Chemical Society* **2019**, *141*, 6519–6526.

(19) 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. *Journal of Chemical Theory and Computation* **2014**, *10*, 2658–2667.

(20) 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. *Journal of Chemical Theory and Computation* **2011**, *7*, 1189–1197.

(21) 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. *The Journal of Physical Chemistry Letters* **2016**, *7*, 3440–3445.

(22) Saglam, A. S.; Chong, L. T. Protein-protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chemical Science* **2019**, *10*, 2360–2372.

(23) Adelman, J. L.; Grabe, M. Simulating Current–Voltage Relationships for a Narrow Ion Channel Using the Weighted Ensemble Method. *Journal of Chemical Theory and Computation* **2015**, *11*, 1907–1918.

(24) Spiriti, J.; Zuckerman, D. M. Tabulation as a high-resolution alternative to coarse-graining protein interactions: Initial application to virus capsid subunits. *The Journal of Chemical Physics* **2015**, *143*, 243159.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(25) Dickson, A.; Brooks, C. L. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *The Journal of Physical Chemistry B* **2014**, *118*, 3532–3542.

(26) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. *Journal of Chemical Physics* **2019**, *150*, 244112.

(27) Suárez, E.; Adelman, J. L.; Zuckerman, D. M. Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. *Journal of Chemical Theory and Computation* **2016**, *12*, 3473–3481.

(28) Degrave, A. J.; Bogetti, A. T.; Chong, L. T. The RED scheme: Rate-constant estimation from pre-steady state weighted ensemble simulations. *Journal of Chemical Physics* **2021**, *154*, 100401.

(29) Torrillo, P. A.; Bogetti, A. T.; Chong, L. T. A Minimal, Adaptive Binning Scheme for Weighted Ensemble Simulations. *Journal of Physical Chemistry A* **2021**, *125*, 27.

(30) Copperman, J.; Zuckerman, D. M. Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian "microbin" Analysis. *Journal of Chemical Theory and Computation* **2020**, *16*, 6763–6775.

(31) Dickson, A.; Lotz, S. D. Ligand Release Pathways Obtained with WExplore: Residence Times and Mechanisms. *Journal of Physical Chemistry B* **2016**, *120*, 5377–5385.

(32) Dickson, A.; Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophysical Journal* **2017**, *112*, 620–629.

(33) Dixon, T.; Uyar, A.; Ferguson-Miller, S.; Dickson, A. Membrane-mediated ligand unbinding of the PK-11195 ligand from the translocator protein (TSPO). *bioRxiv* **2020**, 2020.01.21.914127.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(34) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. *Journal of the American Chemical Society* **2018**, *140*, 618–628.

(35) Faradjian, A. K.; Elber, R. Computing time scales from reaction coordinates by milestoning. *The Journal of Chemical Physics* **2004**, *120*, 10880–10889.

(36) West, A. M. A.; Elber, R.; Shalloway, D. Extending molecular dynamics time scales with milestoning: Example of complex kinetics in a solvated peptide. *The Journal of Chemical Physics* **2007**, *126*, 145104.

(37) Bello-Rivas, J. M.; Elber, R. Exact milestoning. *The Journal of Chemical Physics* **2015**, *142*, 094102.

(38) Elber, R.; Bello-Rivas, J.; Ma, P.; Cardenas, A.; Fathizadeh, A. Calculating Iso-Committor Surfaces as Optimal Reaction Coordinates with Milestoning. *Entropy* **2017**, *19*, 219.

(39) Wei, W.; Elber, R. ScMile: A Script to Investigate Kinetics with Short Time Molecular Dynamics Trajectories and the Milestoning Theory. *Journal of Chemical Theory and Computation* **2020**, *16*, 860–874.

(40) 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. *The Journal of Physical Chemistry B* **2017**, *121*, 3597–3606.

(41) Elber, R. A Milestoning Study of the Kinetics of an Allosteric Transition: Atomically Detailed Simulations of Deoxy Scapharca Hemoglobin. *Biophysical Journal* **2007**, *92*, L85–L87.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(42) Votapka, L. W.; Lee, C. T.; Amaro, R. E. Two Relations to Estimate Membrane Permeability Using Milestoning. *The Journal of Physical Chemistry B* **2016**, *120*, 8606–8616.

(43) Fathizadeh, A.; Elber, R. Ion Permeation through a Phospholipid Membrane: Transition State, Path Splitting, and Calculation of Permeability. *Journal of Chemical Theory and Computation* **2019**, *15*, 720–730.

(44) Cardenas, A. E.; Elber, R. Markovian and Non-Markovian Modeling of Membrane Dynamics with Milestoning. *The Journal of Physical Chemistry B* **2016**, *120*, 8208–8216.

(45) Vanden-Eijnden, E.; Venturoli, M. Markovian milestoning with Voronoi tessellations. *The Journal of Chemical Physics* **2009**, *130*, 194101.

(46) Votapka, L. W.; Amaro, R. E. Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. *PLoS Computational Biology* **2015**, *11*.

(47) 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. *Journal of Chemical Theory and Computation* **2020**, *16*, 1882–1895.

(48) Narayan, B.; Buchete, N.-V.; Elber, R. Computer Simulations of the Dissociation Mechanism of Gleevec from Abl Kinase with Milestoning. *The Journal of Physical Chemistry B* **2021**, acs.jpcb.1c00264.

(49) Jagger, B. R.; Lee, C. T.; Amaro, R. E. Quantitative Ranking of Ligand Binding Kinetics with a Multiscale Milestoning Simulation Approach. *Journal of Physical Chemistry Letters* **2018**, *9*, 4941–4948.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(50) 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. *Journal of the American Chemical Society* **2017**, *139*, 14837–14840.

(51) 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. *Journal of Physical Chemistry B* **2018**, *122*, 10296–10305.

(52) Wang, H.; Elber, R. Catalytic Magnesium as a Door Stop for DNA Sliding. *The Journal of Physical Chemistry B* **2021**, *125*, 3494–3500.

(53) 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. *Biochimica et Biophysica Acta - General Subjects* **2020**, *1864*, 129508.

(54) Maragliano, L.; Vanden-Eijnden, E.; Roux, B. Free energy and kinetics of conformational transitions from voronoi tessellated milestoning with restraining potentials. 2009; `https://pubs.acs.org/sharingguidelines`.

(55) Alberini, G.; Benfenati, F.; Maragliano, L. Molecular Dynamics Simulations of Ion Selectivity in a Claudin-15 Paracellular Channel. *Journal of Physical Chemistry B* **2018**, *122*, 10783–10792.

(56) Cottone, G.; Chiodo, L.; Maragliano, L. Thermodynamics and kinetics of ion permeation in wild-type and mutated open active conformation of the human α7 nicotinic receptor. *Journal of Chemical Information and Modeling* **2020**, *60*, 5045–5056.

(57) Alberini, G.; Benfenati, F.; Maragliano, L. Structural Mechanism of ω-Currents in a Mutated Kv7.2 Voltage Sensor Domain from Molecular Dynamics Simulations. *Journal of Chemical Information and Modeling* **2021**, *61*, 1354–1367.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(58) Jiang, W.; Lin, Y.-C.; Botello-Smith, W.; Contreras, J.; Harris, A. L.; Maragliano, L.; Luo, Y. L. Free energy and kinetics of cAMP permeation through connexin26 hemichannel with and without voltage. *bioRxiv* **2021**,

(59) Ma, W.; Schulten, K. Mechanism of substrate translocation by a ring-shaped ATPase motor at millisecond resolution. *Journal of the American Chemical Society* **2015**, *137*, 3031–3040.

(60) 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. *Journal of the American Chemical Society* **2015**, *137*, 3041–3050.

(61) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *Journal of Chemical Theory and Computation* **2020**, *16*, 5348–5357.

(62) Grazioli, G.; Andricioaei, I. Advances in milestoning. I. Enhanced sampling via wind-assisted reweighted milestoning (WARM). *The Journal of Chemical Physics* **2018**, *149*, 084103.

(63) Wang, H.; Elber, R. Milestoning with wind: Exploring the impact of a biasing potential in exact calculation of kinetics. *The Journal of Chemical Physics* **2020**, *152*, 224105.

(64) Ray, D.; Andricioaei, I. Weighted ensemble milestoning (WEM): A combined approach for rare event simulations. *The Journal of Chemical Physics* **2020**, *152*, 234114.

(65) 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. *Journal of Chemical Theory and Computation* **2017**, *13*, 3372–3377.

(66) Ray, D.; Gokey, T.; Mobley, D. L.; Andricioaei, I. Kinetics and free energy of ligand

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

dissociation using weighted ensemble milestoning. *The Journal of Chemical Physics* **2020**, *153*, 154117.

(67) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 10184–9.

(68) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature Communications* **2015**, *6*, 1–10.

(69) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences of the United States of America* **2015**, *112*, E386–E391.

(70) Teo, I.; Mayne, C. G.; Schulten, K.; Lelièvre, T. Adaptive Multilevel Splitting Method for Molecular Dynamics Calculation of Benzamidine-Trypsin Dissociation Time. *Journal of Chemical Theory and Computation* **2016**, *12*, 2983–2989.

(71) Guillain, F.; Thusius, D. The use of Proflavin as an Indicator in Temperature-Jump Studies of the Binding of a Competitive Inhibitor to Trypsin. 1970; `https://pubs.acs.org/doi/abs/10.1021/ja00721a051`.

(72) Májek, P.; Elber, R. Milestoning without a Reaction Coordinate. *Journal of Chemical Theory and Computation* **2010**, *6*, 1805–1817.

(73) Vanden-Eijnden, E.; Venturoli, M.; Ciccotti, G.; Elber, R. On the assumptions underlying milestoning. *The Journal of Chemical Physics* **2008**, *129*, 174102.

(74) Noé, F. Probability distributions of molecular observables computed from Markov models. *Journal of Chemical Physics* **2008**, *128*, 244103.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(75) McCammon, J. A.; Northrup, S. H.; Allison, S. A. Diffusional dynamics of ligand-receptor association. 1986; `https://pubs.acs.org/doi/abs/10.1021/j100408a015`.

(76) Berezhkovskii, A. M.; Szabo, A. Committors, first-passage times, fluxes, Markov states, milestones, and all that. *Journal of Chemical Physics* **2019**, *150*, 54106.

(77) Ma, N.; Van Der Vaart, A. Free Energy Coupling between DNA Bending and Base Flipping. *Journal of Chemical Information and Modeling* **2017**, *57*, 2020–2026.

(78) Müller, K.; Brown, L. D. Location of saddle points and minimum energy paths by a constrained simplex optimization procedure. *Theoretica Chimica Acta* **1979**, *53*, 75–93.

(79) Huang, J.; Mackerell, A. D. CHARMM36 all-atom additive protein force field: Validation based on comparison to NMR data. *Journal of Computational Chemistry* **2013**, *34*, 2135–2145.

(80) 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. *Journal of computational physics* **1977**, *23*, 327–341.

(81) 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. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.

(82) Fiorin, G.; Klein, M. L.; Hénin, J. Using collective variables to drive molecular dynamics simulations. *Molecular Physics* **2013**, *111*, 3345–3362.

(83) 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 Crystallographica Section B* **1983**, *39*, 480–490.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.26.450057; this version posted June 28, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

(84) 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. *Journal of Chemical Theory and Computation* **2015**, *11*, 3696–3713.

(85) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general Amber force field. *Journal of Computational Chemistry* **2004**, *25*, 1157–1174.

(86) 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. *Journal of Chemical Physics* **2004**, *120*, 9665–9678.

(87) Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free Energy Landscapes Using an Average Force. *Accounts of Chemical Research* **2019**,

(88) Fu, H.; Chen, H.; Wang, X.; Chai, H.; Shao, X.; Cai, W.; Chipot, C. Finding an optimal pathway on a multidimensional free-energy landscape. *Journal of Chemical Information and Modeling* **2020**, *60*, 5366–5374.

(89) Daura, X. et al. *Peptide Folding: When Simulation Meets Experiment*; 1998; Vol. 31; pp 1387–1404.

(90) 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.