Title

Gaussian-Accelerated Molecular Dynamics with the Weighted Ensemble Method: A Hybrid Method Improves Thermodynamic and Kinetic Sampling

Journal

Journal of Chemical Theory and Computation, 17(12)

ISSN

1549-9618

Authors

Ahn, Surl-Hee

Ojha, Anupam A

Amaro, Rommie E

et al.

Publication Date

2021-12-14

DOI

10.1021/acs.jctc.1c00770

Published in final edited form as:

J Chem Theory Comput. 2021 December 14; 17(12): 7938–7951. doi:10.1021/acs.jctc.1c00770.

Gaussian-Accelerated Molecular Dynamics with the Weighted Ensemble Method: A Hybrid Method Improves Thermodynamic and Kinetic Sampling

Surl-Hee Ahn[§],
Department of Chemistry, University of California San Diego, La Jolla 92093 California, United States

Anupam A. Ojha[§],
Department of Chemistry, University of California San Diego, La Jolla 92093 California, United States

Rommie E. Amaro,
Department of Chemistry, University of California San Diego, La Jolla 92093 California, United States

J. Andrew McCammon
Department of Chemistry, University of California San Diego, La Jolla 92093 California, United States; Department of Pharmacology, University of California San Diego, La Jolla 92093 California, United States

Abstract

Gaussian-accelerated molecular dynamics (GaMD) is a well-established enhanced sampling method for molecular dynamics simulations that effectively samples the potential energy landscape of the system by adding a boost potential, which smoothens the surface and lowers the energy barriers between states. GaMD is unable to give time-dependent properties such as kinetics directly. On the other hand, the weighted ensemble (WE) method can efficiently sample transitions between states with its many weighted trajectories, which directly yield rates and pathways. However, convergence to equilibrium conditions remains a challenge for the WE method. Hence, we have developed a hybrid method that combines the two methods, wherein GaMD is first used to sample the potential energy landscape of the system and WE is subsequently used to further sample the potential energy landscape and kinetic properties of interest. We show that the hybrid method can sample both thermodynamic and kinetic properties more accurately and quickly compared to using either method alone.

Graphical Abstract

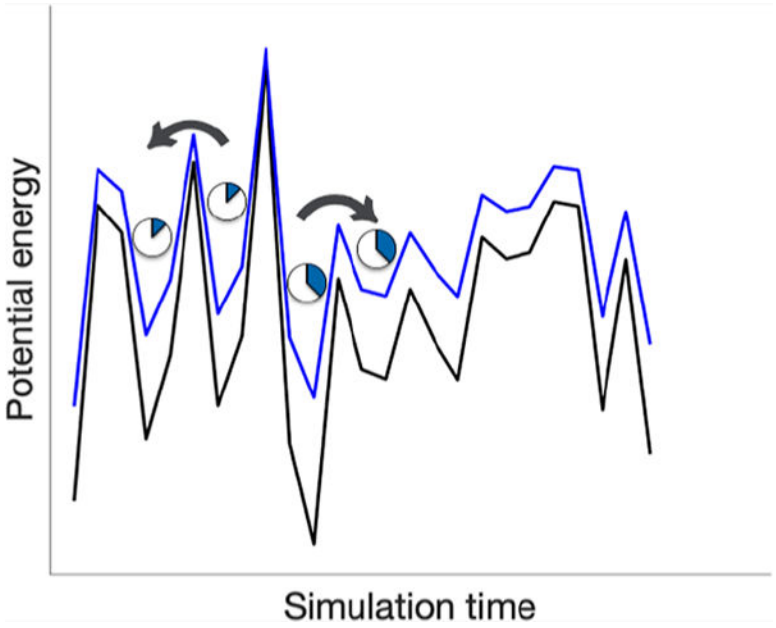
Corresponding Author: Surl-Hee Ahn – *Department of Chemistry, University of California San Diego, La Jolla 92093 California, United States;* s3ahn@ ucsd.edu.

[§] **Author Contributions**

S.-H.A. and A.A.O. have contributed equally.

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.1c00770

The authors declare no competing financial interest.



1. INTRODUCTION

Molecular dynamics (MD) simulations are becoming quintessential tools in many fields, including biology,[1–3] chemistry,[4–7] materials science,[8,9] and chemical and biological engineering.[10,11] An increasing number of researchers have used MD simulations to uncover mechanisms of their biological system of interest in atomistic detail. Applications of MD simulations range from studying protein folding[3,6,12–14] and protein–protein or protein–ligand interactions[15,16] to computer-aided drug design (virtual screening and ligand docking).[4,5,7] However, MD simulations are not without their challenges. MD simulations have to be run using femtosecond time steps due to being limited by the fastest motions in the system. In contrast, biological processes of interest are on the order of microseconds or longer. Additionally, systems often get "stuck" in metastable states and do not change conformations for an extended period. Hence, MD simulations can be computationally costly when attempting to observe rare events, which is often the case of interest.

Fortunately, researchers have developed several "enhanced sampling methods" to overcome this timescale gap between MD simulations and biological processes. Many enhanced sampling methods work by adding a biasing potential to force the system away from metastable states. These include but are not limited to Gaussian-accelerated molecular dynamics (GaMD),[17–20] metadynamics,[21–25] umbrella sampling,[26–29] and adaptive biasing force (ABF).[30–34] Among these, GaMD has the advantage of not requiring any collective variables (CVs) to steer the simulation. Rather, it allows unrestrained sampling of configuration space. Another similar class of methods changes the system's temperature instead to sample states difficult to reach at room temperature, including replica exchange molecular dynamics (REMD) or parallel tempering.[35–39] Although both these classes of methods are effective at obtaining thermodynamic properties like the free-energy landscape of the system, they alter the actual kinetics of the system, preventing them from directly

getting kinetic properties from the system. Note that there are methods to derive kinetic properties like rate constants from simulations that used these methods. Still, they need to be obtained either by using Kramers' rate theory in the high friction or "overdamping" regime,[40] constructing a master equation,[41,42] or assuming a low residence time in the transition states,[43] which are either approximations or are limiting conditions for these methods to be used for all cases. Moreover, since the real kinetics of the system is altered, continuous pathways cannot be obtained from these methods.

As a result, several path-sampling methods focus on sampling kinetic properties, such as rate constants from the reactant state to the product state, including milestoning,[44–48] forward flux sampling,[49–52] transition interface sampling,[53–56] and others. These methods divide the path space from the reactant state to the product state into many interfaces and run many short simulations to efficiently obtain rate constants and free-energy landscape of the path space. However, since these methods primarily focus on sampling the path of interest, the rest of the free-energy landscape is not extensively sampled. If a more comprehensive picture of the entire configuration space is needed along with its thermodynamic and kinetic properties, then the weighted ensemble (WE) method[57–67] can be used on the system of interest. WE decomposes the configuration space into small volume elements called "bins" with appropriate CVs and run many short simulations with probabilities or "weights" for τ amount of time ("resampling time") for many iterations to obtain good statistics. WE has proven to be useful in uncovering insights into the mechanisms of important biophysical systems.[68–76] However, WE requires a sufficient sampling of the configuration space by having many initial configurations with close to steady-state probabilities to get accurate results quickly.

As a result, we have developed a hybrid enhanced sampling method that combines GaMD and WE called GaMD–WE. There also exist hybrid methods that combine REMD and GaMD[77,78] and well-tempered metadynamics and GaMD,[79] but both aim to improve sampling of thermodynamic properties. In contrast, GaMD–WE aims to enhance sampling of both thermodynamic and kinetic properties. There is also another hybrid method called Markovian WE milestoning (M-WEM)[80] that combines milestoning and WE and several reweighting methods for WE[81] that also aim to accelerate sampling of both thermodynamic and kinetic properties. However, M-WEM cannot generate continuous pathways since WE is used to accelerate convergence of milestones in milestoning. The reweighting methods are iterative methods that can be used additionally for any WE simulation, including GaMD–WE simulations. In GaMD–WE, GaMD is initially run to sample the free-energy landscape efficiently with its harmonic boost potentials. Then, after reweighting is performed to recover the original free-energy landscape, WE is run with many initial configurations produced from the GaMD run. This way, the two methods complement each other and reduce each other's limitations. This paper will introduce both methods, the hybrid method, and the results that demonstrate the hybrid method's power to obtain thermodynamic and kinetic properties more accurately and more quickly than one method by itself.

2. METHODS

2.1. GaMD.

GaMD is an enhanced sampling method for MD simulations that can efficiently sample thermodynamic proper ties such as the free-energy landscape of the system. When the system potential $V(\mathbf{r})$, where $\mathbf{r}$ denotes the position vector of an N-atom system, is lower than the threshold energy E, GaMD fills the energy wells by adding a harmonic boost potential $\Delta V(\mathbf{r})$, that is

$$\Delta V(\mathbf{r}) = \frac{1}{2}k(E - V(\mathbf{r}))^2 \quad (1)$$

where k denotes the harmonic force constant. If $V(\mathbf{r}) \geq E$, then no boost potential is added. There are several criteria that the boost potential $\Delta V(\mathbf{r})$ needs to satisfy for GaMD to work, and readers can refer to the original GaMD paper[17] for specific details about the boost potential $\Delta V(\mathbf{r})$ and the harmonic force constant k.

If the anharmonicity of the harmonic boost potential $\Delta V(\mathbf{r})$ is small, then $\Delta V(\mathbf{r})$ follows a near Gaussian distribution and the cumulant expansion to the second order can be used to approximate the exponential average term $\langle e^{\beta \Delta V(\mathbf{r})} \rangle$, where β denotes the thermodynamic beta or $1/k_B T$. This exponential average term $\langle e^{\beta \Delta V(\mathbf{r})} \rangle$ is needed to reweight and recover the original free-energy landscape from GaMD. Readers can refer to the original GaMD paper[17] for specific details about energetic reweighting with cumulant expansion to the second order.

A significant advantage of GaMD is that CVs, which describe the state of a molecular system, are not needed. Identifying the appropriate CVs for a particular system is still an active area of research and can be difficult for new or unfamiliar systems.[82–84] In contrast, metadynamics requires CVs to be chosen a priori, which similarly fills energy wells with repulsive Gaussian potentials. However, metadynamics does not suffer from having unconverged high-energy regions like GaMD since metadynamics recovers the original free-energy landscape as the opposite sum of all Gaussians. Nonetheless, GaMD is one of the few enhanced sampling methods that does not require tuning of many parameters and can be easily applied to various systems. Additionally, GaMD is fully implemented in Amber[85] (starting from Amber16) and NAMD[86] (starting from 2.13),[18] which makes it easier for users to use the method.

2.2. WE Method.

The WE method is another enhanced sampling method for MD simulations that runs many short simulations instead of one long simulation to sample thermodynamic and kinetic properties efficiently. These short simulations or “walkers” carry probabilities or “weights” that evolve throughout the simulation via “resampling,” a statistical procedure to maintain a number of these short simulations at visited regions of the configuration space. More details can be found in the original WE paper,[57] in a review article,[58] and in the Weighted Ensemble Simulation Toolkit with Parallelization and Analysis (WESTPA) papers,[87,88] but the general scheme is as follows.

1. The following parameters are chosen a priori for the WE simulation: CVs, resampling time τ for walkers, partitioning of bins (small volume elements of the configuration space), and target number of walkers per bin n_w. If there is only one initial state, then there will be n_w walkers, each with a weight of $1/n_w$. Otherwise, there will be multiple n_w walkers, each with an appropriate weight that sums up to 1 for the entire system.
2. Walkers are run for τ amount of time and binned to appropriate bins depending on their CV values.
3. Walkers go through "resampling," that is, merged or replicated in a statistically correct way so that the target number of walkers per bin n_w is maintained for each bin. Each walker ends up with a weight between P_i/n_w and $2P_i/n_w$ where P_i denotes the sum of the weights in bin i.
4. Steps 2 and 3 are repeated until the desired convergence is reached.

With resampling, the walkers are maintained in each visited bin regardless of its energy barrier height. Computational cost is also curtailed since walkers are merged in oversampled, low-energy regions and replicated in rare, high-energy regions. Since no statistical bias is added to the system, one can directly obtain both thermodynamic and kinetic properties of the system from the evolution of walkers' weights in each bin.

Although several parameters need to be chosen a priori as stated in step 1, the resampling time τ can be selected without having to worry about fulfilling the Markovian property, a requirement that other enhanced sampling methods have such as milestoning. The resampling time τ should be chosen to be short enough so that WE does not inadvertently miss transitions.[61,76,89] However, since many bins have to reach convergence to extract correct thermodynamic and kinetic properties of the system, WE can be computationally costly if the initial states are not close to the steady state.

In our study, equilibrium WE was used to sample the thermodynamic and kinetic properties of the system since we were interested in obtaining rate constants between more than two states. The rate constants were obtained by defining states post simulation. If one is interested in obtaining kinetic properties between two states, however, then steady-state WE can be used, which "recycles" or feeds back walkers to the initial state once they reach the target state. This way, WE focuses its computational effort in sampling the transition of interest in nonequilibrium steady-state conditions.

2.3. GaMD–WE Method.

The hybrid GaMD–WE method aims to combine strengths and mitigate weaknesses of both methods. By initially running GaMD to sample the free-energy landscape of the system, one can obtain a well-sampled initial state distribution for WE. Then with WE, one can get a more refined free-energy landscape closer to steady state and sample kinetic properties such as rate constants from one state to another state. We show that the hybrid method is significantly more effective than running a conventional WE to sample thermodynamic and kinetic properties within the same amount of simulation time in the subsequent Results section.

We have developed a GaMD–WE package for users to run GaMD and prepare initial states for WE, specifically for the WESTPA.[87,88] The current package is not fully integrated with WESTPA, that is, the user needs to use the GaMD–WE package for the GaMD portion and run a WESTPA simulation separately for the WE portion using initial states from the GaMD–WE package, but we plan to make it fully integrated in the future. The GaMD–WE package is fully customizable, that is, the desired force fields, water models, and others can be added, and it follows a series of scripts as the following.

1. System is prepared for simulation after the Protein Data Bank (PDB) structure is downloaded from the PDB server. Appropriate force field parameters are added followed by system solvation.
2. Solvated system is then minimized, heated, and equilibrated using the OpenMM[90] simulation engine. Directories are created for the subsequent GaMD simulations.
3. Six GaMD simulations are run using the Amber[85] simulation engine for varying degrees of potential boosts, that is, lower bound dihedral potential boost, upper bound dihedral potential boost, lower bound total potential boost, upper bound total potential boost, lower bound dual (dihedral + total) potential boost, and upper bound dual potential boost, which are all of the possible combinations for potential boosts, for the desired amount of simulation time.
4. Simulation data are then extracted from the output log files for each of the six GaMD simulations. Reweighting is then performed with several reweighting methods, that is, cumulant expansion to the first order, second order, and third order, to recover the original free-energy landscape. Reweighting is done for bins with more than 10 frames. Then, with the desired target number of walkers per bin n_w, initial structures for WE are saved with appropriate weights from the reweighted probabilities. If there are more frames in the chosen bin than the target number of walkers per bin, then the frames are chosen at random.
5. Initial structures for WE are minimized in two steps (Step 1: Heavy atoms C_a, N, C, and O of the protein are minimized. Step 2: Entire system including the solvent is minimized) so that none of them “crash” during WE. User can set the minimization steps.
6. WE simulation directory is created with proper initial structures. Number of initial structures can be compared among the six simulation outputs (see the Supporting Information). GaMD simulation that yielded the largest number of initial structures can be subsequently used for WE simulation since that would indicate the greatest coverage of the free-energy landscape. All of the initial structures from GaMD are used for WE. Hence, the current GaMD–WE implementation is suitable for equilibrium WE simulations that are focused on sampling the entire free-energy landscape and rate constants for regions defined post simulation.

In the subsequent Results section, we show that GaMD has greater coverage of the free-energy landscape and closer to steady-state probabilities compared to WE within the

same simulation time. Even if WE has a similar amount of coverage of the free-energy landscape compared to GaMD for some systems, GaMD has an advantage over WE with reweighting since appropriate probabilities or weights can be recovered from the added biasing potentials. In contrast, since WE does not add any statistical bias to the system, the system needs to evolve naturally or reach convergence to appropriate probabilities or weights, which in most cases would take longer than adding a biasing potential to the system.

3. RESULTS

We have tested our hybrid method on two systems: alanine dipeptide in explicit solvent and chignolin in implicit solvent. We show how GaMD–WE outperforms either method in obtaining thermodynamic and kinetic properties. To further illustrate that GaMD surpasses WE in getting the free-energy landscape, we show the free-energy landscapes of bovine pancreatic trypsin inhibitor (BPTI) in the explicit solvent obtained from the two methods. The three systems are illustrated in Figure 1a–c. Amber ff14SB force field parameters[91] were used for all three systems. Simulations were run under the canonical ensemble with the temperature T set to 300 K using the Langevin thermostat with friction coefficient $\gamma = 1.0$ ps^{-1} for all three systems.

Three simulations were run for each system (alanine dipeptide and chignolin) using GaMD–WE and WE, and the average of the three is shown as the final result. The simulations were run until the rate constants had leveled off and remained consistent. Error bars for WE and GaMD–WE rate constants represent 95% confidence intervals (i.e., $1.96 \times \frac{\sigma}{\sqrt{3}}$ where σ denotes standard deviation). Each point in the WE and GaMD–WE rate constant graphs was calculated cumulatively in 50 iteration blocks.

3.1. Alanine Dipeptide.

Alanine dipeptide is a 22-atom system that is commonly used as a test system for new methods. Initial structure was obtained from https://markovmodel.github.io/mdshare/ALA2/. TIP3P water model[92] was used to solvate the system explicitly. For GaMD–WE, GaMD was run for 50 ns and WE was run for 11.95 μs, so that the total simulation time amounted to 12 μs. The total simulation time for conventional WE was also 12 μs (three independent 4 μs runs). Resampling time τ for WE was set to 10 ps, equal to GaMD's sampling frequency. The target number of walkers per bin n_w was set to 4. CVs were set to be the dihedral angles, ϕ and ψ. Bins were evenly spaced in intervals of 10° for both ϕ and ψ (ranging from −180 to 180°).

First, we show that GaMD covers the free-energy landscape more than brute force simulation and WE within the same simulation time. Figure 2a–c shows the average free-energy landscape ($-k_B T \ln P$, where P denotes probability) of alanine dipeptide after 50 ns of brute force simulation, GaMD, and WE, respectively. The lowest energy state was set to be zero for each of the free-energy landscapes. In particular, the GaMD run with the upper bound of the dihedral boost potential yielded the maximum number of initial structures on average as compared to the other five GaMD potential settings (see the Supporting

Information). Cumulant expansion to the second order was used for GaMD reweighting. As seen in Figure 2, GaMD covered most of the metastable regions that includes P_{II}, the rest of the β-sheet region, and the right-handed α-helix region α_R on the left side compared to brute force simulation and WE within 50 ns of the simulation time. This shows that it can be more beneficial to use GaMD instead of brute force simulation or WE to sample the free-energy landscape, even for this simple system. Figure 2d,e shows that the free-energy landscapes are comparable after 12 μs of brute force simulation and WE, respectively.

Second, we show that GaMD–WE can converge to the correct rate constants faster and more accurately than WE. In particular, rate constants between the three regions of interest shown in Figure 2e were measured over simulation time. α_R region was defined to be $-120° \leq \phi \leq 0°$, $-100° \leq \psi \leq 50°$, P_{II} region was defined to be $-120° \leq \phi \leq 0°$, $100° \leq \psi \leq 180°$, and α_L region was defined to be $0° \leq \phi \leq 120°$, $-50° \leq \psi \leq 100°$. Initial structures for the three GaMD–WE runs were from one of the three GaMD runs with an upper bound of dihedral boost potential that yielded the largest number of initial structures. WE and brute force simulations used the same initial structure as the GaMD runs. In addition, since GaMD covered a wider free-energy landscape within 50 ns, another set of GaMD–WE simulations was run with equal weights. Although reweighting would give more accurate weights for each region, we wanted to investigate whether there will be any improvements in obtaining kinetics solely from covering more of the free-energy landscape with GaMD versus WE. Figure 3 shows the evolution of rate constants over aggregate simulation time, and Table 1 summarizes the final rate constants for brute force, WE, and GaMD–WE simulations after 12 μs of simulation time. Reference brute force simulation values were obtained from averaging all first passage times from three independent 4 μs runs and performing Bayesian bootstrapping for 95% confidence intervals.[88] The first 200 ns of simulation time was cut off in the rate constant calculation to eliminate the initial structure bias for brute force, WE, and GaMD–WE simulations.

Figure 3a,b shows that the convergence is comparable between WE and GaMD–WE for the rate constants between the two central metastable states α_R and P_{II} since both methods covered both regions well. However, GaMD–WE under-estimated both rate constants (see Table 1), which WE obtained more accurately. This might be due to GaMD–WE having lower weights than actual for the initial α_R and P_{II} structures, which would slow down the rate of convergence for GaMD–WE. The GaMD–WE simulations with equal weights, on the other hand, had larger error bars than regular GaMD–WE and performed similarly to GaMD–WE.

For the rate constants that involved the higher-energy region α_L, however, GaMD–WE performed better than WE. Figure 3c,e highlights that GaMD–WE having the rate constants go from α_L to either primary metastable state α_R or P_{II} converge faster with smaller error bars (see Table 1) compared to WE. GaMD–WE simulations with equal weights, on the other hand, did not perform better than either WE or GaMD–WE. As for the reverse rate constants that go from either primary metastable state α_R or P_{II} to α_L, GaMD–WE and WE have comparable performances, with GaMD–WE slightly underestimating both rate constants as seen from Figure 3d,f and Table 1. GaMD–WE simulations with equal weights

performed marginally better than GaMD–WE by underestimating the rate constants lesser. This might be due to having higher weights for the regions of interest than GaMD–WE.

These results indicate that GaMD–WE can obtain kinetics involving higher-energy regions like α_L faster and more accurately than WE alone. In contrast, WE performs as well as GaMD–WE in getting kinetics involving central metastable states, which both methods can sample sufficiently well. In addition, GaMD–WE needs reweighting to have more accurate weights and have an advantage over WE, that is, GaMD–WE does not necessarily have an advantage over WE from solely covering more of the free-energy landscape within the same simulation time. However, this might not be the case if GaMD had covered a region with a higher energy barrier that is difficult for conventional WE to sample quickly, so this hypothesis needs to be tested on more complex systems in the future.

3.2. Chignolin.

To investigate whether GaMD–WE will be significantly more effective than WE for more complex systems, we tested the two methods on chignolin, a 138-atom system with 10 residues (PDB: 1UAO). The modified generalized Born implicit model with the model II radii[93] was used to solvate the system implicitly. For GaMD–WE, GaMD was run for 500 ns, and WE was run for 39.5 μs, so that the total simulation time amounted to 40 μs. The total simulation time for conventional WE was also 40 μs (five independent 8 μs runs). Resampling time τ for WE was set to 20 ps, equal to GaMD's sampling frequency and same as in ref 88, which had a chignolin example. Target number of walkers per bin n_w was set to 4. CVs were set to be the mass-weighted root-mean-square-deviation (RMSD) of C_α atoms from the initial folded state (PDB: 1UAO) and the mass-weighted radius of gyration (R_g) of C_α atoms. Bins were evenly spaced in intervals of 0.2 Å for both RMSD and R_g (ranging from 0 to 8 Å).

Figure 4a–c shows the average free-energy landscape ($-k_B T \ln P$, where P denotes probability) of chignolin after 500 ns of brute force simulation (three out of five simulations), GaMD, and WE, respectively. The lowest energy state was set to be zero for each free-energy landscape. In particular, the GaMD run with the upper bound of the dihedral boost potential yielded the maximum number of initial structures on average as compared to the other five GaMD potential settings (see Supporting Information). Cumulant expansion to the second order was used for GaMD reweighting. Although GaMD and WE have comparable free-energy landscape coverage as seen in Figure 4b,c, GaMD has probabilities closer to actual values due to reweighting as seen in Figure 4d,e (or 4f), which show that the free-energy landscapes are comparable after 40 μs of brute force simulation and WE, respectively. In chignolin's case, however, brute force simulation had the most coverage of the free-energy landscape out of the three cases. It is also a common practice to run a short brute force simulation to sample the free-energy landscape and run WE using the sampled states. Hence, using either brute force simulation or GaMD would have been more beneficial than WE to initially sample the free-energy landscape in chignolin's case.

We also show that GaMD–WE can converge to the correct rate constants faster and more accurately than WE, more notably than in the alanine dipeptide case. Specifically, the rate constants between the four regions of interest shown in Figure 4 e,f were measured over the

simulation time. The folded region was defined to be 0.0 Å ≤ RMSD ≤ 1.0 Å, the unfolded region was defined to be 4.0 Å ≤ RMSD, the intermediate I region was defined to be 3.5 Å ≤ RMSD ≤ 4.5 Å, 4.5 Å ≤ R_g ≤ 6.5 Å, and the intermediate II region was defined to be 6.0 Å ≤ RMSD ≤ 7.0 Å, 7.0 Å ≤ R_g ≤ 8.0 Å. Initial structures for the three GaMD–WE runs were from one of the three GaMD runs with an upper bound of the dihedral boost potential that yielded the largest number of initial structures. WE and brute force simulations used the same initial structure as the GaMD runs. Figures 5 and 6 show the evolution of rate constants over aggregate simulation time, and Table 2 summarizes the final rate constants for brute force, WE, and GaMD–WE simulations after 40 μs of simulation time. The reference brute force simulation values were obtained from averaging all first passage times from five independent 8 μs runs and performing Bayesian bootstrapping for 95% confidence intervals.[88] The first 2 μs of simulation time was cut off in the rate constant calculation to eliminate the initial structure bias for brute force, WE, and GaMD–WE simulations.

Figure 5a,b shows that GaMD–WE is faster than WE at converging to the reference rate constants between the folded and unfolded regions. GaMD–WE is significantly better at obtaining the rate constants from unfolded to folded since GaMD–WE had closer to actual probabilities for the unfolded region with reweighting. Without reweighting, WE takes a significantly longer time to converge to the reference rate constants. Figure 6a–d also shows similar results that highlight GaMD–WE converging faster to the reference rate constants, especially for rate constants that go from either intermediate I or II region to the folded region. Performance for GaMD–WE is only slightly better than WE in obtaining the rate constants between the intermediate I region and the intermediate II region; However, as seen in Figure 6e,f, the two regions are close to each other, making sampling between these two regions and converging to the actual rate constants easier than the other cases. Finally, Table 2 indicates that the only rate constant that WE had within the reference confidence interval was for the folded → intermediate II rate constant. In contrast, the GaMD–WE had all of the rate constants fall within the confidence intervals except for the intermediate I → intermediate II rate constant. Error bars for GaMD−WE were lower than WE for all rate constants except for the folded → unfolded, unfolded → folded, and folded → intermediate I rate constants. These results indicate that GaMD–WE's performance in obtaining kinetics is significantly better than conventional WE for more complex systems than alanine dipeptide. Nonetheless, as previously mentioned, the brute force simulation had more coverage of the free-energy landscape than GaMD and had close to true probabilities. Hence, it is plausible that the common practice of running WE, that is, run a short brute force simulation to initially sample the free-energy landscape before running WE, would have been as good or better than GaMD–WE in obtaining the rate constants in chignolin's case.

3.3. BPTI.

As a final test, we ran a brute force simulation, GaMD, and WE of BPTI, an 892-atom system with 58 residues (PDB: 5PTI). This was to test whether GaMD will be more effective than brute force simulation and WE at covering the free-energy landscape of a bigger protein system than alanine dipeptide or chignolin. TIP4P-Ew water model[94] was used to solvate the system explicitly. The total simulation time for all three simulations was 500 ns. Resampling

time τ for WE was set to 40 ps since BPTI is a bigger system than chignolin, and the resampling time needs to be long enough to cross over to the next bin. Simulation points were sampled every 2 ps to match GaMD's sampling frequency. Target number of walkers per bins n_w was set to 4. CVs were set to be the dihedral angles χ_1 − C14 and χ_1 − C38 associated with the disulfide bond formed between cysteine 14 and cysteine 38.[95,96] Bins were evenly spaced in intervals of 10° for both (ranging from −180 to 180°).

Figure 7a–c shows the free-energy landscape ($-k_B T \ln P$, where P denotes probability) of BPTI after one 500 ns run of brute force simulation, GaMD, and WE, respectively. The lowest energy state was set to be zero for each of the free-energy landscapes. In particular, the GaMD run with the upper bound of the dual boost potential yielded the maximum number of initial structures on average as compared to the other five GaMD potential settings (see Supporting Information). Maclaurin expansion to the 10th order was used for GaMD reweighting since using cumulant expansion to the second order is limited for small proteins with 40 residues or less.[17] Figure 7b shows a free-energy landscape similar to the one obtained from accelerated molecular dynamics (aMD), which was also obtained using a dual boost potential and Maclaurin expansion to the 10th order for reweighting.[96] Slight differences between the two free-energy landscapes may be from using different force field parameters: aMD used the modified Amber ff99SB-ILDN force field,[97,98] which removed modifications to leucine, aspartic acid, and asparagine to mimic the Anton simulation of BPTI,[99] and GaMD used the Amber ff14SB force field.[91]

Metastable states of interest, including the major state M and two minor or excited states m_{C14} and m_{C38},[100] are marked in Figure 7a. M region was defined to be −120° < χ_1 − C14 < 0°, 0° < χ_1 − C38 < 120°, 0° < χ_3 < 180°, m_{C14} was defined to be 0° < χ_1 – C14 < 120°, 0° < χ_1 – C38 < 120°, − 180° < χ3 < 0°, and m_{C38} was defined to be −120° < χ_1 − C14 < 0°, −120° < χ_1 − C38 < 0°, −180° < χ_3 < 0° as in the BPTI aMD refs 95 and 96 Dihedral angle χ_3 is associated with the disulfide bond formed between cysteine 14 and cysteine 38. It is clear that GaMD is more effective at exploring other metastable states present in BPTI including m_{C14} as compared to brute force simulation or WE. However, note that the other five GaMD runs yielded similar results as brute force simulation or WE (see the Supporting Information), and only the upper bound of the dual boost potential was able to explore the various metastable states. In addition, WE is better than the brute force simulation at sampling metastable states. However, even after extending the WE simulation for 2 μs (total simulation time: 2.5 μs), WE is still not able to sample m_{C14} as seen in Figure 7d. These results are similar to those in ref 96, which highlighted the fact that a 500 ns aMD simulation of BPTI was able to sample important metastable states and as much as a 1 ms brute force simulation of BPTI from Anton. This highlights the power of GaMD being able to sample more of the configuration space than WE. GaMD can sample orthogonal modes to the chosen CVs χ_1 − C14 and χ_1 − C38 since it is a CV-free enhanced sampling method. In contrast, WE mainly samples along the chosen CVs and can encounter difficulties in the sampling regions when there are orthogonal modes present to the chosen CVs. Although rate constants between these metastable states were not measured, it is expected that GaMD–WE will sample them significantly faster than conventional WE.

4. DISCUSSION

Three examples mentioned in the previous sections highlight how GaMD–WE can be more powerful than either GaMD or WE by itself. On the other hand, this hybrid method also reveals both methods' advantages and disadvantages. GaMD is more effective at sampling the configuration space than WE by being a CV-free method. By adding boost potentials to "fill" the energy wells in a CV-free manner, GaMD can sample the configuration space more evenly across different modes in the system. On the other hand, WE is mainly limited to efficiently sampling along the chosen CVs. This is not problematic if the chosen CVs sufficiently describe the dynamics of the system, but in most cases, it is difficult to know the best CVs a priori. In such cases, WE (and other enhanced sampling methods that need CVs a priori) could miss sampling orthogonal modes and be slow at sampling the configuration space. For alanine dipeptide and chignolin, the chosen CVs have been commonly used in existing literature.[17,59,88] They are small enough systems for brute force simulation, GaMD, and WE to sufficiently sample well. However, for BPTI, the dihedral angles χ_1 − C14 and χ_1 − C38 are not commonly used CVs and have shown to be insufficient for WE to sample as effectively as GaMD. BPTI is also a much bigger system than alanine dipeptide or chignolin. In this case, principal component analysis (PCA) vectors have been commonly used instead as CVs for BPTI.[96,99,101]

However, GaMD's conventional reweighting method, which is used in GaMD–WE, is reliable for small systems up to 100 residues since the energetic noise becomes too high for accurate reweighting for larger systems.[17] A longer simulation or many simulation frames for each conformation would be necessary for larger systems to get good statistics and low error for reweighting. Hence, reweighting has been done for each structural cluster with many simulation frames, instead of individual frames, for larger systems like G-protein-coupled receptors protein complexes.[102] We plan to implement this for larger systems in our next installment of GaMD–WE.

Although WE could use PCA vectors as CVs in theory, prior brute force simulation data will be needed to accurately calculate PCA vectors, time-structure-based independent component analysis vectors, and other dimensionality reduction vectors to use them as CVs and describe the system.[103,104] Nonetheless, WE can have nondifferentiable CVs such as the number of hydrogen bonds, which can be helpful for many systems. In contrast, other methods such as metadynamics and ABF need differentiable CVs. Moreover, WE does not add any statistical bias to the system and is exact regardless of the parameters,[105] so it can reliably obtain the actual kinetics of the system. Since a long simulation is typically needed to get reasonable estimates of the kinetics, researchers have recently developed methods for WE to estimate the actual kinetics faster.[60,106,107] If GaMD–WE is combined with these current methods, more improvements will be seen in obtaining thermodynamic and kinetic properties.

5. CONCLUSIONS

We have combined two well-established enhanced sampling methods, GaMD and WE, into a hybrid method GaMD–WE to create a more powerful enhanced sampling method for MD simulations. GaMD is used to sample the free-energy landscape initially, and WE is used

to further sample the free-energy landscape and ascertain rate constants between two states of interest in the system of interest. We have shown how the hybrid method performs better than conventional WE in sampling thermodynamic and kinetic properties for two systems, and its performance significantly improves as the system size grows. We have also noted that running GaMD initially before WE will be beneficial for BPTI due to its greater coverage of the free-energy landscape. For future directions, we plan to fully integrate the hybrid method with a WE simulation toolkit such as WESTPA and possibly combine it with other WE enhancing algorithms to create state-of-the-art enhanced sampling methods. The GaMD–WE package is available at https://github.com/anandojha/gamd_we, and the documentation is available at https://gamd-we.readthedocs.io/en/latest/.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

S.H.A. acknowledges support from NIH GM31749 and the University of California San Diego. A.A.O. acknowledges support from the fellowship from the Molecular Sciences Software Institute (MolSSI) under NSF grant OAC-1547580. R.E.A. acknowledges support from NSF XSEDE CHE060063. The authors thank Lillian Chong, Daniel Zuckerman, Jeremy Copperman, and Yinglong Miao for giving helpful suggestions and feedback and Saumya Thakur for help with the supplementary cover art. All simulations were done using the Triton Shared Computing Cluster (TSCC), San Diego Supercomputing Center (SDSC).

REFERENCES

(1). Karplus M; McCammon JA Molecular dynamics simulations of biomolecules. Nat. Struct. Biol 2002, 9, 646–652. [PubMed: 12198485]

(2). Hollingsworth SA; Dror RO Molecular dynamics simulation for all. Neuron 2018, 99, 1129–1143. [PubMed: 30236283]

(3). Karplus M; Kuriyan J Molecular dynamics and protein function. Proc. Natl. Acad. Sci. U.S.A 2005, 102, 6679–6685. [PubMed: 15870208]

(4). Hospital A; Goñi JR; Orozco M; Gelpí JL Molecular dynamics simulations: advances and applications. Adv. Appl. Bioinf. Chem 2015, 8, 37.

(5). Durrant JD; McCammon JA Molecular dynamics simulations and drug discovery. BMC Biol 2011, 9, 71. [PubMed: 22035460]

(6). Dror RO; Dirks RM; Grossman JP; Xu H; Shaw DE Biomolecular simulation: a computational microscope for molecular biology. Annu. Rev. Biophys 2012, 41, 429–452. [PubMed: 22577825]

(7). De Vivo M; Masetti M; Bottegoni G; Cavalli A Role of molecular dynamics and related methods in drug discovery. J. Med. Chem 2016, 59, 4035–4061. [PubMed: 26807648]

(8). Steinhauser M; Hiermaier S A review of computational methods in materials science: examples from shock-wave and polymer physics. Int. J. Mol. Sci 2009, 10, 5135–5216. [PubMed: 20054467]

(9). Reed EJ; Soljačić M; Gee R; Joannopoulos JD Molecular dynamics simulations of coherent optical photon emission from shock waves in crystals. Phys. Rev. B: Condens. Matter Mater. Phys 2007, 75, 174302.

(10). Maginn EJ; Elliott JR Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. Ind. Eng. Chem. Res 2010, 49, 3059–3078.

(11). Xu J; Li X; Hou C; Wang L; Zhou G; Ge W; Li J Engineering molecular dynamics simulation in chemical engineering. Chem. Eng. Sci 2015, 121, 200–216.

(12). Chodera JD; Swope WC; Pitera JW; Dill KA Long-time protein folding dynamics from short-time molecular dynamics simulations. Multiscale Model. Simul 2006, 5, 1214–1226.

(13). Levitt M Protein folding by restrained energy minimization and molecular dynamics. J. Mol. Biol 1983, 170, 723–764. [PubMed: 6195346]

(14). Lindorff-Larsen K; Piana S; Dror RO; Shaw DE How fast-folding proteins fold. Science 2011, 334, 517–520. [PubMed: 22034434]

(15). Pan AC; Jacobson D; Yatsenko K; Sritharan D; Weinreich TM; Shaw DE Atomic-level characterization of protein–protein association. Proc. Natl. Acad. Sci. U.S.A 2019, 116, 4244–4249. [PubMed: 30760596]

(16). Zhu J; Lv Y; Han X; Xu D; Han W Understanding the differences of the ligand binding/unbinding pathways between phosphorylated and non-phosphorylated ARH1 using molecular dynamics simulations. Sci. Rep 2017, 7, 12439. [PubMed: 28963484]

(17). Miao Y; Feher VA; McCammon JA Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. J. Chem. Theory Comput 2015, 11, 3584–3595. [PubMed: 26300708]

(18). Pang YT; Miao Y; Wang Y; McCammon JA Gaussian accelerated molecular dynamics in NAMD. J. Chem. Theory Comput 2017, 13, 9–19. [PubMed: 28034310]

(19). Miao Y; McCammon JA Annual Reports in Computational Chemistry; Elsevier, 2017; Vol. 13; pp 231–278. [PubMed: 29720925]

(20). Wang J; Arantes PR; Bhattarai A; Hsu RV; Pawnikar S; Huang Y.-m. M.; Palermo G; Miao Y Gaussian Accelerated Molecular Dynamics: Principles and Applications. Wiley Interdiscip. Rev.: Comput. Mol. Sci 2021, 11, No. e1521. [PubMed: 34899998]

(21). Barducci A; Bonomi M; Parrinello M Metadynamics. Wiley Interdiscip. Rev.: Comput. Mol. Sci 2011, 1, 826–843.

(22). Laio A; Rodriguez-Fortea A; Gervasio FL; Ceccarelli M; Parrinello M Assessing the accuracy of metadynamics. J. Phys. Chem. B 2005, 109, 6714–6721. [PubMed: 16851755]

(23). Bussi G; Laio A; Parrinello M Equilibrium free energies from nonequilibrium metadynamics. Phys. Rev. Lett 2006, 96, 090601. [PubMed: 16606249]

(24). Barducci A; Bussi G; Parrinello M Well-tempered metadynamics: a smoothly converging and tunable free-energy method. Phys. Rev. Lett 2008, 100, 020603. [PubMed: 18232845]

(25). Laio A; Gervasio FL Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. Rep. Prog. Phys 2008, 71, 126601.

(26). Torrie GM; Valleau JP Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. J. Comput. Phys 1977, 23, 187–199.

(27). Kästner J Umbrella sampling. Wiley Interdiscip. Rev.: Comput. Mol. Sci 2011, 1, 932–942.

(28). Virnau P; Müller M Calculation of free energy through successive umbrella sampling. J. Chem. Phys 2004, 120, 10925–10930. [PubMed: 15268122]

(29). Warmflash A; Bhimalapuram P; Dinner AR Umbrella sampling for nonequilibrium processes. J. Chem. Phys 2007, 127, 154112. [PubMed: 17949137]

(30). Darve E; Rodríguez-Gómez D; Pohorille A Adaptive biasing force method for scalar and vector free energy calculations. J. Chem. Phys 2008, 128, 144120. [PubMed: 18412436]

(31). Comer J; Gumbart JC; Hénin J; Lelièvre T; Pohorille A; Chipot C The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. J. Phys. Chem. B 2015, 119, 1129–1151. [PubMed: 25247823]

(32). Darve E; Pohorille A Calculating free energies using average force. J. Chem. Phys 2001, 115, 9169–9183.

(33). Chipot C; Pohorille A Springer Series in Chemical Physics; Springer, 2007; Vol. 86; pp 159–184.

(34). Lelièvre T; Rousset M; Stoltz G Long-time convergence of an adaptive biasing force method. Nonlinearity 2008, 21, 1155.

(35). Sugita Y; Okamoto Y Replica-exchange molecular dynamics method for protein folding. Chem. Phys. Lett 1999, 314, 141–151.

(36). Sindhikara D; Meng Y; Roitberg AE Exchange frequency in replica exchange molecular dynamics. J. Chem. Phys 2008, 128, 024103. [PubMed: 18205439]

(37). Zhang W; Wu C; Duan Y Convergence of replica exchange molecular dynamics. J. Chem. Phys 2005, 123, 154105. [PubMed: 16252940]

(38). Sindhikara DJ; Emerson DJ; Roitberg AE Exchange often and properly in replica exchange molecular dynamics. J. Chem. Theory Comput 2010, 6, 2804–2808. [PubMed: 26616081]

(39). Rosta E; Hummer G Error and efficiency of replica exchange molecular dynamics simulations. J. Chem. Phys 2009, 131, 165102. [PubMed: 19894977]

(40). Miao Y Acceleration of biomolecular kinetics in Gaussian accelerated molecular dynamics. J. Chem. Phys 2018, 149, 072308. [PubMed: 30134710]

(41). Stelzl LS; Hummer G Kinetics from replica exchange molecular dynamics simulations. J. Chem. Theory Comput 2017, 13, 3927–3935. [PubMed: 28657736]

(42). Buchete N-V; Hummer G Peptide folding kinetics from replica exchange molecular dynamics. Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys 2008, 77, 030902.

(43). Tiwary P; Parrinello M From metadynamics to dynamics. Phys. Rev. Lett 2013, 111, 230602. [PubMed: 24476246]

(44). Faradjian AK; Elber R Computing time scales from reaction coordinates by milestoning. J. Chem. Phys 2004, 120, 10880–10889. [PubMed: 15268118]

(45). Vanden-Eijnden E; Venturoli M; Ciccotti G; Elber R On the assumptions underlying milestoning. J. Chem. Phys 2008, 129, 174102. [PubMed: 19045328]

(46). Bello-Rivas JM; Elber R Exact milestoning. J. Chem. Phys 2015, 142, 094102. [PubMed: 25747056]

(47). Májek P; Elber R Milestoning without a reaction coordinate. J. Chem. Theory Comput 2010, 6, 1805–1817. [PubMed: 20596240]

(48). Vanden-Eijnden E; Venturoli M Markovian milestoning with Voronoi tessellations. J. Chem. Phys 2009, 130, 194101. [PubMed: 19466815]

(49). Allen RJ; Valeriani C; Rein ten Wolde P Forward flux sampling for rare event simulations. J. Condens. Matter Phys 2009, 21, 463102.

(50). Allen RJ; Frenkel D; Ten Wolde PR Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. J. Chem. Phys 2006, 124, 194111. [PubMed: 16729807]

(51). Valeriani C; Allen RJ; Morelli MJ; Frenkel D; Rein ten Wolde P Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. J. Chem. Phys 2007, 127, 114109. [PubMed: 17887830]

(52). Hussain S; Haji-Akbari A Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook. J. Chem. Phys 2020, 152, 060901. [PubMed: 32061206]

(53). Van Erp TS; Bolhuis PG Elaborating transition interface sampling methods. J. Comput. Phys 2005, 205, 157–181.

(54). Moroni D; van Erp TS; Bolhuis PG Investigating rare events by transition interface sampling. Phys. A 2004, 340, 395–401.

(55). Borrero EE; Weinwurm M; Dellago C Optimizing transition interface sampling simulations. J. Chem. Phys 2011, 134, 244118. [PubMed: 21721623]

(56). Du W-N; Marino KA; Bolhuis PG Multiple state transition interface sampling of alanine dipeptide in explicit solvent. J. Chem. Phys 2011, 135, 145102. [PubMed: 22010733]

(57). Huber GA; Kim S Weighted-ensemble Brownian dynamics simulations for protein association reactions. Biophys. J 1996, 70, 97–110. [PubMed: 8770190]

(58). Zuckerman DM; Chong LT Weighted ensemble simulation: review of methodology, applications, and software. Annu. Rev. Biophys 2017, 46, 43–57. [PubMed: 28301772]

(59). Zhang BW; Jasnow D; Zuckerman DM The “weighted ensemble” path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. J. Chem. Phys 2010, 132, 054107. [PubMed: 20136305]

(60). Copperman J; Zuckerman DM Accelerated Estimation of Long-Timescale Kinetics from Weighted Ensemble Simulation via Non-Markovian “Microbin” Analysis. J. Chem. Theory Comput 2020, 16, 6763–6775. [PubMed: 32990438]

(61). Bhatt D; Zhang BW; Zuckerman DM Steady-state simulations using weighted ensemble path sampling. J. Chem. Phys 2010, 133, 014110. [PubMed: 20614962]

(62). Dickson A; Brooks CL III WExplore: hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. J. Phys. Chem. B 2014, 118, 3532–3542. [PubMed: 24490961]

(63). Donyapour N; Roussey NM; Dickson A REVO: Resampling of ensembles by variation optimization. J. Chem. Phys 2019, 150, 244112. [PubMed: 31255090]

(64). Ahn S-H; Grate JW; Darve EF Efficiently sampling conformations and pathways using the concurrent adaptive sampling (CAS) algorithm. J. Chem. Phys 2017, 147, 074115. [PubMed: 28830168]

(65). Abdul-Wahid B; Feng H; Rajan D; Costaouec R; Darve E; Thain D; Izaguirre JA AWE-WQ: fast-forwarding molecular dynamics using the accelerated weighted ensemble. J. Chem. Inf. Model 2014, 54, 3033–3043. [PubMed: 25207854]

(66). Yu L; Rajan D; Feng H; Darve E; Thain D; Izaguirre JA, et al. Folding Proteins at 500 ns/hour with Work Queue. 2012 IEEE 8th International Conference on E-Science, 2012; pp 1–8.

(67). Costaouec R; Feng H; Izaguirre J; Darve E Analysis of the Accelerated Weighted Ensemble Methodology; Conference Publications, 2013; p 171.

(68). Adelman JL; Grabe M Simulating current–voltage relationships for a narrow ion channel using the weighted ensemble method. J. Chem. Theory Comput 2015, 11, 1907–1918. [PubMed: 26392816]

(69). Adhikari U; Mostofian B; Copperman J; Subramanian SR; Petersen AA; Zuckerman DM Computational estimation of microsecond to second atomistic folding times. J. Am. Chem. Soc 2019, 141, 6519–6526. [PubMed: 30892023]

(70). Saglam AS; Chong LT Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. Chem. Sci 2019, 10, 2360–2372. [PubMed: 30881664]

(71). Dixon T; Lotz SD; Dickson A Predicting ligand binding affinity using on-and off-rates for the SAMPL6 SAMPLing challenge. J. Comput-Aided Mol. Des 2018, 32, 1001–1012. [PubMed: 30141102]

(72). Lotz SD; Dickson A Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing inter-actions. J. Am. Chem. Soc 2018, 140, 618–628. [PubMed: 29303257]

(73). Sztain T; Ahn S-H; Bogetti AT; Casalino L; Goldsmith JA; Seitz E; McCool RS; Kearns FL; Acosta-Reyes F; Maji S; Mashayekhi G; McCammon JA; Ourmazd A; Frank J; McLellan JS; Chong LT; Amaro RE A glycan gate controls opening of the SARS-CoV-2 spike protein. Nat. Chem 2021, 13, 963–968. [PubMed: 34413500]

(74). Ahn S-H; Grate JW; Darve EF Investigating the role of non-covalent interactions in conformation and assembly of triazine-based sequence-defined polymers. J. Chem. Phys 2018, 149, 072330. [PubMed: 30134719]

(75). Ahn S-H; Grate JW Foldamer Architectures of Triazine-Based Sequence-Defined Polymers Investigated with Molecular Dynamics Simulations and Enhanced Sampling Methods. J. Phys. Chem. B 2019, 123, 9364–9377. [PubMed: 31603686]

(76). Ahn S-H; Jagger BR; Amaro RE Ranking of ligand binding kinetics using a weighted ensemble approach and comparison with a multiscale milestoning approach. J. Chem. Inf. Model 2020, 60, 5340–5352. [PubMed: 32315175]

(77). Huang Y.-m. M.; McCammon JA; Miao Y Replica exchange Gaussian accelerated molecular dynamics: Improved enhanced sampling and free energy calculation. J. Chem. Theory Comput 2018, 14, 1853–1864. [PubMed: 29489349]

(78). Oshima H; Re S; Sugita Y Replica-exchange umbrella sampling combined with gaussian accelerated molecular dynamics for free-energy calculation of biomolecules. J. Chem. Theory Comput 2019, 15, 5199–5208. [PubMed: 31539245]

(79). Chen H; Fu H; Chipot C; Shao X; Cai W Overcoming Free-Energy Barriers with a Seamless Combination of a Biasing Force and a Collective Variable-Independent Boost Potential. J. Chem. Theory Comput 2021, 17, 3886. [PubMed: 34106706]

(80). Ray D; Stone SE; Andricioaei I Markovian Weighted Ensemble Milestoning (M-WEM): Long-time Kinetics from Short Trajectories. 2021, bioRxiv:2021.06.26.450057.

(81). Russo JD; Copperman J; Zuckerman DM Iterative trajectory reweighting for estimation of equilibrium and non-equilibrium observables. 2020, arXiv preprint arXiv:2006.09451

(82). Trapl D; Horvacanin I; Mareska V; Ozcelik F; Unal G; Spiwok V Anncolvar: approximation of complex collective variables by artificial neural networks for analysis and biasing of molecular simulations. Front. Mol. Biosci 2019, 6, 25. [PubMed: 31058167]

(83). Fiorin G; Klein ML; Hénin J Using collective variables to drive molecular dynamics simulations. Mol. Phys 2013, 111, 3345–3362.

(84). Sittel F; Stock G Perspective: Identification of collective variables and metastable states of protein dynamics. J. Chem. Phys 2018, 149, 150901. [PubMed: 30342445]

(85). Case DA; Belfon K; Ben-Shalom I; Brozell SR; Cerutti D; Cheatham T; Cruzeiro VWD; Darden T; Duke RE; Giambasu G; et al. Amber 2020, 2020.

(86). Phillips JC; Hardy DJ; Maia JDC; Stone JE; Ribeiro JV; Bernardi RC; Buch R; Fiorin G; Hénin J; Jiang W; et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys 2020, 153, 044130. [PubMed: 32752662]

(87). Zwier MC; Adelman JL; Kaus JW; Pratt AJ; Wong KF; Rego NB; Suárez E; Lettieri S; Wang DW; Grabe M; et al. WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. J. Chem. Theory Comput 2015, 11, 800–809. [PubMed: 26392815]

(88). Bogetti AT; Mostofian B; Dickson A; Pratt A; Saglam AS; Harrison PO; Adelman JL; Dudek M; Torrillo PA; DeGrave AJ; et al. A suite of tutorials for the WESTPA rare-events sampling software [Article v1. 0]. Living J. Comp. Mol. Sci 2019, 1, 10607.

(89). Pratt A; Suárez E; Zuckerman DM; Chong LT Extensive Evaluation of Weighted Ensemble Strategies for Calculating Rate Constants and Binding Affinities of Molecular Association/ Dissociation Processes. 2019, bioRxiv:671172

(90). Eastman P; Swails J; Chodera JD; McGibbon RT; Zhao Y; Beauchamp KA; Wang L-P; Simmonett AC; Harrigan MP; Stern CD; et al. OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. PLoS Comput. Biol 2017, 13, No. e1005659. [PubMed: 28746339]

(91). Maier JA; Martinez C; Kasavajhala K; Wickstrom L; Hauser KE; Simmerling C ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. J. Chem. Theory Comput 2015, 11, 3696–3713. [PubMed: 26574453]

(92). MacKerell AD Jr; Bashford D; Bellott M; Dunbrack RL Jr; Evanseck JD; Field MJ; Fischer S; Gao J; Guo H; Ha S; et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J. Phys. Chem. B 1998, 102, 3586–3616. [PubMed: 24889800]

(93). Onufriev A; Bashford D; Case DA Modification of the generalized Born model suitable for macromolecules. J. Phys. Chem. B 2000, 104, 3712–3720.

(94). Horn HW; Swope WC; Pitera JW; Madura JD; Dick TJ; Hura GL; Head-Gordon T Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew. J. Chem. Phys. 2004, 120, 9665–9678. [PubMed: 15267980]

(95). Xue Y; Ward JM; Yuwen T; Podkorytov IS; Skrynnikov NR Microsecond time-scale conformational exchange in proteins: using long molecular dynamics trajectory to simulate NMR relaxation dispersion data. J. Am. Chem. Soc 2012, 134, 2555–2562. [PubMed: 22206299]

(96). Pierce LCT; Salomon-Ferrer R; Augusto F. de Oliveira C; McCammon JA; Walker RC Routine access to millisecond time scale events with accelerated molecular dynamics. J. Chem. Theory Comput 2012, 8, 2997–3002. [PubMed: 22984356]

(97). Hornak V; Abel R; Okur A; Strockbine B; Roitberg A; Simmerling C Comparison of multiple Amber force fields and development of improved protein backbone parameters. Proteins: Struct., Funct., Bioinf 2006, 65, 712–725.

(98). Lindorff-Larsen K; Piana S; Palmo K; Maragakis P; Klepeis JL; Dror RO; Shaw DE Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins: Struct., Funct., Bioinf 2010, 78, 1950–1958.

(99). Shaw DE; Maragakis P; Lindorff-Larsen K; Piana S; Dror RO; Eastwood MP; Bank JA; Jumper JM; Salmon JK; Shan Y; et al. Atomic-level characterization of the structural dynamics of proteins. Science 2010, 330, 341–346. [PubMed: 20947758]

(100). Grey MJ; Wang C; Palmer AG Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by CPMG relaxation dispersion and chemical shift modeling. J. Am. Chem. Soc 2003, 125, 14324–14335. [PubMed: 14624581]

(101). Ernst M; Sittel F; Stock G Contact-and distance-based principal component analysis of protein dynamics. J. Chem. Phys 2015, 143, 244114. [PubMed: 26723658]

(102). Wang J; Miao Y Mechanistic insights into specific G protein interactions with adenosine receptors. J. Phys. Chem. B 2019, 123, 6462–6473. [PubMed: 31283874]

(103). Pérez-Hernández G; Paul F; Giorgino T; De Fabritiis G; Noé F Identification of slow molecular order parameters for Markov model construction. J. Chem. Phys 2013, 139, 015102. [PubMed: 23822324]

(104). Schwantes CR; Pande VS Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. J. Chem. Theory Comput 2013, 9, 2000–2009. [PubMed: 23750122]

(105). Aristoff D; Zuckerman DM Optimizing weighted ensemble sampling of steady states. Multiscale Model. Simul 2020, 18, 646–673. [PubMed: 34421402]

(106). DeGrave AJ; Bogetti AT; Chong LT The RED scheme: Rate-constant estimation from pre-steady state weighted ensemble simulations. J. Chem. Phys 2021, 154, 114111. [PubMed: 33752378]

(107). Torrillo PA; Bogetti AT; Chong LT A minimal, adaptive binning scheme for weighted ensemble simulations. J. Phys. Chem. A 2021, 125, 1642–1649. [PubMed: 33577732]

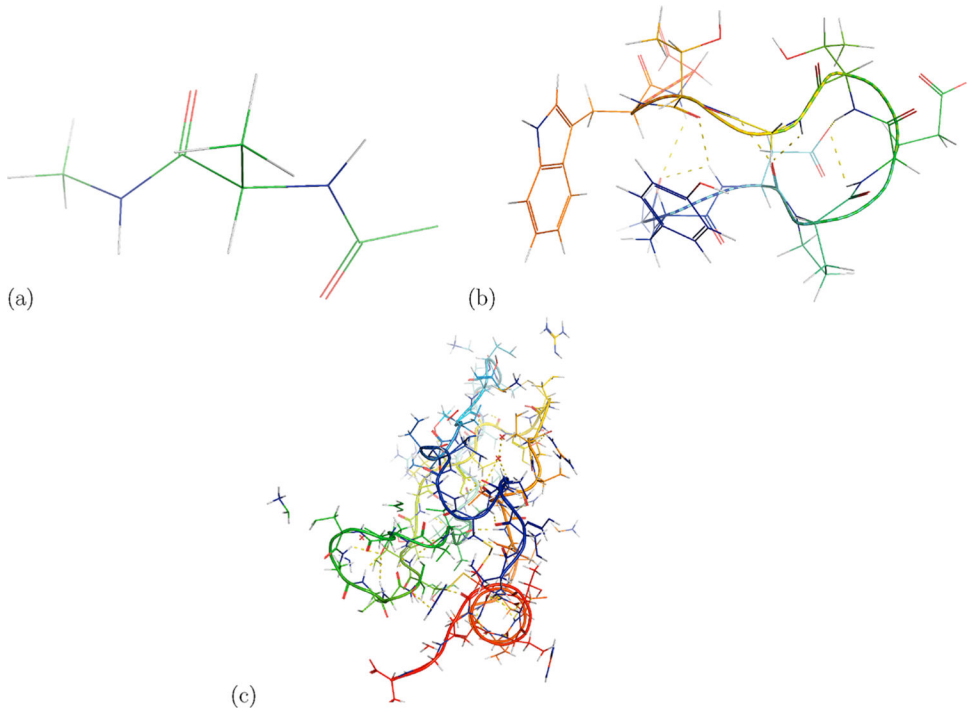


Figure 1.
Representative pictures of the systems tested: (a) alanine dipeptide, (b) chignolin, and (c) BPTI.

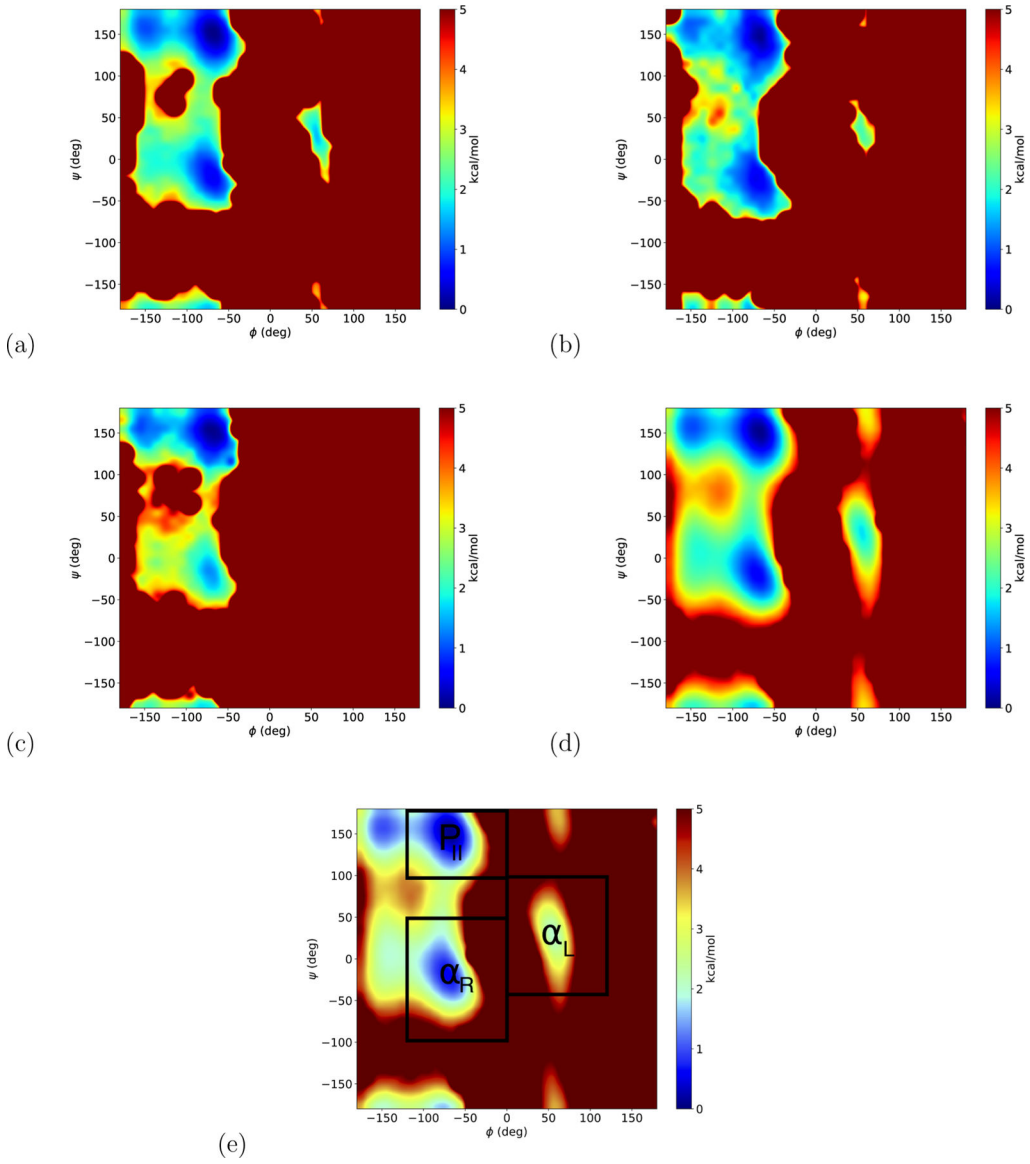


Figure 2.

Average free-energy landscape ($-k_B T \ln P$, where P denotes probability) of alanine dipeptide after (a) 50 ns of brute force simulations, (b) 50 ns of GaMD (with the upper bound of the dihedral boost potential), and (c) 50 ns of WE. (d,e) show the average free-energy landscape obtained after 12 μs of brute force simulation and WE (after cutting out the first 200 ns of simulation time to eliminate the initial structure bias), respectively. (e) shows the regions of interest (α_R, α_L, and P_{II}) marked. The lowest energy state was set to be zero for each free-energy landscape.

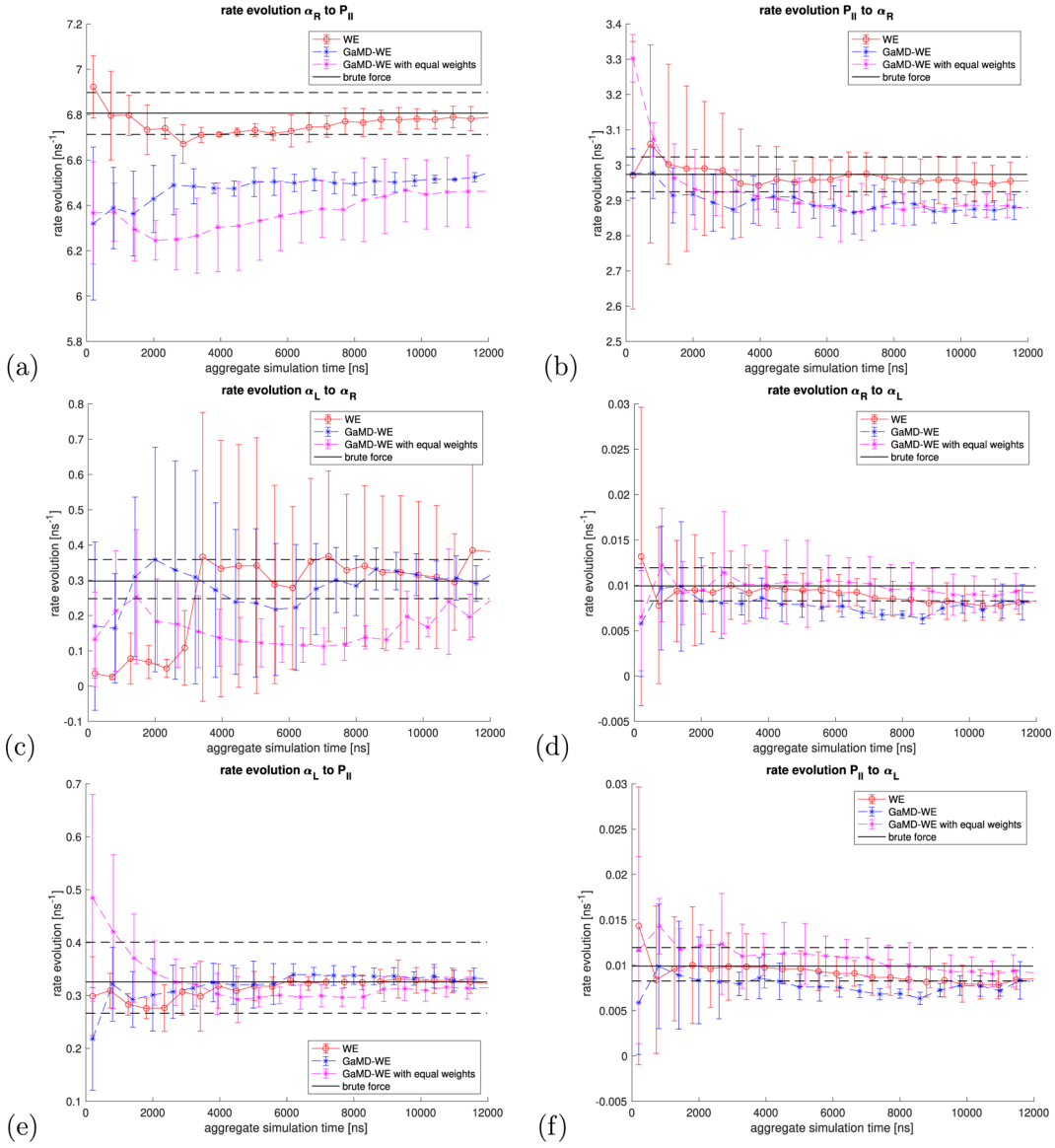


Figure 3.

Evolution of rate constants over aggregate simulation time for WE (in red), GaMD–WE (in blue), and GaMD–WE with equal weights (in magenta). The reference brute force values are in black. (a,b) show the rate constants between the two major metastable states α_R and P_{II}. (c,d) show the rate constants between α_R and α_L, a higher-energy region, and (e,f) show the rate constants between P_{II} and α_L.

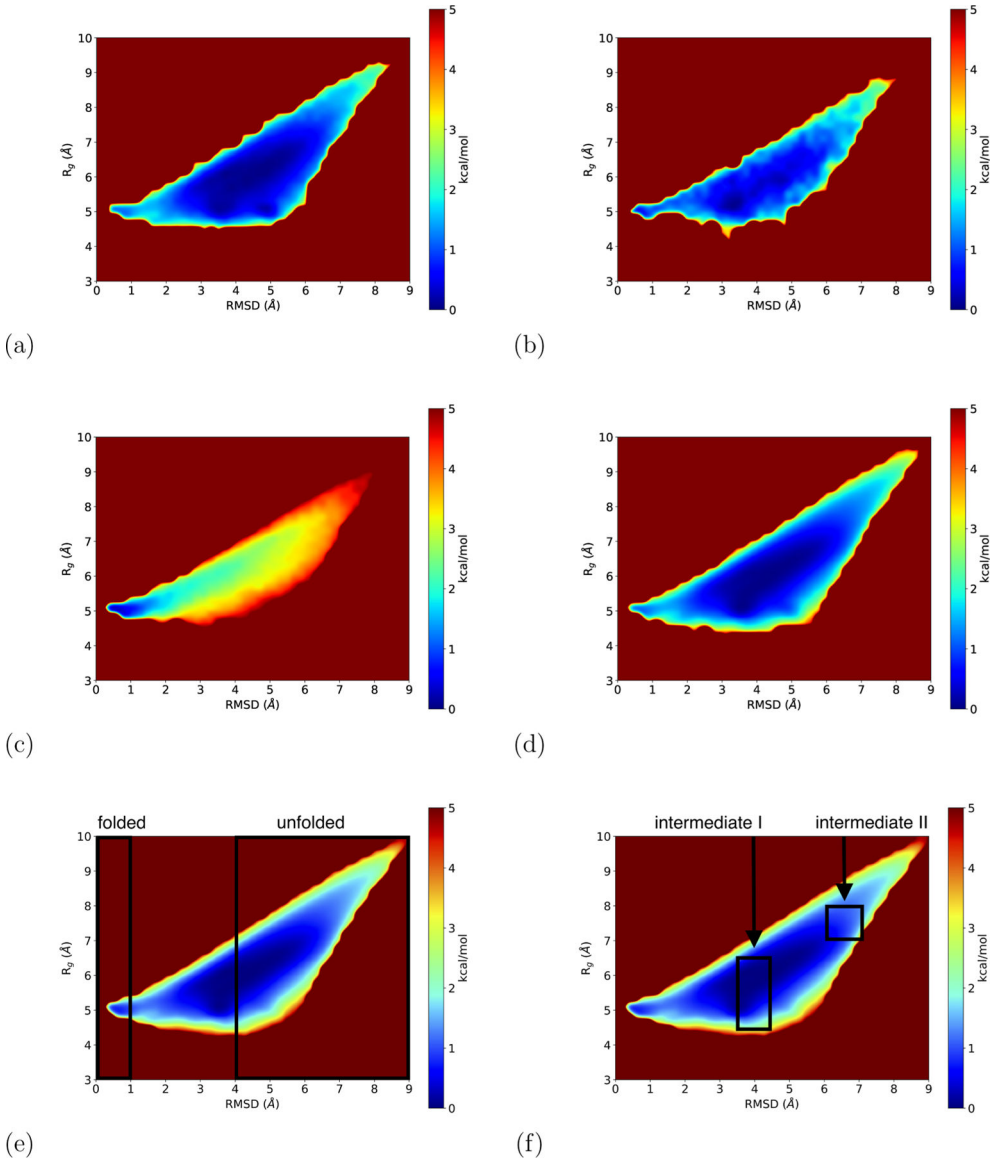


Figure 4.

Average free-energy landscape ($-k_B T \ln P$, where P denotes probability) of chignolin after (a) 500 ns of brute force simulation, (b) 500 ns of GaMD (with the upper bound of the dihedral boost potential), and (c) 500 ns of WE (d,e) (or (f)) show the average free-energy landscape obtained after 40 μs of brute force simulation and WE (after cutting out the first 2 μs of simulation time to eliminate the initial structure bias), respectively. (e,f) show the regions of interest marked. The lowest energy state was set to be zero for each free-energy landscape.

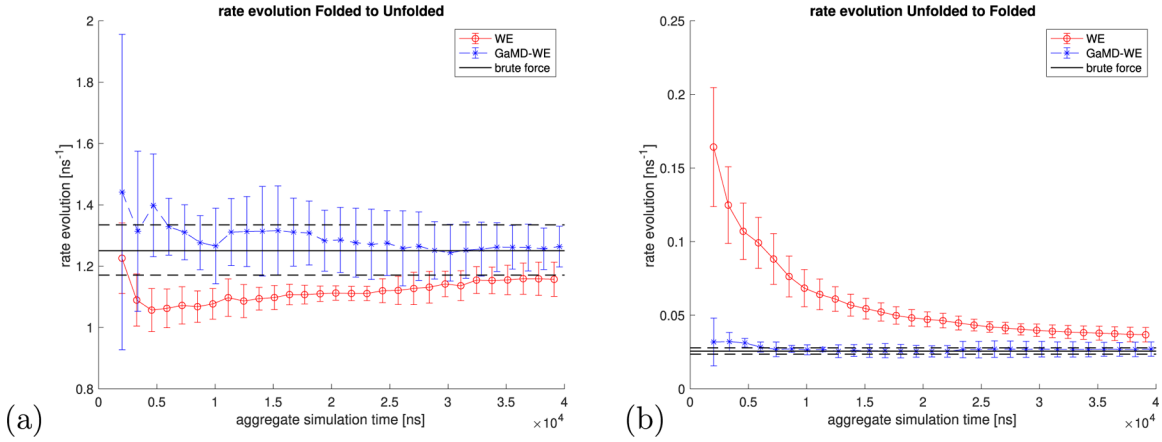


Figure 5.

Evolution of rate constants over aggregate simulation time for WE (in red) and GaMD–WE (in blue). The reference brute force values are in black. (a,b) show the rate constants between the folded region and the unfolded region.

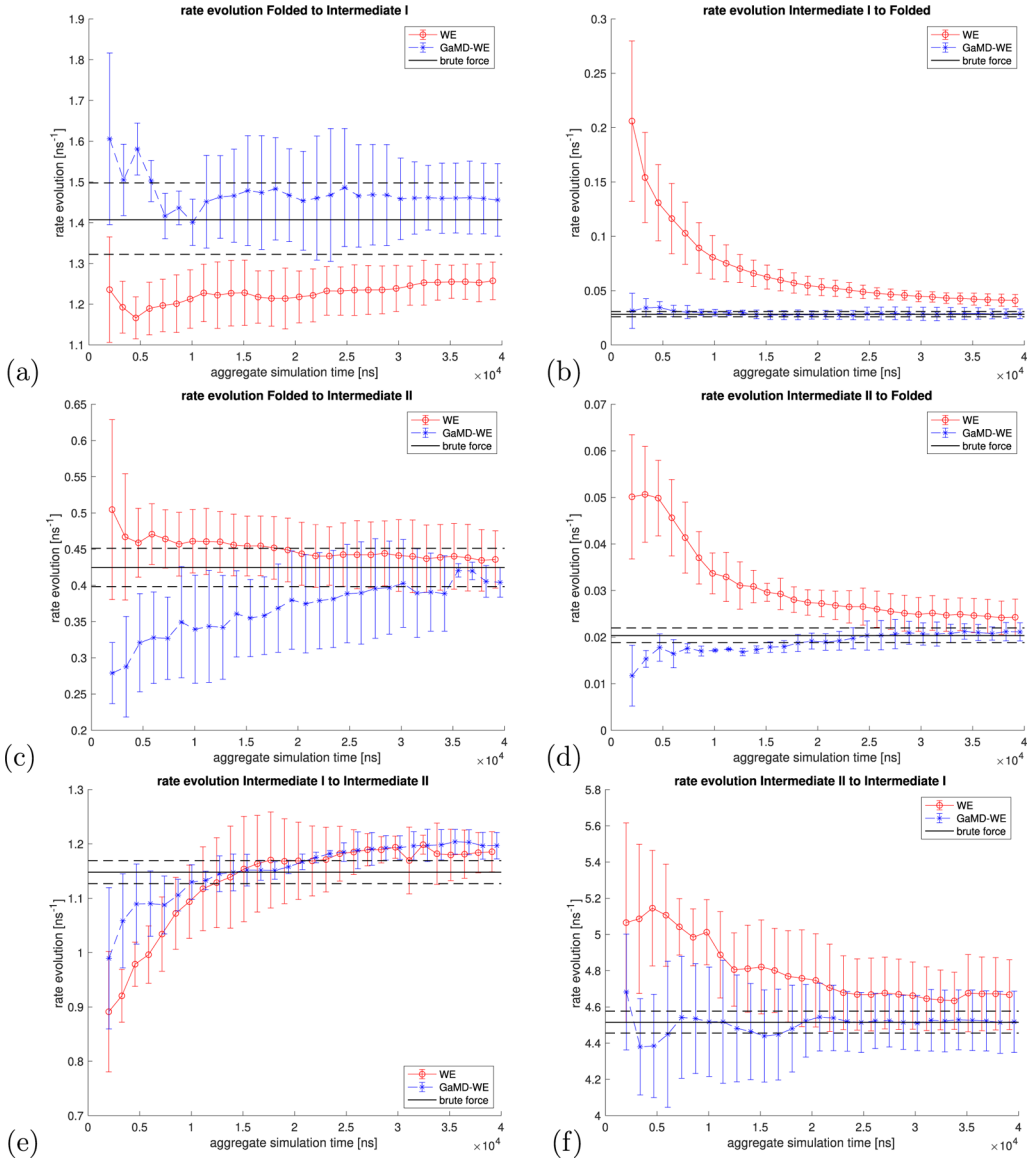


Figure 6.

Evolution of rate constants over aggregate simulation time for WE (in red) and GaMD–WE (in blue). The reference brute force values are in black. (a,b) show the rate constants between the folded region and the intermediate I region. (c,d) show the rate constants between the folded region and the intermediate II region. (e,f) show the rate constants between the intermediate I region and the intermediate II region.

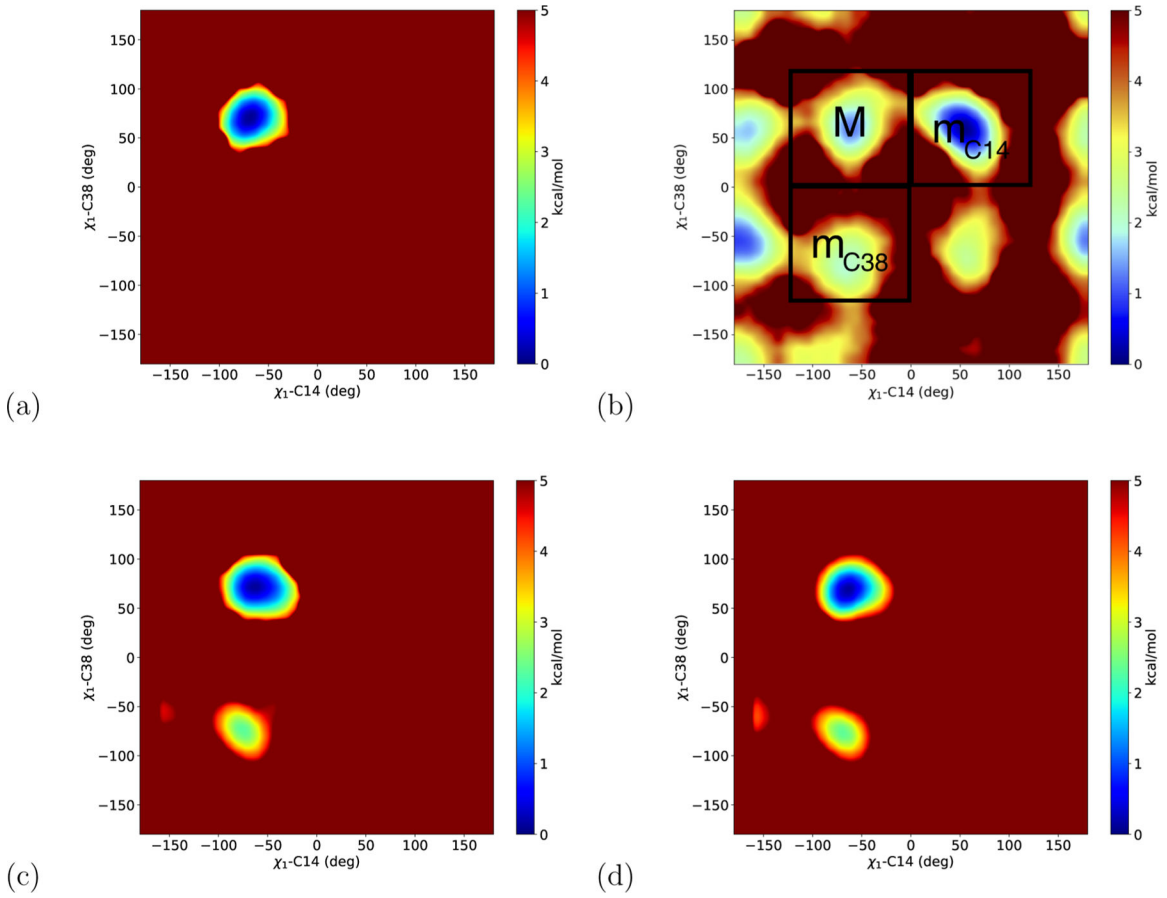


Figure 7.

Free-energy landscape ($-k_B T \ln P$, where P denotes probability) of BPTI after (a) one 500 ns brute force simulation run, (b) one 500 ns run of GaMD (with the upper bound of dual boost potential) with metastable states of interest M, m_{C14}, and m_{C38} marked, (c) one 500 ns run of WE, and (d) extending WE run for 2 μs (total simulation time: 2.5 μs), respectively. The lowest energy state was set to be zero for each of the free-energy landscapes.

Table 1.

Alanine Dipeptide Rate Constants (in ns^{-1}) after 12 μs of Simulation Time[a]

	brute force	WE	GaMD–WE	GaMD–WE with equal weights
αR → PII	6.81, [6.71, 6.90]	6.79 ± 0.052	6.54 ± 0.088	6.47 ± 0.12
PII → αR	2.97, [2.92, 3.02]	2.96 ± 0.055	2.88 ± 0.035	2.88 ± 0.040
αL → αR	0.30, [0.25, 0.36]	0.38 ± 0.25	0.34 ± 0.087	0.23 ± 0.14
α_R → α_L	0.0099, [0.0083, 0.012]	0.0083 ± 0.0011	0.0081 ± 0.0021	0.0092 ± 0.0017
α_L → P_{II}	0.33, [0.27, 0.40]	0.32 ± 0.020	0.33 ± 0.012	0.31 ± 0.019
P_{II} → α_L	0.0099, [0.0083, 0.012]	0.0085 ± 0.0014	0.0082 ± 0.0020	0.0090 ± 0.0018

[a] In the brute force simulation column, the first value indicates the average rate constant value, and the second value indicates the 95% confidence interval calculated from Bayesian bootstrapping. For WE and GaMD–WE, the error bars represent 95% confidence intervals calculated from the standard deviation of three independent runs.

Table 2.

Chignolin Rate Constants after 40 μs of Simulation Time[a]

	brute force [ns−1]	WE [ns−1]	GaMD–WE [ns−1]
folded → unfolded	1.25, [1.17, 1.33]	1.16 ± 0.056	1.26 ± 0.067
unfolded → folded	0.026, [0.024, 0.028]	0.037 ± 0.0049	0.027 ± 0.0049
folded → intermediate I	1.41, [1.32, 1.50]	1.26 ± 0.046	1.46 ± 0.089
intermediate I → folded	0.028, [0.026, 0.031]	0.041 ± 0.0056	0.029 ± 0.0045
folded → intermediate II	0.42, [0.40, 0.45]	0.44 ± 0.040	0.40 ± 0.020
intermediate II → folded	0.020, [0.019, 0.022]	0.024 ± 0.0039	0.021 ± 0.0019
intermediate I → intermediate II	1.15, [1.13, 1.17]	1.19 ± 0.037	1.20 ± 0.024
intermediate II → intermediate I	4.52, [4.46, 4.58]	4.67 ± 0.19	4.52 ± 0.17

[a]In the brute force simulation column, the first value indicates the average rate constant value, and the second value indicates the 95% confidence interval calculated from Bayesian bootstrapping. For WE and GaMD–WE, the error bars represent 95% confidence intervals calculated from the standard deviation of three independent runs.