Title

Hybrid Gaussian Accelerated Molecular Dynamics and the Weighted Ensemble Methods for Biomolecular Simulations

Authors

Thyagatur, Anugraha

Wan, Hung-Yu

Sonti, Siddharth

et al.

Publication Date

2026-07-31

DOI

10.1002/9781394316632.ch13

Chapter 1

Hybrid Gaussian accelerated molecular dynamics and the weighted ensemble methods for biomolecular simulations

Anugraha Thyagatur,[2] Hung-Yu Wan,[1] Siddharth Sonti,[1] Roland Faller,[1,3] and Surl-Hee Ahn[1*]

[1] *Department of Chemical Engineering, University of California, Davis, 95616, California, Davis, 1 Shields Ave, USA*
[2] *Department of Materials Science and Engineering, University of California, Davis, 95616, California, Davis, 1 Shields Ave., USA*
[3] *Edward E. Whitacre Jr. College of Engineering, Texas Tech University, 79409, Texas, Lubbock, 2500 Broadway W, USA*

*Corresponding Author: Surl-Hee Ahn; sahn@ucdavis.edu

G:aussian accelerated molecular dynamics (GaMD) enhances sampling without predefined collective variables (CVs), but it scales poorly on GPUs. The weighted-ensemble (WE) method is statistically unbiased and parallelizable, yet it requires CVs and can converge slowly from inadequate initial states. To mitigate these bottlenecks, we developed two hybrid methods: (1) GaMD–WE and (2) parallelizable GaMD (ParGaMD). GaMD–WE uses short GaMD simulations to generate diverse, well-sampled starting configurations with reweighted probabilities for subsequent WE

calculations, improving efficient estimation of both thermodynamic and kinetic properties. ParGaMD integrates GaMD dynamics within the WE framework: many short GaMD trajectories act as WE walkers in parallel, which improves computational scalability and accelerates convergence of the free energy landscape. We discuss both hybrids in detail and demonstrate their superiority in sampling compared to standalone GaMD and WE on complex biomolecular systems.

Keywords: Molecular dynamics, Gaussian accelerated molecular dynamics, weighted ensemble method, enhanced sampling methods

1.1. Introduction

Molecular dynamics (MD) simulations have become indispensable tools in many fields, including biology [Karplus and McCammon, 2002, Karplus and Kuriyan, 2005, Hollingsworth and Dror, 2018], chemistry [Durrant and McCammon, 2011, Dror et al., 2012, Hospital et al., 2015, De Vivo et al., 2016], materials science [Reed et al., 2007, Steinhauser and Hiermaier, 2009], and chemical and biological engineering [Maginn and Elliott, 2010, Xu et al., 2015]. A common application of MD simulations is to elucidate the mechanisms of a biological system of interest, providing insights with atomistic detail, e.g., to study protein folding [Levitt, 1983, Karplus and Kuriyan, 2005, Chodera et al., 2006, Lindorff-Larsen et al., 2011, Dror et al., 2012], to study protein-protein or protein-ligand interactions [Zhu et al., 2017, Pan et al., 2019], and for computer-aided drug design (virtual screening and ligand docking) [Durrant and McCammon, 2011, Hospital et al., 2015, De Vivo et al., 2016]. A major limitation of MD is the necessity of using femtosecond time steps, a constraint imposed by the fastest motions in the system, such as atomic bond vibrations.

This creates a significant challenge, as the biological processes being studied often unfold on timescales of microseconds or longer. Furthermore, a common challenge in routine MD simulations is kinetic trapping in metastable states. This results in insufficient conformational sampling, making it computationally costly to observe rare events (often the primary scientific goal).

To address the timescale gap between MD simulations and biological processes, various "enhanced sampling methods" have been developed. A common strategy is to apply a biasing potential that lowers energy barriers, thereby promoting transitions between metastable states. These include Gaussian accelerated dynamics (GaMD) [Miao et al., 2015, Pang et al., 2017, Miao and McCammon, 2017, Wang et al., 2021], metadynamics [Laio et al., 2005, Bussi et al., 2006, Barducci et al., 2008, Laio and Gervasio, 2008, Barducci et al., 2011], umbrella sampling [Torrie and Valleau, 1977, Virnau and Müller, 2004, Warmflash et al., 2007, Kästner, 2011], and adaptive biasing force (ABF) [Darve and Pohorille, 2001, Chipot and Pohorille, 2007, Darve et al., 2008, Lelièvre et al., 2008, Comer et al., 2015]. Among these methods, GaMD is a notable method because it operates without predefined collective variables (CVs) to guide the simulation, which permits a broad sampling of the configuration space. Other enhanced sampling techniques, such as replica exchange molecular dynamics (REMD or parallel tempering), employ a different approach by running multiple replicas of the system in parallel at various temperatures. This strategy helps to sample states that are difficult to access at the desired target temperature [Sugita and Okamoto, 1999, Zhang et al., 2005, Sindhikara et al., 2008, Rosta and Hummer, 2009, Sindhikara et al., 2010, Paschek and García, 2004, Bunker and Dünweg, 2000, Faller et al., 2002]. Although both these classes of methods are effective at

obtaining thermodynamic properties (e.g., free energy landscape), they alter the actual kinetics of the system, thus precluding the direct extraction of kinetic parameters from simulations employing these techniques. Nevertheless, kinetic properties, including rate constants, can be indirectly derived using theoretical frameworks such as Kramers' Rate Theory in the high friction regime [Miao, 2018], the formulation of master equations [Buchete and Hummer, 2008, Stelzl and Hummer, 2017], or assumptions involving short residence times within transition states [Tiwary and Parrinello, 2013]. However, these indirect methods often rely on specific approximations or conditions that may not be universally applicable to all systems or scenarios.

Consequently, several path-sampling methods have been developed to sample kinetic properties specifically. These include techniques such as milestoning, [Faradjian and Elber, 2004, Vanden-Eijnden et al., 2008, Vanden-Eijnden and Venturoli, 2009, Májek and Elber, 2010, Bello-Rivas and Elber, 2015], forward flux sampling (FFS) [Allen et al., 2006, Valeriani et al., 2007, Allen et al., 2009, Hussain and Haji-Akbari, 2020], transition interface sampling (TIS) [Moroni et al., 2004, Van Erp and Bolhuis, 2005, Borrero et al., 2011, Du et al., 2011], among others. These path-sampling methods work by dividing the pathway from a reactant to a product state into a series of interfaces. By running many short simulations between these interfaces, they can efficiently determine rate constants and the free energy profile for that specific path.

However, a key limitation is their narrow focus; since they concentrate on the pathway of interest, the rest of the conformational space remains largely unexplored. When a more complete picture of the entire system is required, including both its thermodynamic and kinetic properties, the

weighted ensemble method (WE) is a suitable approach. [Huber and Kim, 1996, Zhang et al., 2010, Bhatt et al., 2010, Yu et al., 2012, Costaouec et al., 2013, Abdul-Wahid et al., 2014, Dickson and Brooks III, 2014, Zuckerman and Chong, 2017, Ahn et al., 2017, Donyapour et al., 2019, Copperman and Zuckerman, 2020]. WE divides the configuration space into discrete regions called "bins" using collective variables (CVs). Numerous short simulations, each assigned a statistical weight, are then run for a fixed duration, for a τ amount of time ("resampling time"). This cycle is repeated over many iterations to achieve robust statistical sampling. WE has been valuable for providing mechanistic insights into important biophysical systems [Adelman and Grabe, 2015, Dixon et al., 2018, Lotz and Dickson, 2018, Ahn et al., 2018, Adhikari et al., 2019, Saglam and Chong, 2019, Ahn and Grate, 2019, Ahn et al., 2020, Sztain et al., 2021]. However, for WE to yield accurate results quickly, it must begin with a well-sampled set of initial configurations that closely approximate the system's true probability distribution.

To address the limitation of WE's dependence on well-distributed initial configurations, we developed a hybrid approach called GaMD-WE, which combines the GaMD and WE methods. While other hybrid strategies have been reported, such as those combining REMD with GaMD [Huang et al., 2018, Oshima et al., 2019] or well-tempered metadynamics with GaMD [Chen et al., 2021], these methods focus mainly on enhancing the sampling of thermodynamic properties. GaMD-WE, in contrast, is designed to improve the sampling efficiency of both thermodynamic and kinetic properties. Other methodologies also aim to accelerate the comprehensive sampling of both these property types. These include the Markovian weighted ensemble milestoning (M-WEM) method [Ray et al., 2021], which pairs milestoning with WE,

and various reweighting techniques developed for WE [Russo et al., 2020]. However, M-WEM cannot generate continuous pathways, as its WE component primarily serves to accelerate milestone convergence within the milestoning framework rather than to preserve overall path continuity.

In the GaMD-WE method, GaMD is first used to broadly explore the free energy landscape by applying its harmonic boost potentials. Following this stage, a reweighting procedure is performed to recover the original, unbiased free energy landscape. Subsequently, the WE simulations are conducted, utilizing numerous initial configurations derived from the GaMD simulation, to sample both thermodynamic and kinetic properties effectively. This integrated approach allows the two methodologies to complement each other and mitigate their respective limitations (GaMD's limitation being that the real kinetics of the system are altered).

While GaMD efficiently samples the system's configuration space with its harmonic boost potential, specific MD simulation engines in which GaMD is implemented, such as the Assisted Model Building with Energy Refinement (AMBER) [Case et al., 2020] and OpenMM [Eastman et al., 2017], are optimized for single GPU execution and exhibit suboptimal performance scaling with multiple GPUs. To address this scaling limitation, we have developed parallelizable GaMD (ParGaMD), a distinct hybrid enhanced sampling method that integrates GaMD with the WE framework using a strategy different from that of GaMD-WE. The WE component of the ParGaMD framework facilitates the parallel execution of numerous short GaMD trajectories. This approach achieves a marked reduction in wall-clock time, and by consolidating the sampling data from these runs, it allows for the effective exploration of a wide range of conformations. Based on

this, the ParGaMD framework accelerates the sampling of rare transitions across significant energy barriers through a synergy between the global energy landscape flattening of GaMD and the directed exploration of WE along predefined collective variables (CV). The ParGaMD method reconstructs the free energy landscape by applying a rigorous reweighting algorithm that integrates weights from both its WE and GaMD frameworks.

In this chapter, we will provide the theoretical foundations of the GaMD and WE methods. Subsequently, we will discuss the implementation of the hybrid GaMD-WE and ParGaMD approaches (Figure 1.1). Finally, we will demonstrate the applications of these hybrid methods, showing that they can yield thermodynamic and kinetic properties with computational efficiency compared to those of the standalone GaMD or WE approaches.

1.2. Methods

1.2.1. Gaussian accelerated molecular dynamics (GaMD)

GaMD is designed to facilitate the exploration of the configuration space of biomolecular systems, which often face substantial potential energy barriers in transitioning from one configuration to another [Miao et al., 2015]. Such energy barriers can trap systems in local minima, restricting the comprehensive sampling of diverse molecular conformations. GaMD addresses this limitation by adding a harmonic boost potential to smooth the system's potential energy surface. This boost is applied when the system's instantaneous potential energy

falls below a predefined threshold, smoothing the modified potential energy landscape and lowering the effective energy barriers. Specifically, for a system of N atoms with a potential energy $V(\mathbf{r})$ at a position $\mathbf{r} = \{\vec{r}_1, \vec{r}_2, \vec{r}_3, \ldots, \vec{r}_N\}$, when $V(\mathbf{r}) < E$ (a threshold energy), a boost potential $\Delta V(\mathbf{r})$ is added:

$$\Delta V(\mathbf{r}) = \frac{k}{2}(E - V(\mathbf{r}))^2 \tag{1.1}$$

where k is the harmonic force constant. The boost potential can be added to the total potential energy, dihedral potential energy, or both (referred to as "dual"). For further methodological details, readers are referred to the original GaMD publication [Miao et al., 2015].

The ligand Gaussian accelerated molecular dynamics (LiGaMD) is a specialized extension of GaMD that is designed to enhance the sampling of protein-ligand interactions [Miao et al., 2020, Wang and Miao, 2023, 2024]. LiGaMD employs a harmonic boost potential similar to that of GaMD but targets the ligand and its immediate interactions with the protein or receptor more precisely. Specifically, for a system with N atoms at a position $\mathbf{r}$, the potential energy $V(\mathbf{r})$ can be expressed as:

$$\begin{aligned} V(\mathbf{r}) = {} & V_{P,b}(\mathbf{r}_P) + V_{L,b}(\mathbf{r}_L) + V_{E,b}(\mathbf{r}_E) \\ & + V_{PP,nb}(\mathbf{r}_P) + V_{LL,nb}(\mathbf{r}_L) + V_{EE,nb}(\mathbf{r}_E) \\ & + V_{PL,nb}(\mathbf{r}_{PL}) + V_{PE,nb}(\mathbf{r}_{PE}) + V_{LE,nb}(\mathbf{r}_{LE}) \end{aligned} \tag{1.2}$$

where $V_{P,b}$, $V_{L,b}$, and $V_{E,b}$ are the bonded potential energies for the protein (P), ligand (L), and environment (E), respectively. $V_{PP,nb}$, $V_{LL,nb}$, and $V_{EE,nb}$ represent the non-bonded potential energies within the protein, ligand, and environment, respectively, while $V_{PL,nb}$, $V_{PE,nb}$, and $V_{LE,nb}$ represent the non-bonded potential energies for interactions between the protein and the

ligand, the protein and its surrounding environment, and the ligand and its surrounding environment, respectively.

The boost potential in LiGaMD is applied to the ligand non-bonded interactions $V_{L,nb}$, which include $V_{LL,nb}$, $V_{PL,nb}$, and $V_{LE,nb}$ (i.e., electrostatic and van der Waals interactions). Similar to GaMD, when the ligand non-bonded potential energy $V_{L,nb}(\mathbf{r})$ is below a threshold energy $E_{L,nb}$, a boost potential $\Delta V_{L,nb}$ is added:

$$\Delta V_{L,nb}(\mathbf{r}) = \frac{1}{2} k_{L,nb}(E_{L,nb} - V_{L,nb}(\mathbf{r}))^2 \tag{1.3}$$

where $k_{L,nb}$ is the harmonic force constant. Another boost potential ΔV_D, where D stands for "dual," can be added in LiGaMD using the system's total potential energy other than $V_{L,nb}$ when V_D is below a threshold energy E_D:

$$\Delta V_D(\mathbf{r}) = \frac{1}{2} k_D(E_D - V_D(\mathbf{r}))^2 \tag{1.4}$$

where k_D is the harmonic force constant. Readers can refer to the original LiGaMD paper [Miao et al., 2020, Wang and Miao, 2023, 2024] for more specific details. Recently, several extensions of LiGaMD have been developed. In LiGaMD2 [Wang and Miao, 2023], boost potentials are applied to both the ligand and the protein within the binding pocket. In LiGaMD3 [Wang and Miao, 2024], three individual boost potentials are applied to enhance ligand dissociation, rebinding, and accompanying broader system conformational changes. By applying these boost potentials, LiGaMD enhances the exploration of the configuration space associated with ligand binding and dissociation events, capturing transient and rare intermediate states that are critical for understanding binding mechanisms.

After completing GaMD and/or LiGaMD simulations, an unbiased free energy profile for the process of interest is obtained from the simulation data through a reweighting procedure. This procedure involves using the boost potential $\Delta V(\mathbf{r})$ for each frame in the MD simulation trajectory and the modified probability distribution $P^*(A)$ along a reaction coordinate A. The canonical ensemble distribution $P(A)$ can then be expressed as:

$$P(A_j) = \frac{P^*(A_j) \times \langle e^{\beta \Delta V(\mathbf{r})} \rangle_j}{\sum_{i=1}^{M} \langle P^*(A_i) e^{\beta \Delta V(\mathbf{r})} \rangle_i}, \quad j = 1, \ldots, M \tag{1.5}$$

where M is the number of bins that are evenly divided intervals for the reaction coordinate A, $\beta = 1/(k_B T)$, and $\langle e^{\beta \Delta V(\mathbf{r})} \rangle_j$ is the ensemble-average Boltzmann factor of $\Delta V(\mathbf{r})$ for MD simulation trajectory frames found in the j-th bin. The cumulant expansion can approximate the ensemble-average Boltzmann factor:

$$\langle e^{\beta \Delta V(\mathbf{r})} \rangle = \exp \left\{ \sum_{k=1}^{\infty} \frac{\beta^k}{k!} C_k \right\} \tag{1.6}$$

where the first two cumulants are defined as:

$$C_1 = \langle \Delta V \rangle, \tag{1.7}$$

$$C_2 = \langle \Delta V^2 \rangle - \langle \Delta V \rangle^2 = \sigma_V^2 \tag{1.8}$$

When the boost potential follows a near-Gaussian distribution, using a second-order cumulant expansion provides a more accurate reweighting than the Maclaurin series expansion. Consequently, the reweighted free energy $F(A_j)$ can be calculated as:

$$F(A_j) = -\frac{1}{\beta}\ln(P(A_j)) \tag{1.9}$$

Finally, Kramers' rate theory can be used for GaMD and LiGaMD simulations to estimate kinetic rate constants. This theory describes the rate of barrier crossing k_R in the large viscosity limit as a function of the system's viscosity and temperature:

$$k_R \approx \frac{2\pi\omega_m\omega_b}{\xi} e^{-\Delta G/(k_B T)} \tag{1.10}$$

where ω_m and ω_b are the harmonic oscillator frequencies near the energy minimum and barrier, respectively, ξ is the friction constant, and ΔG is the free energy barrier of the transition.

In summary, a key strength of GaMD and LiGaMD is their ability to operate without setting collective variables (CVs), which makes them broadly applicable even without prior knowledge of the system's reaction coordinates. GaMD and LiGaMD have been successfully applied to a wide range of systems, including intrinsically disordered proteins [Wang et al., 2021, Bao et al., 2024], enzyme catalysis [Miao and McCammon, 2017, Wang et al., 2021], and drug-target interactions [Bhattarai and Miao, 2018], highlighting their robustness and versatility in computational chemistry and biophysics.

1.2.2. Weighted ensemble method (WE)

The weighted ensemble method (WE) is an enhanced sampling technique used in molecular simulations to efficiently compute the rates and pathways of rare events, such as protein folding and ligand binding. It achieves this by focusing

computational resources on barrier regions of a molecule's conformational space without introducing any biasing forces. The WE strategy orchestrates numerous short, parallel simulations or "walkers." Each walker possesses a statistical weight that is periodically adjusted via a strict resampling algorithm consisting of cloning and merging steps. This resampling directs computational power to less-sampled regions of conformational space, enhancing overall sampling efficiency. The comprehensive details of WE are available in key publications such as the foundational paper [Huber and Kim, 1996], in a review article [Zuckerman and Chong, 2017], and the weighted ensemble simulation toolkit with parallelization and analysis (WESTPA) software papers for more extensive details [Zwier et al., 2015, Russo et al., 2022]. The general scheme proceeds as follows:

1. A WE simulation is initiated by defining several parameters *a priori*, including a set of collective variables (CVs) that partition the conformational space into discrete bins, a resampling time interval τ, and a target number of walkers per bin n_w. The initial ensemble of walkers is then generated with statistical weights that are normalized to sum to unity for the entire system.
2. Each walker in the ensemble is independently propagated for the time interval τ after which it is assigned to a bin based on its final CV coordinates.
3. A resampling procedure is applied within each bin, where walkers are merged or replicated in a statistically rigorous manner to maintain the target population, n_w.
4. The propagation and resampling steps (2 and 3) are iteratively repeated until statistical convergence is achieved.

The resampling process maintains a continuous population of walkers across all visited bins, which facilitates the effective navigation of high energy barriers. This strategy enhances computational efficiency by merging walkers in well-sampled, low energy regions and cloning them in sparsely populated regions that are often kinetically challenging. Because the underlying dynamics of the walkers are not altered, this statistically unbiased approach allows for the rigorous determination of both thermodynamic and kinetic properties from the simulation data.

While several parameters are defined *a priori*, the choice of the resampling time τ is notably flexible. It is not constrained by the Markovian property required by other enhanced sampling techniques, such as milestoning. The primary criterion for selecting τ is that it should be short enough to resolve critical transition events during propagation [Bhatt et al., 2010, Pratt et al., 2019, Ahn et al., 2020]. Nevertheless, a potential limitation is that WE can be computationally demanding if the simulation begins from initial states that are far from the system's steady-state population, as convergence across numerous bins is required.

1.2.3. Gaussian accelerated molecular dynamics-weighted ensemble method (GaMD-WE)

The hybrid GaMD-WE methodology [Ahn et al., 2021] is designed to combine the advantages of both the GaMD and the WE methods while overcoming their respective limitations. The protocol begins with an initial GaMD simulation to perform a broad exploration of the system's conformational landscape. The resulting trajectory provides a diverse ensemble of well-sampled configurations

that serve as the starting points for a subsequent WE simulation. This second stage yields a more refined free energy landscape that is closer to the steady-state distribution and enables the rigorous calculation of kinetic properties, such as interstate rate constants. An implementation of this hybrid approach is available as the GaMD-WE package (https://github.com/anandojha/gamd_we), which operates via a series of customizable scripts as detailed below.

1. The system of interest is first parameterized using an appropriate force field and subsequently solvated in an explicit or implicit solvent model.
2. Following solvation, the system is subjected to energy minimization, thermalization to a target temperature, and equilibration via conventional molecular dynamics (MD). The necessary directories for the GaMD production runs are then created.
3. Using the AMBER [Case et al., 2020] simulation package, a set of six independent GaMD simulations is conducted for a designated period. These simulations systematically explore the full range of potential boost combinations by applying lower and upper bound boosts to the dihedral potential, the total potential, and a dual-boost combination of both.
4. Data from the six GaMD trajectories are extracted and processed. The original free energy landscape is recovered by applying a reweighting scheme based on a third-order cumulant expansion to all bins populated with at least ten frames. Based on these reweighted probabilities, an ensemble of initial structures for the WE simulation is selected to satisfy the target number of walkers per bin n_w, with each structure assigned a corresponding weight. For bins containing an excess of frames, a random selection is performed.

5. The selected seed structures for the WE simulation are subjected to a two-stage minimization protocol in AMBER to prevent numerical instabilities. This involves an initial minimization of the protein backbone heavy atoms (C_α, N, C, O of the protein are minimized), followed by a full minimization of the entire system, including solvent.
6. The minimized structures are then organized into a directory structure suitable for initiating the WE simulation. The GaMD run that yielded the highest count of initial seed structures is chosen for the WE production phase, as this is indicative of the broadest exploration of the conformational space.

By using GaMD to generate well-distributed initial states, GaMD-WE aims to accelerate the convergence of WE simulations, particularly for complex systems where adequate initial sampling is a bottleneck. The GaMD phase provides a global overview of the energy landscape, guiding the more focused and statistically unbiased WE calculations. This combination allows for a more efficient determination of both equilibrium properties and kinetics of rare events compared to using either GaMD or WE in isolation, especially when prior knowledge of the system's energy landscape is limited for initiating WE.

1.2.4. Parallelizable Gaussian accelerated molecular dynamics (ParGaMD)

Despite GaMD's effectiveness in smoothing the potential energy surface to sample the configuration space efficiently, its conventional use within standard MD simulation engines (e.g., AMBER, OpenMM, and NAMD) encounters

significant hurdles in terms of parallel efficiency. Efforts to accelerate a single, extended GaMD simulation by distributing it over multiple GPUs often result in performance gains that are less than proportional to the added resources. This under-performance stems from established difficulties in standard MD parallelization strategies, including the considerable inter-GPU communication required for domain decomposition and force computations, intrinsic algorithmic components that do not lend themselves to parallel execution, and dependencies on system size, where more compact systems derive diminished advantages from multi-GPU configurations [Phillips et al., 2020]. Conversely, due to its fundamental design, WE presents an intrinsically powerful solution for leveraging parallel computing resources. WE simulations proceed by concurrently evolving numerous brief, statistically weighted trajectories termed "walkers" (further detailed in Section 1.2.2). The advancement of each walker constitutes an independent computational operation, rendering this part of the simulation "embarrassingly parallel" and exceptionally well-suited for deployment across multiple GPUs.

Exploiting these complementary strengths, parallelizable GaMD (ParGaMD) integrates GaMD-accelerated simulation techniques into the operational framework of the weighted ensemble method (WE) (see Figure 1.1b) [Sonti et al., 2025]. In a ParGaMD simulation, each WE walker is propagated using a GaMD-based MD simulation engine. This integration allows GaMD's harmonic boost potential to accelerate the conformational exploration of each walker by lowering energy barriers. Simultaneously, the WE framework facilitates large-scale parallelization by distributing these GaMD-enhanced walkers across multiple GPUs, while its ensemble oversight effectively captures rare events and reconstructs unbiased thermodynamics

(see Section 1.2.1). This synergy results in significantly improved sampling efficiency and a broader exploration of the configuration space compared to using either method alone. A significant practical benefit of ParGaMD is its ability to achieve near-linear performance scaling as the number of deployed GPUs increases, as shown in Figure 1.2. This translates into a considerable decrease in the wall-clock time required to reach equilibrium distributions of states. This enhanced efficiency makes it feasible to investigate complex biomolecular processes that were previously computationally prohibitive or extremely challenging to fully characterize, such as detailed ligand binding/unbinding pathways (including those involving deeply buried binding sites), the characterization of transient intermediate states, and the exploration of diverse conformational transitions. The ParGaMD package is implemented as a collection of customizable scripts. Versions are publicly available at https://github.com/Sonti974948/ParGaMD (for AMBER) and https://github.com/anugrahat/ParGaMD_MPS_TACC (for OpenMM). A summary of the workflow is outlined below.

1. **System preparation and initial GaMD simulations:** Prepare the biomolecular system as you would for a standard MD simulation (e.g., define force field, solvate, ionize). Then, perform the standard GaMD protocol (see [Miao et al., 2015]) to obtain the necessary finalized GaMD boost parameters (E, V_{max}, V_{min}, k_0). These parameters and energy statistics are typically saved in a GaMD restart file (`gamd_restart.dat`) and will be needed for the subsequent ParGaMD run.
2. **Set up the WE simulation for ParGaMD:** Choose CVs that effectively describe the process of interest. Divide the CV space into discrete "bins" of the configuration space. Set the resampling time τ for

the walkers and the target number of walkers per bin n_w.

3. **Run the ParGaMD simulation with the weighted ensemble simulation toolkit with parallelization and analysis (WESTPA) [Zwier et al., 2015, Russo et al., 2022]:** Each WE walker will run an independent GaMD simulation for the duration τ. The following steps are repeated until the desired convergence is reached:

   - **Initialize walkers:** Extract the initial coordinates for all walkers from the final frame of the parent GaMD trajectory. For each walker, generate initial velocities by sampling from a Maxwell-Boltzmann distribution at the target temperature, using a unique random seed to ensure statistical independence and promote diverse exploration of the conformational space.
   - **Apply GaMD boost potentials:** The finalized GaMD boost potential parameters (obtained from `gamd_restart.dat` in Step 1) are applied to each GaMD segment run by the WE walkers. MD simulation engines interfaced with WESTPA (such as AMBER or OpenMM with GaMD capabilities) can read these GaMD parameters.
   - **End of each τ interval or WE iteration:** WESTPA assesses and manages each walker's progress along the defined CVs, assigning the walkers to their respective bins. The WE resampling algorithm is then executed to split and merge walkers strategically, thereby maintaining the target population in each bin while ensuring a statistically rigorous redistribution of weights. Finally, all GaMD-specific parameters required for subsequent reweighting calculations are recorded in a dedicated log file for each walker.

4. **Data analysis and reweighting:** After a sufficient number of WE iterations is obtained to achieve the desired convergence, the log files from each walker can be used to carry out reweighting to reconstruct the true free energy landscape along the chosen CVs and rate constants using Kramer's rate theory.

Finally, to maximize the performance of ParGaMD on GPU clusters, effective resource management is important. Technologies like NVIDIA's multi-process service (MPS) can be beneficial, as they allow multiple CUDA processes (each corresponding to a separate WE/GaMD walker) to share a single GPU context. This can reduce the overhead from context switching and improve overall GPU utilization, especially when the computational demand of a single walker does not saturate a full GPU.

In essence, ParGaMD offers a scalable and robust method for enhanced sampling. By having each parallel WE walker execute GaMD dynamics, ParGaMD effectively flattens the energy landscape while globally directing the sampling towards regions of interest, making it a powerful tool for complex biomolecular simulations.

1.3. Applications of GaMD-WE

The GaMD-WE approach has been shown to enhance sampling efficiency for both thermodynamic and kinetic properties in biomolecular systems, exemplified by the folding of the mini-protein chignolin [Ahn et al., 2021]. Figures 1.3a, 1.3b, and 1.3c show the average free energy landscape ($-k_BT \ln P$, where P is probability) of chignolin from after 40 μs of brute-force simulation, 500 ns of GaMD simulation, and 500 ns of WE simulation, respectively. Three

independent simulations were run for each type of simulation, and the CVs were defined as the root mean square deviation (RMSD) from the folded state and the radius of gyration (R_g). Although GaMD and WE have comparable free energy landscape coverage, GaMD has probabilities closer to actual values due to reweighting, demonstrating that GaMD is faster at obtaining the true free energy landscape. This advantage allows the hybrid GaMD-WE method to determine rate constants with greater speed and accuracy than WE alone. Specifically, rate constants were calculated for transitions between the folded (0.0 Å $\leq$ RMSD $\leq$ 1.0 Å) and unfolded (RMSD $\geq$ 4.0 Å) regions depicted in Figure 1.3d. The evolution of these rate constants over time (Figures 1.3e and 1.3f) confirms that GaMD-WE converges more quickly to the reference values. The convergence is notably improved for the unfolded-to-folded rate constant (Figure 1.3f), as the initial populations generated by GaMD provide a superior sampling of the unfolded basin, which otherwise requires extensive simulation time for standalone WE to equilibrate.

The GaMD-WE method has also demonstrated potential to study complex biomolecular systems, including the bovine pancreatic trypsin inhibitor (BPTI) [Ahn et al., 2021]. For the BPTI system, the CVs were set as the dihedral angles χ_1-C14 and χ_1-C38 dihedral angles of the Cys14-Cys38 disulfide bond [Xue et al., 2012, Pierce et al., 2012]. Figures 1.4a, 1.4b, and 1.4c show the free energy landscape ($-k_BT\ln P$, where P is probability) of BPTI from 500 ns simulations of brute force MD, GaMD, and WE, respectively. Notably, the GaMD simulation employing the upper bound of the dual boost potential proved most effective in exploring the conformational space, outperforming five other GaMD potential settings (see Supplementary Information in [Ahn et al., 2021]), brute force MD, and WE. Specifically,

the metastable states of interest, including the major state M and two minor or excited states m_{C14} and m_{C38} [Grey et al., 2003] that are marked in Figure 1.4b, are present in GaMD. In contrast, the WE simulation failed to sample the sample m_{C14} state even after the simulation was extended to 2 μs (total simulation time: 2.5 μs) as seen in Figure 1.4d. These observations are consistent with previous reports where a 500 ns simulation using accelerated molecular dynamics (aMD), sampled key metastable states of BPTI that required a 1 ms brute-force simulation on the specialized Anton supercomputer [Pierce et al., 2012]. The superior sampling power of GaMD can be attributed to its CV-free enhanced sampling method, which allows it to explore motions orthogonal to the defined CVs. WE, on the other hand, primarily samples along the chosen CVs and can face challenges if important transitions involve orthogonal degrees of freedom. Although the rate constants for BPTI were not calculated, these results strongly suggest that the GaMD-WE framework would sample such kinetic events significantly faster than conventional WE.

1.4. Applications of ParGaMD

The ParGaMD approach has been applied to several biomolecular systems, including two protein-ligand systems [Sonti et al., 2025]. Specifically, ParGaMD has been applied to characterize the interaction between an α-synuclein monomer and a small-molecule inhibitor in development called paramagnetic amyloid ligand (PAL) [Hilt et al., 2022, Estaun-Panzano et al., 2023] (Figure 1.5). Elucidating the interactions between intrinsically disordered proteins (IDPs) implicated in neurodegenerative diseases, such as Alzheimer's

and Parkinson's, and small-molecule inhibitors in development is crucial for designing effective therapeutics. IDPs are particularly difficult to study effectively using conventional experimental and computational approaches since IDPs lack stable tertiary structures and exist as dynamic ensembles of conformations [Estaun-Panzano et al., 2023, Viegas et al., 2024], which gives rise to broad and shallow energy basins and a wide distribution of transition states [Estaun-Panzano et al., 2023]. ParGaMD addresses these challenges through its highly efficient sampling strategy. By simultaneously running multiple replicas and enabling frequent transitions between the bound and unbound states, ParGaMD generates a large number of statistically independent events [Sonti et al., 2025].

A comparison of ParGaMD and LiGaMD simulations revealed key differences in their abilities to explore the α-synuclein-PAL interactions thoroughly. Specifically, the ParGaMD simulations provided a more well-defined characterization of the interactions between the different states over a simulation period of 1.0 μs by capturing the free energy landscape and more binding and unbinding pathways compared to LiGaMD simulations that were run for the same simulation period [Sonti et al., 2025]. By comprehensively capturing various binding and unbinding states and pathways, including weak and transient interactions that are common in IDP systems [Chebaro et al., 2015, Robustelli et al., 2022], ParGaMD enables a detailed characterization of underlying mechanisms, providing valuable information to guide rational design strategies targeting IDPs.

ParGaMD has also been successfully applied to the simulation of complex ligand binding, particularly in systems featuring deeply buried and predominantly hydrophobic receptor pockets, such as that of the peroxisome

proliferator-activated receptor alpha (PPAR-α) [Sonti et al., 2025]. PPAR-α is a ligand-activated transcription factor critical in lipid metabolism, and its interaction with an endogenous fatty acid 10-hydroxystearic acid (10-HSA) presents a complex ligand-binding problem due to its buried orthosteric pocket and substantial conformational rearrangements that are needed for ligand entry and rearrangement [Marx et al., 1999, Torra et al., 2001, Chinetti et al., 2001, Lemieux et al., 2002]. Since the timescales associated with these events often exceed those accessible by classical MD simulations, the binding pathways for PPAR-α have been explored using GaMD and ParGaMD [Sonti et al., 2025]. Conventional GaMD simulations provided valuable information on 10-HSA binding to PPAR-α, which consistently identified a "tail-up" conformation of 10-HSA within the binding pocket (Figure 1.6b), a binding pose that has already been known from crystallographic studies of other agonists targeting the PPAR family [Xu et al., 1999, Chua and Bruning, 2021]. However, even with an extended simulation time of 5 μs with dual boost, the conventional GaMD simulations were unable to sample the canonical "tail-down" pose (similar to the conformation shown in Figure 1.6c), which has been observed in other crystal structures of the PPAR family [Kamata et al., 2020]. Instead, GaMD simulations often resulted in the ligand being trapped in intermediate states, such as the one depicted in Figure 1.6a, which prevented progression to the fully bound tail-down pose. This raised questions about whether the sampling was exhaustive or if the tail-down state was intrinsically less stable for 10-HSA.

To address these sampling limitations and achieve a more comprehensive understanding of the accessible conformations, ParGaMD was employed. Similar to the α-synuclein-PAL system as shown in Figure 1.2, ParGaMD

achieved a cumulative sampling speed of approximately 1406 ns/day with 12 GPUs, a nearly 8-fold increase compared to conventional GaMD with a speed of 179 ns/day on a single GPU [Sonti et al., 2025]. Furthermore, ParGaMD was able to sample the tail-up conformation of 10-HSA (Figure 1.6b), consistent with the conventional GaMD results, as well as the tail-down state within 2.5 μs of simulation time (Figure 1.6c) [Sonti et al., 2025]. The ability of ParGaMD to sample this conformation, which remained elusive in much longer conventional GaMD simulations, highlights ParGaMD's enhanced sampling power and more comprehensive exploration. Moreover, ParGaMD was able to sample multiple binding events and diverse bound structures, providing a more definitive mechanistic understanding of the complex ligand-receptor interaction. This case highlights the utility of ParGaMD in addressing complex biomolecular systems in drug discovery and molecular biophysics.

1.5. Conclusions

To address the inherent limitations of standalone GaMD and WE methods in exploring complex biomolecular phenomena, this chapter has detailed two hybrid strategies: GaMD-WE and ParGaMD. GaMD-WE strategically employs initial GaMD simulations to generate well-distributed starting configurations for subsequent WE calculations, thereby enhancing the efficient characterization of both thermodynamic and kinetic properties for systems ranging from fast-folding miniproteins to larger proteins with intricate conformational landscapes. ParGaMD, conversely, integrates GaMD dynamics directly within the WE framework, distributing numerous GaMD-accelerated walkers across multiple GPUs, which achieves significant computational

scalability and accelerates the construction of the free energy landscape, facilitating the detailed study of challenging systems, such as intrinsically disordered protein-ligand interactions and complex ligand binding pathways. Together, these hybrid methods offer marked improvements in sampling efficiency and provide reliable mechanistic insights, underscoring their substantial utility for advancing research in drug discovery and molecular biophysics.

Bibliography

Badi' Abdul-Wahid, Haoyun Feng, Dinesh Rajan, Ronan Costaouec, Eric Darve, Douglas Thain, and Jesus A Izaguirre. Awe-wq: fast-forwarding molecular dynamics using the accelerated weighted ensemble. *Journal of Chemical Information and Modeling*, 54(10):3033–3043, 2014.

Joshua L Adelman and Michael Grabe. Simulating current–voltage relationships for a narrow ion channel using the weighted ensemble method. *Journal of Chemical Theory and Computation*, 11(4):1907–1918, 2015.

Upendra Adhikari, Barmak Mostofian, Jeremy Copperman, Sundar Raman Subramanian, Andrew A Petersen, and Daniel M Zuckerman. Computational estimation of microsecond to second atomistic folding times. *Journal of the American Chemical Society*, 141(16):6519–6526, 2019.

Surl-Hee Ahn and Jay W Grate. Foldamer architectures of triazine-based sequence-defined polymers investigated with molecular dynamics simulations and enhanced sampling methods. *Journal of Physical Chemistry B*, 123(44): 9364–9377, 2019.

Surl-Hee Ahn, Jay W Grate, and Eric F Darve. Efficiently sampling conformations and pathways using the concurrent adaptive sampling (cas) algorithm. *Journal of Chemical Physics*, 147(7):074115, 2017.

Surl-Hee Ahn, Jay W Grate, and Eric F Darve. Investigating the role of non-covalent interactions in conformation and assembly of triazine-based sequence-defined polymers. *Journal of Chemical Physics*, 149(7):072330, 2018.

Surl-Hee Ahn, Benjamin R Jagger, and Rommie E Amaro. Ranking of ligand binding kinetics using a weighted ensemble approach and comparison with a multiscale milestoning approach. *Journal of Chemical Information and Modeling*, 60(11):5340–5352, 2020.

Surl-Hee Ahn, Anupam A. Ojha, Rommie E. Amaro, and J. Andrew McCammon. Gaussian-accelerated molecular dynamics with the weighted ensemble method: A hybrid method improves thermodynamic and kinetic sampling. *Journal of Chemical Theory and Computation*, 17(12):7938–7951, 2021.

Rosalind J Allen, Daan Frenkel, and Pieter Rein Ten Wolde. Forward flux sampling-type schemes for simulating rare events: Efficiency analysis. *Journal of Chemical Physics*, 124(19):194111, 2006.

Rosalind J Allen, Chantal Valeriani, and Pieter Rein Ten Wolde. Forward flux sampling for rare event simulations. *Journal of Condensed Matter Physics*, 21(46):463102, 2009.

Huayin Bao, Wei Wang, Haibo Sun, and Jianzhong Chen. The switch states of the gdp-bound hras affected by point mutations: A study from gaussian

accelerated molecular dynamics simulations and free energy landscapes. *Journal of Biomolecular Structure and Dynamics*, 42(7):3363–3381, 2024.

Alessandro Barducci, Giovanni Bussi, and Michele Parrinello. Well-tempered metadynamics: a smoothly converging and tunable free-energy method. *Physical Review Letters*, 100(2):020603, 2008.

Alessandro Barducci, Massimiliano Bonomi, and Michele Parrinello. Metadynamics. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(5):826–843, 2011.

Juan M Bello-Rivas and Ron Elber. Exact milestoning. *Journal of Chemical Physics*, 142(9):03B602_1, 2015.

Divesh Bhatt, Bin W Zhang, and Daniel M Zuckerman. Steady-state simulations using weighted ensemble path sampling. *Journal of Chemical Physics*, 133(1):014110, 2010.

Apurba Bhattarai and Yinglong Miao. Gaussian accelerated molecular dynamics for elucidation of drug pathways. *Expert opinion on drug discovery*, 13(11):1055–1065, 2018.

Ernesto E Borrero, Marcus Weinwurm, and Christoph Dellago. Optimizing transition interface sampling simulations. *Journal of Chemical Physics*, 134 (24):244118, 2011.

Nicolae-Viorel Buchete and Gerhard Hummer. Peptide folding kinetics from replica exchange molecular dynamics. *Physical Review E*, 77(3):030902, 2008.

Alex Bunker and Burkhard Dünweg. Parallel excluded volume tempering for polymer melts. *Physical Review E*, 63(1):016701, 2000.

Giovanni Bussi, Alessandro Laio, and Michele Parrinello. Equilibrium free energies from nonequilibrium metadynamics. *Physical Review Letters*, 96 (9):090601, 2006.

David A Case, Kellon Belfon, Ido Ben-Shalom, Scott R Brozell, David Cerutti, Thomas Cheatham, Vinícius Wilian D Cruzeiro, Tom Darden, Robert E Duke, George Giambasu, et al. Amber 2020. 2020.

Yassmine Chebaro, Andrew J. Ballard, Debayan Chakraborty, and David J. Wales. Intrinsically disordered energy landscapes. *Scientific Reports*, 5: 10386, 2015. doi: 10.1038/srep10386.

Haochuan Chen, Haohao Fu, Christophe Chipot, Xueguang Shao, and Wensheng Cai. Overcoming free-energy barriers with a seamless combination of a biasing force and a collective variable-independent boost potential. *Journal of Chemical Theory and Computation*, 2021.

Giulia Chinetti, Sophie Lestavel, Virginie Bocher, Alan T Remaley, Bernadette Neve, Inés Pineda Torra, Elisabeth Teissier, Anne Minnich, Michael Jaye, Nicolas Duverger, et al. Ppar-α and ppar-γ activators induce cholesterol removal from human macrophage foam cells through stimulation of the abca1 pathway. *Nature medicine*, 7(1):53–58, 2001.

Christophe Chipot and Andrew Pohorille. Free energy calculations. 86: 159–184, 2007.

John D Chodera, William C Swope, Jed W Pitera, and Ken A Dill. Long-time protein folding dynamics from short-time molecular dynamics simulations. *Multiscale Modeling & Simulation*, 5(4):1214–1226, 2006.

Benjamin SK Chua and John B Bruning. Pparα and δ ligand design: Honing the traditional empirical method with a more holistic overview. *Nuclear Receptors: The Art and Science of Modulator Design and Discovery*, pages 111–178, 2021.

Jeffrey Comer, James C Gumbart, Jérôme Hénin, Tony Lelièvre, Andrew Pohorille, and Christophe Chipot. The adaptive biasing force method: Everything you always wanted to know but were afraid to ask. *Journal of Physical Chemistry B*, 119(3):1129–1151, 2015.

Jeremy Copperman and Daniel M Zuckerman. Accelerated estimation of long-timescale kinetics from weighted ensemble simulation via non-markovian "microbin" analysis. *Journal of Chemical Theory and Computation*, 16(11):6763–6775, 2020.

Ronan Costaouec, Haoyun Feng, Jesús Izaguirre, and Eric Darve. Analysis of the accelerated weighted ensemble methodology. In *Conference Publications*, volume 2013, page 171. American Institute of Mathematical Sciences, 2013.

Eric Darve and Andrew Pohorille. Calculating free energies using average force. *Journal of Chemical Physics*, 115(20):9169–9183, 2001.

Eric Darve, David Rodríguez-Gómez, and Andrew Pohorille. Adaptive biasing force method for scalar and vector free energy calculations. *Journal of Chemical Physics*, 128(14):144120, 2008.

Marco De Vivo, Matteo Masetti, Giovanni Bottegoni, and Andrea Cavalli. Role of molecular dynamics and related methods in drug discovery. *Journal of Medicinal Chemistry*, 59(9):4035–4061, 2016.

Alex Dickson and Charles L Brooks III. Wexplore: hierarchical exploration of high-dimensional spaces using the weighted ensemble algorithm. *Journal of Physical Chemistry B*, 118(13):3532–3542, 2014.

Tom Dixon, Samuel D Lotz, and Alex Dickson. Predicting ligand binding affinity using on-and off-rates for the sampl6 sampling challenge. *Journal of Computer-Aided Molecular Design*, 32(10):1001–1012, 2018.

Nazanin Donyapour, Nicole M Roussey, and Alex Dickson. Revo: Resampling of ensembles by variation optimization. *Journal of Chemical Physics*, 150 (24):244112, 2019.

Ron O Dror, Robert M Dirks, JP Grossman, Huafeng Xu, and David E Shaw. Biomolecular simulation: a computational microscope for molecular biology. *Annual Review of Biophysics*, 41:429–452, 2012.

Wei-Na Du, Kristen A Marino, and Peter G Bolhuis. Multiple state transition interface sampling of alanine dipeptide in explicit solvent. *Journal of Chemical Physics*, 135(14):10B609, 2011.

Jacob D Durrant and J Andrew McCammon. Molecular dynamics simulations and drug discovery. *BMC Biology*, 9(1):1–9, 2011.

Peter Eastman, Jason Swails, John D Chodera, Robert T McGibbon, Yutong Zhao, Kyle A Beauchamp, Lee-Ping Wang, Andrew C Simmonett, Matthew P Harrigan, Chaya D Stern, et al. Openmm 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology*, 13(7):e1005659, 2017.

Juan Estaun-Panzano, Marie-Laure Arotcarena, and Erwan Bezard.

Monitoring α-synuclein aggregation. *Neurobiology of disease*, 176:105966, 2023.

Roland Faller, Qiliang Yan, and Juan J. de Pablo. Multicanonical parallel tempering. *The Journal of Chemical Physics*, 116(13):5419–5423, apr 2002. doi: 10.1063/1.1456504. URL https://doi.org/10.1063/1.1456504.

Anton K Faradjian and Ron Elber. Computing time scales from reaction coordinates by milestoning. *Journal of Chemical Physics*, 120(23): 10880–10889, 2004.

Michael J Grey, Chunyu Wang, and Arthur G Palmer. Disulfide bond isomerization in basic pancreatic trypsin inhibitor: multisite chemical exchange quantified by cpmg relaxation dispersion and chemical shift modeling. *Journal of the American Chemical Society*, 125(47):14324–14335, 2003.

Silvia Hilt, Ruiwu Liu, Izumi Maezawa, Tatu Rojalin, Hnin H Aung, Madhu Budamagunta, Ryan Slez, Qizhi Gong, Randy P Carney, and John C Voss. Novel stilbene-nitroxyl hybrid compounds display discrete modulation of amyloid beta toxicity and structure. *Frontiers in Chemistry*, 10:896386, 2022.

Scott A Hollingsworth and Ron O Dror. Molecular dynamics simulation for all. *Neuron*, 99(6):1129–1143, 2018.

Adam Hospital, Josep Ramon Goñi, Modesto Orozco, and Josep L Gelpí. Molecular dynamics simulations: advances and applications. *Advances and Applications in Bioinformatics and Chemistry*, 8:37, 2015.

Yu-ming M Huang, J Andrew McCammon, and Yinglong Miao. Replica exchange gaussian accelerated molecular dynamics: Improved enhanced sampling and free energy calculation. *Journal of Chemical Theory and Computation*, 14(4):1853–1864, 2018.

Gary A. Huber and Samuel Kim. Weighted-ensemble brownian dynamics simulations for protein association reactions. *Biophysical Journal*, 70:97–110, 1996.

Sarwar Hussain and Amir Haji-Akbari. Studying rare events using forward-flux sampling: Recent breakthroughs and future outlook. *Journal of Chemical Physics*, 152(6):060901, 2020.

Shotaro Kamata, Takuji Oyama, Kenta Saito, Akihiro Honda, Yume Yamamoto, Keisuke Suda, Ryo Ishikawa, Toshimasa Itoh, Yasuo Watanabe, Takahiro Shibata, et al. Pparα ligand-binding domain structures with endogenous fatty acids and fibrates. *Iscience*, 23(11), 2020.

Martin Karplus and John Kuriyan. Molecular dynamics and protein function. *Proceedings of the National Academy of Sciences of the United States of America*, 102(19):6679–6685, 2005.

Martin Karplus and J Andrew McCammon. Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9):646–652, 2002.

Johannes Kästner. Umbrella sampling. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 1(6):932–942, 2011.

Alessandro Laio and Francesco L Gervasio. Metadynamics: a method to simulate rare events and reconstruct the free energy in biophysics, chemistry and material science. *Reports on Progress in Physics*, 71(12):126601, 2008.

Alessandro Laio, Antonio Rodriguez-Fortea, Francesco Luigi Gervasio, Matteo Ceccarelli, and Michele Parrinello. Assessing the accuracy of metadynamics. *Journal of Physical Chemistry B*, 109(14):6714–6721, 2005.

Tony Lelièvre, Mathias Rousset, and Gabriel Stoltz. Long-time convergence of an adaptive biasing force method. *Nonlinearity*, 21(6):1155, 2008.

Isabelle Lemieux, Luc Laperrière, Vladimir Dzavik, Gérald Tremblay, Joanne Bourgeois, and Jean-Pierre Després. A 16-week fenofibrate treatment increases ldl particle size in type iia dyslipidemic patients. *Atherosclerosis*, 162(2):363–371, 2002.

Michael Levitt. Protein folding by restrained energy minimization and molecular dynamics. *Journal of Molecular Biology*, 170(3):723–764, 1983.

Kresten Lindorff-Larsen, Stefano Piana, Ron O Dror, and David E Shaw. How fast-folding proteins fold. *Science*, 334(6055):517–520, 2011.

Samuel D Lotz and Alex Dickson. Unbiased molecular dynamics of 11 min timescale drug unbinding reveals transition state stabilizing interactions. *Journal of the American Chemical Society*, 140(2):618–628, 2018.

EJ Maginn and JR Elliott. Historical perspective and current outlook for molecular dynamics as a chemical engineering tool. *Industrial & Engineering Chemistry Research*, 49(7):3059–3078, 2010.

Peter Májek and Ron Elber. Milestoning without a reaction coordinate. *Journal of Chemical Theory and Computation*, 6(6):1805–1817, 2010.

Nikolaus Marx, Galina K Sukhova, Tucker Collins, Peter Libby, and Jorge Plutzky. Pparα activators inhibit cytokine-induced vascular cell adhesion

molecule-1 expression in human endothelial cells. *Circulation*, 99(24): 3125–3131, 1999.

Yinglong Miao. Acceleration of biomolecular kinetics in gaussian accelerated molecular dynamics. *Journal of Chemical Physics*, 149(7):072308, 2018.

Yinglong Miao and J Andrew McCammon. Gaussian accelerated molecular dynamics: Theory, implementation, and applications. *Annual Reports in Computational Chemistry*, 13:231–278, 2017.

Yinglong Miao, Victoria A. Feher, and J. Andrew McCammon. Gaussian accelerated molecular dynamics: Unconstrained enhanced sampling and free energy calculation. *Journal of Chemical Theory and Computation*, 11(8): 3584–3595, 2015.

Yinglong Miao, Ashok Bhattarai, and Jing Wang. Ligand gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *Journal of Chemical Theory and Computation*, 16:5526–5547, 2020.

Daniele Moroni, Titus S van Erp, and Peter G Bolhuis. Investigating rare events by transition interface sampling. *Physica A: Statistical Mechanics and its Applications*, 340(1-3):395–401, 2004.

Hiraku Oshima, Suyong Re, and Yuji Sugita. Replica-exchange umbrella sampling combined with gaussian accelerated molecular dynamics for free-energy calculation of biomolecules. *Journal of Chemical Theory and Computation*, 15(10):5199–5208, 2019.

Albert C Pan, Daniel Jacobson, Konstantin Yatsenko, Duluxan Sritharan, Thomas M Weinreich, and David E Shaw. Atomic-level characterization of

protein–protein association. *Proceedings of the National Academy of Sciences of the United States of America*, 116(10):4244–4249, 2019.

Yui Tik Pang, Yinglong Miao, Yi Wang, and J Andrew McCammon. Gaussian accelerated molecular dynamics in namd. *Journal of Chemical Theory and Computation*, 13(1):9–19, 2017.

Dietmar Paschek and Angel E García. Reversible temperature and pressure denaturation of a protein fragment:¡? format?¿ a replica exchange molecular dynamics simulation study. *Physical review letters*, 93(23):238105, 2004.

James C Phillips, David J Hardy, Julio DC Maia, John E Stone, João V Ribeiro, Rafael C Bernardi, Ronak Buch, Giacomo Fiorin, Jérôme Hénin, Wei Jiang, et al. Scalable molecular dynamics on cpu and gpu architectures with namd. *Journal of Chemical Physics*, 153(4):044130, 2020.

Levi CT Pierce, Romelia Salomon-Ferrer, Cesar Augusto F. de Oliveira, J Andrew McCammon, and Ross C Walker. Routine access to millisecond time scale events with accelerated molecular dynamics. *Journal of Chemical Theory and Computation*, 8(9):2997–3002, 2012.

AJ Pratt, Ernesto Suárez, Daniel M Zuckerman, and Lillian T Chong. Extensive evaluation of weighted ensemble strategies for calculating rate constants and binding affinities of molecular association/dissociation processes. *bioRxiv*, page 671172, 2019.

Dhiman Ray, Sharon Emily Stone, and Ioan Andricioaei. Markovian weighted ensemble milestoning (m-wem): Long-time kinetics from short trajectories. *Journal of Chemical Theory and Computation*, 18(1):79–95, 2021.

Evan J Reed, Marin Soljačić, Richard Gee, and JD Joannopoulos. Molecular dynamics simulations of coherent optical photon emission from shock waves in crystals. *Physical Review B*, 75(17):174302, 2007.

Paul Robustelli, Alain Ibanez-de Opakua, Cecily Campbell-Bezat, Fabrizio Giordanetto, Stefan Becker, Markus Zweckstetter, Albert C Pan, and David E Shaw. Molecular basis of small-molecule binding to α-synuclein. *Journal of the American Chemical Society*, 144(6):2501–2510, 2022.

Edina Rosta and Gerhard Hummer. Error and efficiency of replica exchange molecular dynamics simulations. *Journal of Chemical Physics*, 131(16): 10B615, 2009.

John D Russo, Jeremy Copperman, and Daniel M Zuckerman. Iterative trajectory reweighting for estimation of equilibrium and non-equilibrium observables. *arXiv Preprint arXiv:2006.09451*, 2020.

John D. Russo, She Zhang, Jeremy M. G. Leung, Anthony T. Bogetti, Jeff P. Thompson, Alex J. DeGrave, Paul A. Torrillo, A. J. Pratt, Kim F. Wong, Junchao Xia, Jeremy Copperman, Joshua L. Adelman, Matthew C. Zwier, David N. LeBard, Daniel M. Zuckerman, and Lillian T. Chong. WESTPA 2.0: High-performance upgrades for weighted ensemble simulations and analysis of longer-timescale applications. *Journal of Chemical Theory and Computation*, 18(2):638–649, 2022.

Ali S Saglam and Lillian T Chong. Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chemical Science*, 10(8):2360–2372, 2019.

Daniel Sindhikara, Yilin Meng, and Adrian E Roitberg. Exchange frequency

in replica exchange molecular dynamics. *Journal of Chemical Physics*, 128 (2):01B609, 2008.

Daniel J Sindhikara, Daniel J Emerson, and Adrian E Roitberg. Exchange often and properly in replica exchange molecular dynamics. *Journal of Chemical Theory and Computation*, 6(9):2804–2808, 2010.

Siddharth Sonti, Anugraha Thyagatur, Hung-Yu Wan, Maxen Hamelynck, Roland Faller, and Surl-Hee Ahn. Accelerating free energy exploration using parallelizable gaussian accelerated molecular dynamics (pargamd). *ChemRxiv*, 2025. doi: 10.26434/chemrxiv-2025-rr5v9. This content is a preprint and has not been peer-reviewed.

Martin O Steinhauser and Stefan Hiermaier. A review of computational methods in materials science: examples from shock-wave and polymer physics. *International Journal of Molecular Sciences*, 10(12):5135–5216, 2009.

Lukas S Stelzl and Gerhard Hummer. Kinetics from replica exchange molecular dynamics simulations. *Journal of Chemical Theory and Computation*, 13(8): 3927–3935, 2017.

Yuji Sugita and Yuko Okamoto. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, 314(1-2):141–151, 1999.

Terra Sztain, Surl-Hee Ahn, Anthony T Bogetti, Lorenzo Casalino, Jory A Goldsmith, Evan Seitz, Ryan S McCool, Fiona L Kearns, Francisco Acosta-Reyes, Suvrajit Maji, Ghoncheh Mashayekhi, J. Andrew McCammon, Abbas Ourmazd, Joachim Frank, Jason S. McLellan, Lillian T.

Chong, and Rommie E. Amaro. A glycan gate controls opening of the sars-cov-2 spike protein. *Nature Chemistry*, 13(10):963–968, 2021.

Pratyush Tiwary and Michele Parrinello. From metadynamics to dynamics. *Physical Review Letters*, 111(23):230602, 2013.

Inés Pineda Torra, Giulia Chinetti, Caroline Duval, Jean-Charles Fruchart, and Bart Staels. Peroxisome proliferator-activated receptors: from transcriptional control to clinical practice. *Current Opinion in Lipidology*, 12(3):245–254, 2001.

Glenn M Torrie and John P Valleau. Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics*, 23(2):187–199, 1977.

Chantal Valeriani, Rosalind J Allen, Marco J Morelli, Daan Frenkel, and Pieter Rein ten Wolde. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. *Journal of Chemical Physics*, 127(11):114109, 2007.

Titus S Van Erp and Peter G Bolhuis. Elaborating transition interface sampling methods. *Journal of Computational Physics*, 205(1):157–181, 2005.

Eric Vanden-Eijnden and Maddalena Venturoli. Markovian milestoning with voronoi tessellations. *Journal of Chemical Physics*, 130(19):194101, 2009.

Eric Vanden-Eijnden, Maddalena Venturoli, Giovanni Ciccotti, and Ron Elber. On the assumptions underlying milestoning. *Journal of Chemical Physics*, 129(17):174102, 2008.

Rafael G Viegas, Ingrid BS Martins, and Vitor BP Leite. Understanding the

energy landscape of intrinsically disordered protein ensembles. *Journal of Chemical Information and Modeling*, 64(10):4149–4157, 2024.

Peter Virnau and Marcus Müller. Calculation of free energy through successive umbrella sampling. *Journal of Chemical Physics*, 120(23):10925–10930, 2004.

Jinan Wang and Yinglong Miao. Ligand gaussian accelerated molecular dynamics 2 (ligamd2): Improved calculations of ligand binding thermodynamics and kinetics with closed protein pocket. *Journal of Chemical Theory and Computation*, 19(3):733–745, 2023.

Jinan Wang and Yinglong Miao. Ligand gaussian accelerated molecular dynamics 3 (ligamd3): Improved calculations of binding thermodynamics and kinetics of both small molecules and flexible peptides. *Journal of Chemical Theory and Computation*, 20(14):5829–5841, 2024.

Jinan Wang, Pablo R Arantes, Apurba Bhattarai, Rohaine V Hsu, Shristi Pawnikar, Yu-ming M Huang, Giulia Palermo, and Yinglong Miao. Gaussian accelerated molecular dynamics: Principles and applications. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, page e1521, 2021.

Aryeh Warmflash, Prabhakar Bhimalapuram, and Aaron R Dinner. Umbrella sampling for nonequilibrium processes. *Journal of Chemical Physics*, 127 (15):114109, 2007.

H Eric Xu, Millard H Lambert, Valerie G Montana, Derek J Parks, Steven G Blanchard, Peter J Brown, Daniel D Sternbach, Jürgen M Lehmann, G Bruce Wisely, Timothy M Willson, et al. Molecular recognition of fatty acids by

peroxisome proliferator–activated receptors. *Molecular cell*, 3(3):397–403, 1999.

Ji Xu, Xiaoxia Li, Chaofeng Hou, Limin Wang, Guangzheng Zhou, Wei Ge, and Jinghai Li. Engineering molecular dynamics simulation in chemical engineering. *Chemical Engineering Science*, 121:200–216, 2015.

Yi Xue, Joshua M Ward, Tairan Yuwen, Ivan S Podkorytov, and Nikolai R Skrynnikov. Microsecond time-scale conformational exchange in proteins: using long molecular dynamics trajectory to simulate nmr relaxation dispersion data. *Journal of the American Chemical Society*, 134(5): 2555–2562, 2012.

Li Yu, Dinesh Rajan, Haoyun Feng, Eric Darve, Douglas Thain, Jesus A Izaguirre, et al. Folding proteins at 500 ns/hour with work queue. In *2012 IEEE 8th International Conference on E-Science*, pages 1–8. IEEE, 2012.

Bin W Zhang, David Jasnow, and Daniel M Zuckerman. The "weighted ensemble" path sampling method is statistically exact for a broad class of stochastic processes and binning procedures. *Journal of Chemical Physics*, 132(5):054107, 2010.

Wei Zhang, Chun Wu, and Yong Duan. Convergence of replica exchange molecular dynamics. *Journal of Chemical Physics*, 123(15):154105, 2005.

Jingxuan Zhu, Yishuo Lv, Xiaosong Han, Dong Xu, and Weiwei Han. Understanding the differences of the ligand binding/unbinding pathways between phosphorylated and non-phosphorylated arh1 using molecular dynamics simulations. *Scientific Reports*, 7(1):1–14, 2017.

Daniel M. Zuckerman and Lillian T. Chong. Weighted ensemble simulation: Review of methodology, applications, and software. *Annual Review of Biophysics*, 46:43–57, 2017.

Matthew C Zwier, Joshua L Adelman, Joseph W Kaus, Adam J Pratt, Kim F Wong, Nicholas B Rego, Ernesto Suárez, Steven Lettieri, David W Wang, Michael Grabe, et al. Westpa: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *Journal of Chemical Theory and Computation*, 11(2):800–809, 2015.

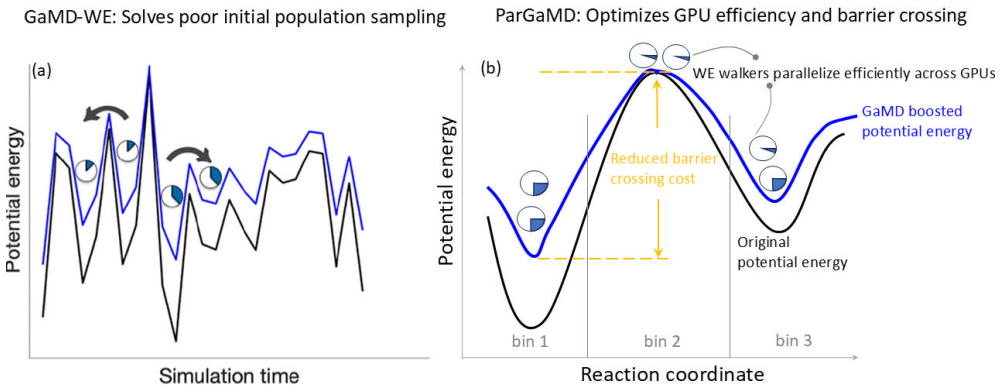

Figure 1.1: Schematic of the two different hybrid Gaussian accelerated molecular dynamics (GaMD) and the weighted ensemble (WE) methods. (a) Schematic of the Gaussian accelerated molecular dynamics-weighted ensemble method (GaMD-WE), where the GaMD simulation first flattens the energy landscape with its harmonic boost potential (blue line) and produces initial populations for the WE simulations to enhance sampling. Image is from [Ahn et al., 2021], which is reproduced with permission. (b) Schematic of parallelizable Gaussian accelerated molecular dynamics (ParGaMD), which integrates GaMD directly within the WE framework and simultaneously smoothens the potential energy surface (green line) and propagates multiple weighted walkers concurrently to overcome energy barriers efficiently.

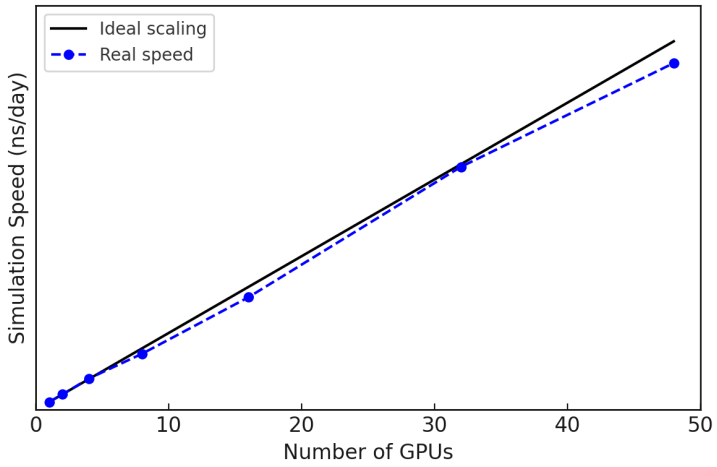

Figure 1.2: GPU scaling performance for an α-synuclein monomer and a small molecule inhibitor (paramagnetic amyloid ligand or PAL) system. The linear scaling line (in black) is a benchmark for perfect parallel efficiency. In contrast, the real speed line (in blue) reflects the practical performance gains observed across different GPU counts. ParGaMD achieves near-linear parallel scaling and demonstrates its excellent computational efficiency in leveraging multiple GPUs.

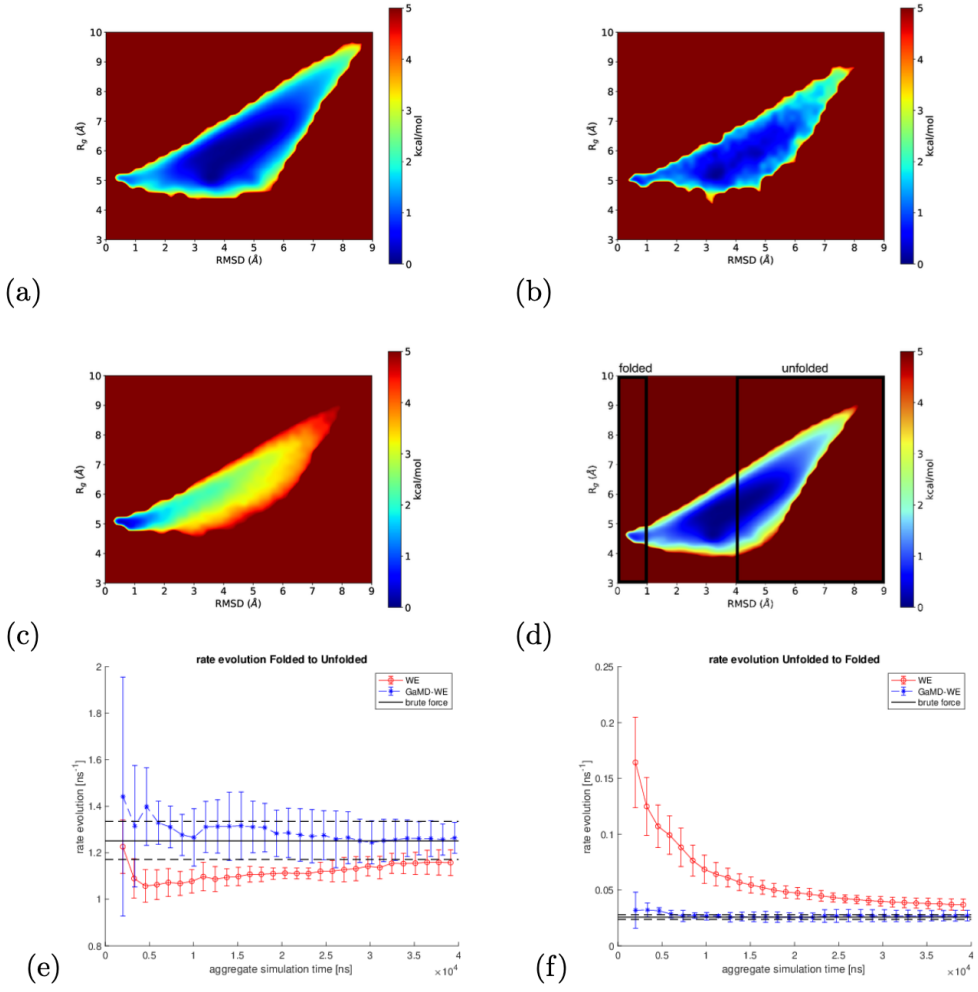

Figure 1.3: Gaussian accelerated molecular dynamics-weighted ensemble method (GaMD-WE) can efficiently sample both thermodynamic and kinetic properties for chignolin. (a), (b), and (c) show the average free energy landscape of chignolin after 40 μs of brute force molecular dynamics (MD) simulation, 500 ns of Gaussian accelerated molecular dynamics (GaMD) simulation, and 500 ns of weighted ensemble method (WE) simulation, respectively. (d) shows the regions of interest marked. (e) and (f) show the evolution of rate constants between the folded and unfolded regions over aggregate simulation time for WE (in red) and GaMD-WE (in blue). The reference brute force values are in black. All images are from [Ahn et al., 2021], which are reproduced with permission.

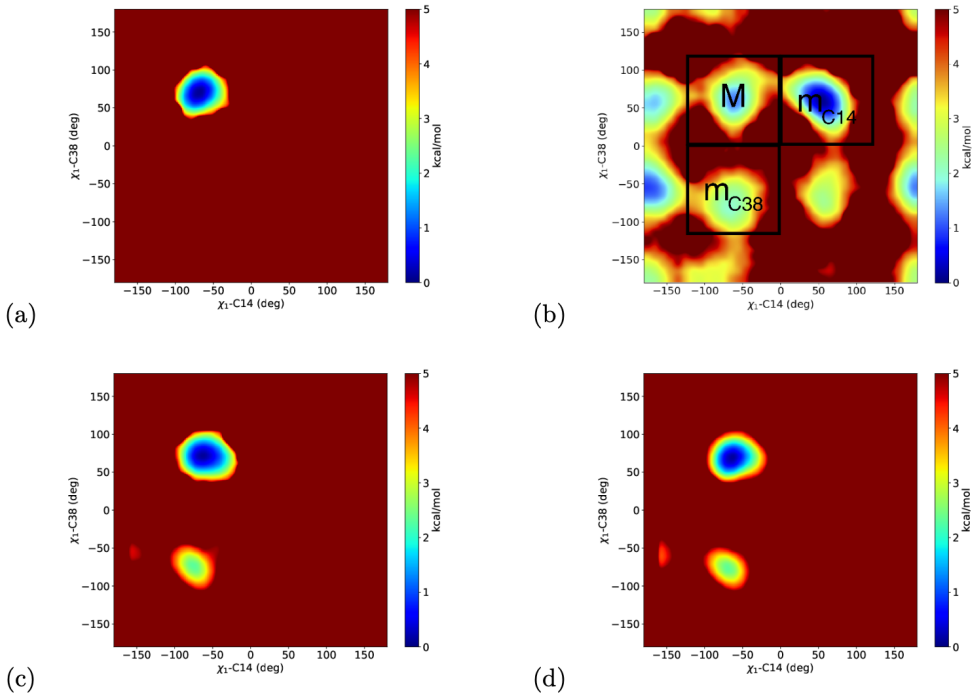

Figure 1.4: Free energy landscape of bovine pancreatic trypsin inhibitor (BPTI) after (a) one 500 ns brute force molecular dynamics (MD) simulation run, (b) one 500 ns run of Gaussian accelerated molecular dynamics (GaMD) with metastable states of interest M, m_{C14}, and m_{C38} marked, (c) one 500 ns run of the weighted ensemble method (WE), and (d) extending WE run for 2 μs (total simulation time: 2.5 μs), respectively. All images are from [Ahn et al., 2021], which are reproduced with permission.

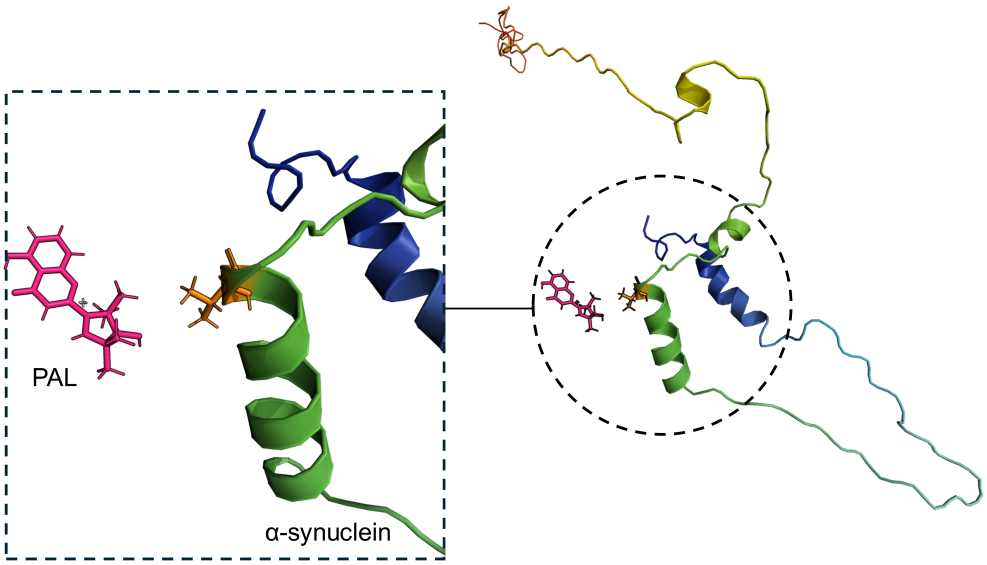

Figure 1.5: Binding of the paramagnetic amyloid ligand (PAL) to an α-synuclein monomer, an intrinsically disordered protein (IDP) associated with Parkinson's disease. The structure shown highlights a representative bound pose, where the ligand or PAL (pink) interacts with a transient state of α-synuclein, which provides mechanistic insight into IDP–ligand recognition.

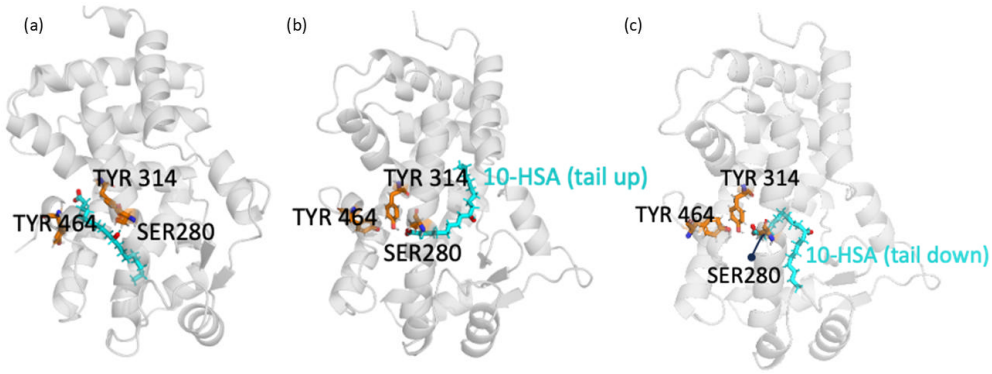

Figure 1.6: Binding conformations of (R)-10-hydroxystearic acid (10-HSA) in peroxisome proliferator-activated receptor α (PPAR-α). (a) An intermediate binding state identified from Gaussian accelerated molecular dynamics (GaMD) simulations, where the 10-HSA tail (cyan) initiates entry towards the "tail-down" conformation but is not yet fully bound. The ligand interacts with key residues (TYR 314, TYR 464, SER280; orange) of PPAR-α (gray). (b) The "tail-up" conformation of 10-HSA, predicted by both conventional GaMD and parallelizable Gaussian accelerated molecular dynamics (ParGaMD) simulations. (c) The "tail-down" conformation of 10-HSA was only captured by ParGaMD simulations.