Challenges and Frontiers of Computational Modeling of Biomolecular Recognition

Jinan Wang, Apurba Bhattarai, Hung Nguyen Do, Yinglong Miao*

Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, KS 66047

*Email: miao@ku.edu

Short Title: Modeling of Biomolecular Recognition

Abstract:

Biomolecular recognition including binding of small molecules, peptides and proteins to their target receptors plays a key role in cellular function and has been targeted for therapeutic drug design. However, the high flexibility of biomolecules and slow binding and dissociation processes have presented challenges for computational modeling. Here, we review the challenges and computational approaches developed to characterize biomolecular binding, including molecular docking, Molecular Dynamics (MD) simulations (especially enhanced sampling) and Machine Learning. Further improvements are still needed in order to accurately and efficiently characterize binding structures, mechanisms, thermodynamics and kinetics of biomolecules in the future.

Keywords: Biomolecular Recognition, Thermodynamics, Kinetics, Molecular dynamics, Enhanced Sampling, Machine Learning.

1. Introduction

Biomolecular recognition plays key roles in many fundamental biological processes, including immune responses, cellular signal transduction, and so on (Nooren & Thornton, 2003). Moreover, these processes are implicated in the development of numerous human diseases and serve as important drug targets (Ferreira et al., 2016; Scott et al., 2016). Experimental techniques (Miura, 2018) including X-ray crystallography, nuclear magnetic resonance (NMR) and cryo-electron microscopy (cryo-EM) have been applied to determine the bound structures of protein-small molecule, protein-peptide and protein-protein complexes. The number of experimental complex structures are significantly increased in recent years (Sussman et al., 1998). However, it is still rather time consuming and resource demanding to obtain high-resolution experimental structures. Moreover, the experimental structures often capture static pictures of protein complexes. Intermediate conformational states that could be relevant for drug design are usually difficult to probe using current experimental techniques.

Computational methods have been developed to model biomolecular recognition and predict the binding free energies and/or kinetics rates, including the widely used molecular docking (Ciemny et al., 2018; Morris et al., 2009; Porter et al., 2017; Vakser, 2020; Wang & Zhu, 2016), Brownian Dynamics (Ermak & McCammon, 1978; Gabdoulline & Wade, 2001; Spaar et al., 2006; Votapka & Amaro, 2015; Wieczorek & Zielenkiewicz, 2008) and Molecular Dynamics (MD) simulations (Basdevant et al., 2013; He et al., 2021; Karplus & McCammon, 2002; Lamprakis et al., 2021; Pan et al., 2019). Molecular docking has been widely used for predicting the *holo* structures of protein-ligand (Wang & Zhu, 2016), protein-peptide (Ciemny et al., 2018) and protein-protein complexes (Vakser, 2020). Although significant improvements have been achieved in developments of the molecular docking algorithms, the accuracy of docking could be

still limited, due to high system flexibility especially in docking of the peptides and proteins. Recently, Deep Learning techniques have been introduced into molecular docking to increase accuracy. One successful example is the AlphaFold-Multimer (Evans et al., 2022), which has significantly increased the accuracy of predicting protein-protein complex structures. However, one is still not able to predict biomolecular binding kinetics from molecular docking.

Molecular Dynamics (MD) is a powerful technique for simulations of biomolecular structural dynamics (Karplus & McCammon, 2002). Remarkable advances in computing hardware (e.g., the Anton supercomputer and GPUs) and software developments have significantly increased the accessible time scale of conventional MD (cMD) from hundreds of nanoseconds to hundreds of microseconds (Harvey et al., 2009; Hollingsworth & Dror, 2018; Johnston & Filizola, 2011; Lane et al., 2013; Shaw et al., 2021; Shaw et al., 2010). Notably, the latest Anton3 (Shaw et al., 2021) has achieved the speed of hundreds of microseconds per day for ATPase and Satellite Tobacco Mosaic Virus (STMV) with total number of atoms ranging from 328 K to 1,067 K, which will significantly facilitate simulations of biomolecular recognition process. The cMD simulations have been widely applied to investigate biomolecular dynamics, including conformational change (Jensen et al., 2012), protein folding (Lindorff-Larsen et al., 2011) and substrate binding (Dror et al., 2013; Robustelli et al., 2020; Shan et al., 2011).

For small-molecule ligand binding, Shan et al. (Shan et al., 2011) observed spontaneous binding of the Dasatinib drug to its target Src kinase during tens of microseconds cMD simulations. However, no dissociation event was observed in the cMD simulations. Pan et al. (Pan et al., 2017) performed tens of microseconds cMD simulations to successfully characterize repetitive binding and dissociation of six small-molecule fragments to the protein FKBP. Based on the large number of binding and dissociation events in the simulations, they were able to accurately calculate the

binding free energies and kinetic rates. Remarkably, the binding free energies calculated from the cMD simulations agreed very well with those predicted from free energy perturbations (FEP) calculations. It is worthy noting that the tested fragments were weak binders with affinities ranging from 200 μM to 20 mM. It is still challenging to simulate both binding and dissociation of typical small-molecule ligands of proteins (usually with higher binding affinities and slow dissociation rates) using cMD, although the ligand residence time (or dissociation rate) has recently been recognized to correlate better with drug efficacy (Schuetz et al., 2017). For protein-protein interactions, tens of microseconds cMD simulations were able to capture barnase binding to barstar (Pan et al., 2019). Accurate barnase binding rate (k_{on}) was predicted based on multiple binding events captured in a total of ~440 μs Anton cMD simulations (Pan et al., 2019). However, it remains challenging to simulate dissociation of the barnase−barstar model system using cMD (Pan et al., 2019).

Weighted Ensemble (Saglam & Chong, 2019) and Markov state model (MSM) (Plattner et al., 2017) have been developed to improve prediction of biomolecular binding thermodynamics and kinetics based on a large number of short cMD trajectories. The kinetic binding rate (k_{on}) of the p53 peptide to the MDM2 protein was accurately predicted with Weighted Ensemble of a total amount of ~120 μs cMD simulations in implicit solvent (Zwier et al., 2016). Another Weighted Ensemble of a total of ~18 μs cMD simulations were able to accurately predict the barnase-barstar binding rate constant (k_{on}) (Saglam & Chong, 2019). However, it is still challenging to model the slow protein/peptide dissociation processes with Weighted Ensemble simulations (Saglam & Chong, 2019; Zwier et al., 2016). MSM (Paul et al., 2017; Plattner et al., 2017; Plattner & Noe, 2015) was able to simultaneously predict the binding and dissociation kinetics through longer aggregated cMD simulations. MSM built with 150 μs MD simulation data was used to accurately

predict benzamidine-trypsin binding kinetics (Plattner & Noe, 2015). Based on a total of two millisecond cMD simulations of barnase binding to barstar, MSM was generated to predict intermediate structures, binding energies and kinetic rates that were consistent with experimental data (Plattner et al., 2017). However, these calculations required very expensive computational resources.

Coarse-grained MD models have been developed to reduce the demand of computational resources and extend simulation time scales (Souza et al., 2021; Souza et al., 2020). Souza et al. performed millisecond cMD simulations to capture the binding of diverse protein-ligand systems (Souza et al., 2020). Accurate binding free energies were predicted through the cMD simulations without *a priori* information (Souza et al., 2020). Millisecond MD simulations with a useful coarse-grained model (PACE) were performed to characterize the binding mechanism of the intrinsically disordered Aβ peptides ($A\beta_{17-42}$) to form Aβ fibril (Han & Schulten, 2014). In addition, coarse-grained models could be incorporated into multiscale computational approaches to improve the efficiency and accuracy of ligand binding thermodynamics and kinetics calculations (Elber, 2020; Huang, 2021; Jagger et al., 2020). For example, simulation enabled estimation of kinetic rates (SEEKR) (Jagger et al., 2020; Votapka & Amaro, 2015) is a multiscale simulation approach combining MD, Brownian dynamics, and milestoning for calculating receptor−ligand binding and dissociation rates. SEEKR has been shown to estimate binding kinetic rates with up to a factor of 10 less simulation time (Jagger et al., 2020).

Enhanced sampling methods have been developed to efficiently simulate biomolecular recognition. They could be generally divided into two categories depending on the usage of collective variables (CVs). The CV-based methods include the widely used Steered MD (Kingsley et al., 2016), Umbrella Sampling (Gumbart et al., 2013b; Joshi & Lin, 2019b; Kingsley et al., 2016),

Metadynamics (Antoszewski et al., 2020; Banerjee & Bagchi, 2020), adaptive biasing force (ABF) (Darve & Pohorille, 2001; Darve et al., 2008), and so on. These methods often use predefined CVs to effectively guide simulations. Thus, *a priori* knowledge of the system is required in CV-based enhanced sampling. Alternatively, when it is difficult to predefine CVs, CV-free enhanced sampling methods could be useful(Kamenik et al., 2022). These methods include Replica Exchange MD (Siebenmorgen & Zacharias, 2020; Sugita et al., 2019; Sugita & Okamoto, 1999), Random Acceleration Molecular Dynamics (RAMD) (Nunes-Alves et al., 2021), Tempered Binding (Pan et al., 2019), Integrated Tempering Sampling (ITS) (Shao & Zhu, 2019; Yang et al., 2015), scaled MD (Deb & Frank, 2019), accelerated MD (aMD) (Hamelberg et al., 2004), Gaussian accelerated MD (GaMD) (Miao et al., 2015b; Wang et al., 2021), and so on. The above-mentioned methodological advances have enabled simulations of millisecond or even longer time scale processes. Here, we will briefly review recent efforts in modeling biomolecular recognition, especially characterization of binding thermodynamics and kinetics.

2. Collective Variable-Based Enhanced Sampling

During CV-based enhanced sampling simulations, a potential or force bias is applied along certain CVs to facilitate energy barrier crossing events among different conformational states. Typical CVs include distances, angle, dihedral, path, eigenvectors generated from the principal component analysis, root-mean square deviation (RMSD) relative to a reference conformation (Bouvier & Grubmuller, 2007), and so on. The bias potential applied to the system is usually around several kcal/mol. Thus one is able to accurately recover the original free energy profiles.

Umbrella Sampling has been applied to predict the ligand/peptide/protein binding and/or dissociation pathways and map the associated free energy landscapes (Gumbart et al., 2013a; Joshi

& Lin, 2019a; Sieker et al., 2008; You et al., 2019). Metadynamics has been applied to investigate ligand/peptide/protein binding in terms of the binding kinetic rates (Casasnovas et al., 2017; Sun et al., 2017) and free energies (Banerjee et al., 2018; Raniolo & Limongelli, 2020; Saleh et al., 2017; Wang et al., 2022a). Metadynamics simulations (Limongelli et al., 2013; Tiwary & Parrinello, 2013) have also been applied to investigate the thermodynamics and kinetics of benzamidine inhibitor binding to trypsin. Multiple Metadynamics trajectories with a total of 5 μs simulations were obtained to predict the ligand unbinding pathways and dissociation rate constant (k_{off}). The predicted k_{off} (9.1 ± 2.5 s^{-1}) was smaller than the experimental value (600 ± 300 s^{-1}). Separate funnel Metadynamics simulations predicted accurate of ligand binding free energies (-8.5 ± 0.7 kcal/mol) for the same system (Limongelli et al., 2013). Infrequent Metadynamics simulations with 3 carefully chosen CVs have successfully predicted the peptide binding and dissociation rates for the system of P53-MDM2 (Zou et al., 2020). Although these methods have shown remarkable improvements in capturing rare events that happen over exceedingly long timescales, users often face a challenge for defining CVs, which requires expert knowledge of the studied systems (Abrams & Bussi, 2014; Zuckerman, 2011). Additionally, the pre-defined CVs could constrain the sampling space, leading to slow convergence of the simulations and suffering from "hidden energy barrier" once important CVs were missed during the simulation setup (Bešker & Gervasio, 2012). To accelerate the convergence of simulations, replica exchange or parallel tempering methods have been incorporated into Metadynamics. For example, bias-exchange Metadynamics simulations with 8 CVs have been performed to predict accurate binding free energy of the p53 peptide to the MDM2 protein. Parallel tempering Metadynamics simulations with well-tempered ensemble (PTMetaD-WTE) successfully captured the binding and dissociation processes of insulin dimer (Antoszewski et al., 2020). In summary, by carefully defining reaction

coordinates, the CV-based enhanced sampling methods could efficiently and accurately predict binding free energies and kinetic rates.

3. Enhanced Sampling without Predefined Collective Variables

In CV-free enhanced sampling methods, bias is often applied on generalized properties of the system (such as the potential energy and atomic forces) in the simulations. Repetitive benzamidine binding and unbinding in trypsin were captured using the selective ITS method (Shao & Zhu, 2019; Yang et al., 2015; Yang & Qin Gao, 2009). Pan et al.(Pan et al., 2019) developed the Tempered Binding method, which significantly accelerates the slow protein dissociation process by dynamically adjusting electrostatic and van der Waals interactions between different groups of protein atoms by a factor λ. The tempered binding simulations have successfully captured repetitive binding and dissociation events for five diverse protein–protein systems (Pan et al., 2019). In the scaled MD simulations (Sinko et al., 2013), a scale factor ranging from 0 to 1 is introduced to smoothen the potential energy surface. Schuetz et al. performed scaled MD simulations to accurately predict the residence time and drug dissociation pathways of different inhibitors of heat shock protein 90 (Hsp90) (Schuetz et al., 2018). In a recent study (Bianciotto et al., 2021), Bianciotto et al. used scaled MD simulations to predict the residence time and ligand unbinding pathways for a set of 27 ligands of Hsp90 protein, being highly consistent with experimental data. Deb et al. developed a selective scaled MD simulation method (Deb & Frank, 2019), where specific energy terms are scaled to promote dissociation of bound ligands from the protein. Particularly, ligand-water interactions are scaled to help the ligands dissociate from its bound state. Selective scaled MD predict accurate residence times and associated free energy change of three inhibitor drugs bound to cyclin-dependent kinase protein complexes. Hence,

selective scaled MD proves to be an important enhanced sampling method for modeling biomolecular dissociation process.

In RAMD, an additional random force is applied on the ligand to promote especially the dissociation. In one recent study, Nunes-Alves et al. performed RAMD simulations to predict ligand dissociation rates of T4 lysozyme (Nunes-Alves et al., 2021). The predicted kinetic rates agreed well with experimental values for various systems with different ligands, temperatures, and protein mutations. Moreover, a ligand with complex dissociation pathways was often associated with longer residence time. In another study, the same group (Kokh & Wade, 2021) performed RAMD simulations to explore ligand dissociation pathways and kinetics of two GPCRs, i.e., the β_2 adrenergic receptor (β_2AR) and M2 muscarinic acetylcholine receptor (M2R). The ligand dissociation pathways observed in the RAMD simulations were similar to those in long cMD and Metadynamics simulations. Additionally, RAMD revealed an allosteric modulation mechanism of the LY2119620 PAM in the M2R. Dissociation of the iperoxo agonist was blocked from one of the possible pathways and hence had increased residence time, being consistent with the experimental data.

The aMD enhanced sampling technique works by adding a non-negative boost potential to smoothen the system potential energy surface (Hamelberg et al., 2004; Voter, 1997). The boost potential (ΔV) decreases the energy barrier to facilitate the system cross different conformational states (Hamelberg et al., 2007; Hamelberg et al., 2004). In one study (Kappel et al., 2015), Kappel et al. performed aMD simulations to study ligand binding to M3 muscarinic receptor (M3R). Three ligands of the receptor: full agonist Ach, partial agonist arecoline (Arc) and antagonist tiotropium (TTP) were used to perform the aMD simulations. Starting from the bulk solvent, aMD captured the binding of Ach to the M3R orthosteric site in significantly less time as compared to the cMD

simulations. The Arc was also observed binding to the orthosteric site whereas the TTP molecule bound to the extracellular vestibule of the receptor. Moreover, all ligands could bind to the extracellular vestibule of the receptor, suggesting the vestibule as metastable binding site for orthosteric ligands. However, aMD suffers from large energetic noise during reweighting as the boost potential is typically on the order of tens to hundreds of kcal/mol (Shen & Hamelberg, 2008).

GaMD is developed to apply a harmonic boost potential to enhance sampling with significantly reduced energetic noise. The boost potential normally exhibits a near Gaussian distribution, which enables proper reweighting of the free energy profiles through cumulant expansion to the second order (Miao et al., 2015b; Wang et al., 2021). GaMD has been successfully applied to simulate important biomolecular processes, including ligand/protein/RNA binding (Chuang et al., 2018; Liao & Wang, 2019; Miao et al., 2015a; Miao et al., 2018b; Miao & McCammon, 2016; Pang et al., 2017; Wang et al., 2022b; Wang & Chan, 2017), protein folding (Miao et al., 2015a; Pang et al., 2017), and protein conformational changes (Miao & McCammon, 2016; Salawu, 2018; Zhang et al., 2018). However, it remained challenging to simulate repetitive substrate binding and dissociation through normal GaMD (Miao & McCammon, 2018; Wang et al., 2021).

Recently, "selective GaMD" algorithms have been developed to allow for more efficient enhanced sampling of biomolecular binding and dissociation processes, including the Ligand GaMD (LiGaMD) (Miao et al., 2020), Peptide GaMD (Pep-GaMD) (Wang & Miao, 2020) and Protein-Protein Interaction - GaMD (PPI-GaMD) (Wang & Miao, 2022). For simulations of biomolecular binding, the system contains substrate L (e.g. small molecule ligands, peptides or ligand protein), protein *P* and the biological environment *E.* Therefore*,* the potential energy of system could be decomposed into the following terms: $V(r) = V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) +$

$V_{PP,nb}(r_P) + V_{LL,nb}(r_L) + V_{EE,nb}(r_E) + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE})$, where $V_{P,b}$, $V_{L,b}$ and $V_{E,b}$ are the bonded potential energies in protein *P*, substrate *L* and environment *E*, respectively. $V_{PP,nb}$, $V_{LL,nb}$ and $V_{EE,nb}$ are the self non-bonded potential energies in protein *P*, substrate *L* and environment *E*, respectively. $V_{PL,nb}$, $V_{PE,nb}$ and $V_{LE,nb}$ are the non-bonded interaction energies between *P-L*, *P-E* and *L-E*, respectively. In order to facilitate the ligand/peptide/protein binding (**Fig 1**), a boost potential is selectively added on the essential energy terms ($V_{select}(r)$) in the LiGaMD, Pep-GaMD and PPI-GaMD, respectively. Presumably, ligand binding mainly involves the non-bonded interaction energies of the ligand. LiGaMD thus selectively boosts on the energy terms of $V_{select}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. In comparison, peptide binding involves in both the bonded and non-bonded interaction energies of the peptide since peptides often undergo large conformational changes during binding to the target proteins. Thus, the essential energy term in Pep-GaMD is $V_{select}(r) = V_{LL,b}(r_L) + V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. While protein-protein binding and unbinding processes mainly involve the non-bonded interaction energies between protein partners, one can apply a selective boost to the essential energy term $V_{select}(r) = V_{PL,nb}$ in PPI-GaMD. In addition to selectively boost the essential energy term $V_{select}(r)$, another boost potential could be applied on the remaining energy of the system to facilitate substrate rebinding in a dual-boost scheme. These new algorithms have been implemented in the GPU version of AMBER22 (D.A. Case).

Repetitive binding and dissociation of small-molecule ligands were captured in the LiGaMD simulations of host-guest and protein-ligand binding model systems (Miao et al., 2020), which enabled us to calculate ligand binding thermodynamics and kinetics calculations. Repetitive guest binding and dissociation in the β-cyclodextrin host were observed in hundreds-of-nanoseconds LiGaMD simulations. The binding free energies of guest molecules predicted from

LiGaMD simulations agreed excellently with experimental data (< 1.0 kcal/mol error). In comparison with previous microsecond-timescale cMD simulations, accelerations of ligand kinetic rate constants in LiGaMD simulations were properly estimated using Kramers' rate theory. Furthermore, microsecond LiGaMD simulations observed repetitive benzamidine binding and dissociation in trypsin. Trypsin-benzamidine ligand binding free energy was calculated from the 3D PMF profile to be -6.13 ± 0.35 kcal/mol, being highly consistent with the experimental value of -6.2 kcal/mol (Guillain & Thusius, 1970). Similarly, the ligand binding and dissociation time periods were recorded to calculate the reweighted k_{on} and k_{off} values to be $1.15 \pm 0.79 \times 10^7$ $M^{-1} \cdot s^{-1}$ and 3.53 ± 1.41 s^{-1}, respectively. These data were comparable to the values calculated from experiments (Guillain & Thusius, 1970).

Pep-GaMD (Wang & Miao, 2020) has been demonstrated on binding of three model peptides to the SH3 domains (Ahmad & Helms, 2009; Ball et al., 2005), including "PAMPAR" (PDB: 1SSH), "PPPALPPKK" (PDB: 1CKA) and "PPPVPPRR" (PDB: 1CKB). Repetitive dissociation and binding of the three peptides were successfully captured in each of the 1 microsecond Pep-GaMD simulations. The peptide binding free energies calculated from Pep-GaMD simulations were in excellent agreements with those from the experiments. For the 1CKA system, the calculated peptide binding free energy value was -7.72±0.54 kcal/mol, being highly consistent with the experimental value of -7.84 kcal/mol (Wu et al., 1995). For the 1CKB system, the predicting binding free energy was -6.84±0.14 kcal/mol, being closely similar to the experimental value of -7.24 kcal/mol (Wu et al., 1995). In addition, the Pep-GaMD predicted the k_{on} and k_{off} of 1CKA as $4.06 \pm 2.26 \times 10^{10}$ $M^{-1} \cdot s^{-1}$ and $1.45 \pm 1.07 \times 10^3$ s^{-1}, respectively. They were comparable to the experimental data (Xue et al., 2014) of $k_{on}^{exp} = 1.5 \times 10^9$ $M^{-1} \cdot s^{-1}$ and $k_{off}^{exp} = 8.9 \times 10^3$ s^{-1}.

More recently, Pep-GaMD simulations were combined with complementary biochemical experiments to elucidate mechanism of tripeptide trimming of amyloid β-peptide (Aβ peptide) by γ-secretase(Bhattarai et al., 2022). The active model of γ-secretase for ε cleavage was extracted from previous study (Bhattarai et al., 2020) and used as the starting structure for Pep-GaMD simulations. 600 ns Pep-GaMD simulations were able to capture the ζ cleavage activation starting from the ε cleavage activated model, which was suggested to carry out in timescale of minutes (Kamp et al., 2015). During activation, coordinated hydrogen bonds were formed between carbonyl oxygen of Aβ49 Val46 and enzyme catalytic Asp257. The two catalytic aspartates, Asp257 and Asp385 in the active site of the enzyme both formed hydrogen bonds with the water molecule aligned in between them. This activated enzyme conformation was well oriented for the ζ cleavage of amide between Val46 and Ile47 of the Aβ49. Three low energy states including "Final", "Intermediate", and "Initial" were identified from the Pep-GaMD simulations **(Fig. 2A)**. The Final state denoted the activated enzyme conformation for ζ cleavage where the Asp257 - Asp385 distance was ~7 – 8 Å and the Asp257 – Aβ49 Val46 distance was ~ 3 Å (hydrogen bond). The Initial and Intermediate low energy states denoted the starting and transitional conformation during the activation process. Furthermore, Pep-GaMD simulations were performed for three additional FAD mutant Aβ49 bound enzyme systems. Similar to the wildtype system, Pep-GaMD simulations of I45F, A42T and V46F mutant Aβ49 bound enzyme systems were able to capture the ζ cleavage activation starting from the ε cleavage activated model. Free energy profiles of the FAD mutant systems were similar to the wildtype system **(Fig. 2B-2D).** In the I45F mutant system, two low energy states were identified including "Initial" and "Final" **(Fig. 2B)**. The A42T mutant was the most dynamic enzyme system with four distinct low energy states identified in a larger area covered free energy profile including "Initial", "Final", "Inhibited-1", and "Inactive" **(Fig.**

2C). The catalytic aspartates of the "Inhibited-1" conformational state were too close for activation and hence was inhibited. In contrast, the aspartates were too far for their catalytic activity in the "Inactive" low energy state of the enzyme. In the V46F mutant γ-secretase system, two low energy states were identified in the free energy profile including "Final" and "Inhibited-2" **(Fig. 2D)**. The structures were compared between the "Initial" and "Final" low energy conformational states of the enzyme as identified from the free energy profiles **(Figs. 2E-2G)**. The enzyme moved from Initial to Final conformational state, the Aβ49 substrate tilted by ~ 50° **(Fig. 2F)**. Unwinding of helix was observed in the C-terminus of Aβ49 where residues Val44 and Ile45 were observed changing their conformation from helix to a loop **(Fig. 2F)**. Similarly, in the active site of the enzyme, the protonated Asp257 in the Final state was observed moving forward towards the substrate scissile amide bond by 3 Å in comparison to the Initial state **(Fig. 2G)**. In contrast, the deprotonated Asp385 in the Final state and the Initial state were observed in a similar conformation **(Fig. 2G)**. The simulation findings were highly consistent with biochemical experimental data. Taken together, complementary biochemical experiments and Pep-GaMD simulations have enabled elucidation of the mechanism of tripeptide trimming of Aβ49 by γ-secretase.

PPI-GaMD (Wang & Miao, 2020) has been demonstrated on a model system of the ribonuclease barnase binding to barstar. Six independent 2 μs PPI-GaMD simulations have successfully captured repetitive barstar dissociation and rebinding events (**Fig. 3A**). Five binding and six dissociation events were observed in both Sim1 and Sim3. In Sim2, three binding and four dissociation events were captured. For the remaining simulations (Sim4 - Sim6), three binding and three dissociation events were observed (**Fig. 3A**). The barstar binding free energy predicted from PPI-GaMD was -17.79 kcal/mol with a standard deviation of 1.11 kcal/mol, being highly consistent with the experimental value of -18.90 kcal/mol (Schreiber & Fersht, 1993). In addition,

the PPI-GaMD simulations allowed us to calculate the protein binding kinetics. The average reweighted k_{on} and k_{off} were predicted as $21.7\pm13.8\times10^{8}$ $M^{-1}\cdot s^{-1}$ and $7.32\pm4.95\times10^{-6}$ s^{-1}, being highly consistent with the corresponding experimental values of 6.0×10^{8} $M^{-1}\cdot s^{-1}$ and 8.0×10^{-6} s^{-1}, respectively. Furthermore, PPI-GaMD simulations have provided mechanistic insights into barstar binding to barnase, which involve long-range electrostatic interactions and multiple binding pathways (**Fig. 3C-3F**), being consistent with previous experimental and computational findings of this model system. It is worth noting that at least 3 independent replicas of selective GaMD simulations with longer simulation lengths (e.g., microsecond) are required to obtain sufficient statistics for ligand binding, peptide binding and protein-protein interactions. In order to calculate accurate binding free energy and kinetic rates, the length of each simulation should be long enough to capture ≥3 binding and dissociation events as suggested by LiGaMD(Miao et al., 2020), Pep-GaMD(Wang & Miao, 2020) and PPI GaMD(Wang & Miao, 2022) studies.

4. Machine Learning

Machine learning (ML) has been applied to improve computational docking, especially in the scoring functions (Khamis et al., 2015). A scoring function in molecular docking refers to a mathematical predictive model that outputs a representative score of the binding free energy of a bound conformation. Scoring of a docked complex is the final step of the three essential components in molecular docking, with the first two being chemical molecule representation and pose generation (Khamis et al., 2015). A reliable scoring function should have a good scoring power (the ability to produce scores for different binding poses), ranking power (the ability to correctly rank a given set of ligands with known binding poses when bound to a common protein), and docking power (the ability to identify the best binding pose of a given ligand from a set of

computationally generated poses when bound to a specific protein) (Ashtaway & Mahapatra, 2012). Kinnings et al. (Kinnings et al., 2011) used a support vector machine (SVM) to derive a unique set of weights for each individual protein family – the w_i*'s* in the following equation:

$$\Delta G_{binding} = w_0 + w_1 \Delta G_{VdW} + w_2 \Delta G_{h-bond} + w_3 \Delta G_{rotor} + w_4 \Delta G_{hydrophobic} \quad (1)$$

This was shown to improve the binding affinity prediction of the electronic high throughput screening (eHiTS) molecular docking software(Zsoldos et al., 2007) compared with empirical knowledge-based scoring functions(Khamis et al., 2015). Similarly, a force field scoring function can be trained to derive a unique set of parameters for each individual protein family - the A_{ij}'s and B_{ij}'s in the following equation:

$$E_{binding} = \sum_{i=1}^{ligand} \sum_{j=1}^{protein} \left(\frac{A_{ij}}{r_{ij}^{a}} - \frac{B_{ij}}{r_{ij}^{b}} + 332 \frac{q_i q_j}{D r_{ij}} \right) \quad (2)$$

ML could also be used to predict the binding affinity based on a number of features of the protein-ligand complex, including geometric features, physical force field energy terms, pharmacophore features, etc. Specifically, ML could learn the relationship between these features and corresponding known binding affinity to predict the binding affinity of new complexes(Khamis et al., 2015). Recently, Ballester et al. (Ballester & Mitchell, 2010) applied non-parametric ML techniques to generate the functional form of scoring functions given molecular databases. The authors used random forest (RF) (Breiman, 2001) to learn the relationship between the atomic-level description of the complex and the experimental binding affinity. Here, the K_d and K_i measurements were merged into a single binding constant K to represent the experimental binding affinity. The atomic-level description used was of geometric nature and was the occurrence count of 9 common elemental atoms (*C, N, O, F, P, S, Cl, Br, I*) type pair. Even though they completely neglected the energy terms induced by protein-ligand

interactions, Ballester et al. were able to achieve Pearson correlation coefficient of 0.774 on the PDBbind v2007 core set (195 complexes) (Ballester & Mitchell, 2010).

Very recently, deep learning (DL) methods, including RoseTTAFold(Baek et al., 2021) and AlphaFold (Jumper et al., 2021), were developed to achieve structure prediction accuracies far beyond those from classical force-field-based methods (Baek & Baker, 2022). These methods have millions of parameters, much more than the hundreds of parameters in classical approaches, thus better sample the large conformational space of proteins. Furthermore, they make no assumptions about the functional form of the interactions between atoms. In fact, the two DL-based methods learn millions of parameters directly to generate correct 3D structures from input amino acid sequences (Baek & Baker, 2022; Baek et al., 2021; Jumper et al., 2021). AlphaFold and RoseTTAFold are trained to predict structures from alignments of homologous amino acid sequences. In particular, the two DL-based approaches learn to extract rich structural information through a three-track network where information at the 1D sequence level, 2D distance map, and 3D coordinate level is successively transformed and integrated (Baek et al., 2021; Jumper et al., 2021). They were also shown to predict protein structures very accurately from single amino acid sequences (Baek & Baker, 2022; Baek et al., 2021; Jumper et al., 2021).

MD simulations could generate very large data in terms of conformation frames and number of simulated atoms. For example, Weighted Ensemble of the COVID19 spike protein's closed-to-open state generated over 100 terabytes of data (Casalino et al., 2021). This brings a challenge to identify proper CVs to differentiate conformational states from the raw simulation data and to identify corresponding biologically transitions between such states (e.g., open/closed states of spike). In this regard, the machine learning/deep learning has been applied to identify appropriate CV to analysis MD simulation trajectories (Glielmo et al., 2021; Noé, 2020; Sun et al.,

2022; Wang et al., 2020). These linear, non-linear, and hybrid Machine Learning approaches cluster the simulation data along a small number of latent dimensions to identify conformational transitions between states (Bernetti et al., 2020; Ramanathan et al., 2012). Another benefit of MD-coupled machine learning approaches is that the information learned from machine learning can be used to iteratively guide the MD sampling(Wang et al., 2019). Based on the predictive information bottleneck, Wang et al. developed an approach to identify system reaction coordinates and computer the free energy and kinetic rates in biomolecules (Wang et al., 2019). The algorithm was demonstrated on conformational transitions in the alanine dipeptide model system and ligand dissociation from the L99A T4lysome. Thermodynamic and kinetic quantities calculated from short enhanced MD simulations for slow biomolecular processes were in good agreement with the experiments and long unbiased MD simulations.

Recently, we have integrated the GaMD, Deep Learning and free energy prOfiling Workflow (GLOW) to predict important reaction coordinates and map free energy profiles of biomolecules (Do et al., 2022). First, GaMD simulations are performed on the target biomolecules (**Fig. 4A**). The residue contact map is then calculated for each GaMD simulation frame and transformed into images (**Fig. 4B**). The specialized type of neural network for image classification, two-dimensional (2D) convolutional neural network (CNN), is employed to classify the residue contact maps of target biomolecules, from which important residue contacts are identified by classic gradient-based pixel attribution (**Fig. 4C**). Finally, the free energy profiles of these reaction coordinates are calculated through reweighting of GaMD simulations to characterize the biomolecular systems of interest (**Fig. 4D**)(Do et al., 2022). GLOW was successfully demonstrated on characterization of activation and allosteric modulation of a GPCR, using the adenosine A_1 receptor (A_1AR) as a model system. Characterization of the A_1AR activation was achieved by

classification of the A_1AR bound by "Antagonist", "Agonist", and "Agonist-Gi". GLOW achieved an overall accuracy of 99.34% and loss of 1.85%, respectively, on the validation data set after 15 epochs. Meanwhile, characterization of A_1AR allosteric modulation was achieved by classification of the A_1AR bound by "Agonist-Gi" and "Agonist-Gi-PAM". GLOW achieved an overall accuracy of 99.27% and loss of 1.78%, respectively, on the validation data set after 15 epochs. GLOW identified characteristic residue contacts that were highly consistent with previous studies to the residue levels for both A_1AR activation and allosteric modulation. In particular, the ligand-binding extracellular domains (ECL1-ECL3) and intracellular G-protein binding domains (TM3, TM5, TM6, and TM7) were found to be loosely coupled in the GPCR activation. Furthermore, it showed that ECL2 played a critical role in the allosteric modulation of A_1AR, being consistent with previous mutagenesis, structure, and molecule modeling studies (Avlani et al., 2007; Draper-Joyce et al., 2021; Miao et al., 2018a; Nguyen et al., 2016; Peeters et al., 2012). GLOW revealed that binding of a PAM (MIPS521) to the agonist-Gi-A_1AR complex biased the receptor conformational ensemble, especially in the ECL1 and ECL2 regions. PAM binding stabilized agonist binding within the orthosteric pocket of A_1AR, which confined the extracellular mouth of the receptor Furthermore, PAM binding disrupted the $N148^{ECL2}$-$V152^{ECL2}$ α-helical hydrogen bond and distorted this portion of the ECL2 helix(Do et al., 2022).

In addition, DL has been widely applied to optimize force field (Chatterjee et al., 2022; Poltavsky & Tkatchenko, 2021; Unke et al., 2021), binding free energy calculations (Chen et al., 2021; Jiang et al., 2021; Jones et al., 2021) and binding pathway identification (Motta et al., 2022).

5. CONCLUSIONS AND OUTLOOK

With remarkable advances in both computer hardware and software, computational approaches have achieved significant improvement to characterize biomolecular recognition, including molecular docking, MD simulations and Machine Learning. Machine learning has been incorporated into both molecular docking and MD simulations to improve the docking accuracy, simulation efficiency and trajectory analysis, e.g., AlphaFold-Multimer and GLOW. MD simulations have enabled characterization of biomolecular binding thermodynamics and kinetics, attracting increasing attention in recent years. Long time scale cMD simulations have successfully captured biomolecular binding processes, although slow dissociation of biomolecules are still often difficult to simulate using cMD.

Enhanced sampling methods have greatly reduced the computational cost for calculations of biomolecular binding thermodynamics and kinetics. Higher sampling efficiency could be generally obtained using the CV-based methods than using the CV-free methods. However, CV-based enhanced sampling methods require predefined CVs, which is often challenging for simulations of complex biological systems. Nevertheless, machine learning techniques have proven useful to identify proper CVs or reaction coordinates. Alternatively, CV-free methods are usually easy to use without requirement of *a priori* knowledge of the studied systems. Additionally, the CV-based and CV-free methods could be combined to be more powerful. The CV-free methods can enhance the sampling to potentially overcome the hidden energy barriers in orthogonal degrees of freedom relative to the CVs predefined in the CV-based methods, which could enable faster convergence of the MD simulations. Newly developed algorithms in this direction include integration of replica exchange umbrella sampling with GaMD (GaREUS) (Oshima et al., 2019b), replica exchange of solute tempering with umbrella sampling (gREST/REUS) (Kamiya & Sugita, 2018; Re et al., 2019), replica exchange of solute tempering with well-tempered Metadynamics

(ST-MetaD)(Mlýnský et al., 2022) and temperature accelerated molecular dynamics (TAMD) with integrated tempering sampling (ITS/TAMD) (Xie et al., 2017).

Recent years have seen an increasing number of techniques that introduce "selective" boost in the CV-free enhanced sampling methods, including the selective ITS, selective scaled MD, selective aMD and selective LiGaMD, Pep-GaMD and PPI-GaMD. In these methods, only essential energy terms are selectively boosted to further increase the sampling efficiency. Additionally, compatible enhanced sampling methods could be combined to be more powerful. For example, GaMD has been combined with Umbrella Sampling to achieve significantly improved efficiency (Oshima et al., 2019a; Wang et al., 2021). Besides enhanced sampling, the accuracy of force fields and water models play a critical role in predicting the biomolecular binding affinities and kinetics. For example, the TIP4P2015 water model was shown to be more accurate than the TIP3P water model in calculating the kinetics of barnase-barstar binding in cMD simulations(Pan et al., 2019). Nevertheless, biomolecular recognition in systems of increasing sizes (such as viruses and cells) and accurate calculations of binding thermodynamics and kinetics of large biomolecular complexes present grand challenges for computational modelling and enhanced sampling simulations. Further innovations in both computing hardware and method developments may help us to address these challenges in the future.

Acknowledgements

This work used supercomputing resources with allocation awards TG-MCB180049 and BIO210039 through the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1548562 and project M2874 through the National Energy Research Scientific Computing Center (NERSC), which is a

U.S. Department of Energy Office of Science User Facility operated under Contract No. DE-AC02-05CH11231. It also used computational resources provided by the Research Computing Cluster at the University of Kansas.This work was supported in part by the National Institutes of Health (R01GM132572), National Science Foundation (2121063), and the startup funding in the College of Liberal Arts and Sciences at the University of Kansas.

References

ABRAMS, C. & BUSSI, G. (2014). Enhanced Sampling in Molecular Dynamics Using Metadynamics, Replica-Exchange, and Temperature-Acceleration. Entropy, 16(1), 163-199.

AHMAD, M. & HELMS, V. (2009). How do proteins associate? A lesson from SH3 domain. Chem Cent J, 3(S1), O22.

ANTOSZEWSKI, A., FENG, C.-J., VANI, B. P., THIEDE, E. H., HONG, L., WEARE, J., TOKMAKOFF, A. & DINNER, A. R. (2020). Insulin dissociates by diverse mechanisms of coupled unfolding and unbinding. The Journal of Physical Chemistry B, 124(27), 5571-5587.

ASHTAWAY, H. M. & MAHAPATRA, N. R. (2012). A comparative assessment of ranking accuracies of conventional and machine-learning-based scoring functions for protein-ligand binding affinity prediction. IEEE/ACM Trans Comput Biol Bioinform, 9(5), 1301-1313.

AVLANI, V. A., GREGORY, K. J., MORTON, C. J., PARKER, M. W., SEXTON, P. M. & CHRISTOPOULOS, A. (2007). Critical role for the second extracellular loop in the binding of both orthosteric and allosteric G protein-coupled receptor ligands. Journal of Biological Chemistry, 282(35), 25677-25686.

BAEK, M. & BAKER, D. (2022). Deep learning and protein structure modeling. Nature Methods, 19(1), 13-14.

BAEK, M., DIMAIO, F., ANISHCHENKO, I., DAUPARAS, J., OVCHINNIKOV, S., LEE, G. R., WANG, J., CONG, Q., KINCH, L. N. & SCHAEFFER, R. D. (2021). Accurate prediction of protein structures and interactions using a three-track neural network. Science, 373(6557), 871-876.

BALL, L. J., KUHNE, R., SCHNEIDER-MERGENER, J. & OSCHKINAT, H. (2005). Recognition of proline-rich motifs by protein-protein-interaction domains. Angew Chem Int Ed Engl, 44(19), 2852-2869.

BALLESTER, P. J. & MITCHELL, J. B. (2010). A machine learning approach to predicting protein–ligand binding affinity with applications to molecular docking. Bioinformatics, 26(9), 1169-1175.

BANERJEE, P. & BAGCHI, B. (2020). Dynamical control by water at a molecular level in protein dimer association and dissociation. Proceedings of the National Academy of Sciences, 117(5), 2302-2308.

BANERJEE, P., MONDAL, S. & BAGCHI, B. (2018). Insulin dimer dissociation in aqueous solution: A computational study of free energy landscape and evolving microscopic structure along the reaction pathway. The Journal of Chemical Physics, 149(11), 114902.

BASDEVANT, N., BORGIS, D. & HA-DUONG, T. (2013). Modeling protein–protein recognition in solution using the coarse-grained force field SCORPION. Journal of Chemical Theory and Computation, 9(1), 803-813.

BERNETTI, M., BERTAZZO, M. & MASETTI, M. (2020). Data-driven molecular dynamics: a multifaceted challenge. Pharmaceuticals, 13(9), 253.

BEŠKER, N. & GERVASIO, F. L. (2012). Using metadynamics and path collective variables to study ligand binding and induced conformational transitions. In Computational Drug Discovery and Design, pp. 501-513: Springer.

BHATTARAI, A., DEVKOTA, S., BHATTARAI, S., WOLFE, M. S. & MIAO, Y. (2020). Mechanisms of γ-secretase activation and substrate processing. ACS central science, 6(6), 969-983.
BHATTARAI, A., DEVKOTA, S., DO, H. N., WANG, J., BHATTARAI, S., WOLFE, M. S. & MIAO, Y. (2022). Mechanism of Tripeptide Trimming of Amyloid β-Peptide 49 by γ-Secretase. Journal of the American Chemical Society.
BIANCIOTTO, M., GKEKA, P., KOKH, D. B., WADE, R. C. & MINOUX, H. (2021). Contact Map Fingerprints of Protein–Ligand Unbinding Trajectories Reveal Mechanisms Determining Residence Times Computed from Scaled Molecular Dynamics. Journal of Chemical Theory and Computation, 17(10), 6522-6535.
BOUVIER, B. & GRUBMULLER, H. (2007). Molecular dynamics study of slow base flipping in DNA using conformational flooding. Biophysical Journal, 93(3), 770-786.
BREIMAN, L. (2001). Random forests. Machine learning, 45(1), 5-32.
CASALINO, L., DOMMER, A. C., GAIEB, Z., BARROS, E. P., SZTAIN, T., AHN, S.-H., TRIFAN, A., BRACE, A., BOGETTI, A. T., CLYDE, A., MA, H., LEE, H., TURILLI, M., KHALID, S., CHONG, L. T., SIMMERLING, C., HARDY, D. J., MAIA, J. D., PHILLIPS, J. C., KURTH, T., STERN, A. C., HUANG, L., MCCALPIN, J. D., TATINENI, M., GIBBS, T., STONE, J. E., JHA, S., RAMANATHAN, A. & AMARO, R. E. (2021). AI-driven multiscale simulations illuminate mechanisms of SARS-CoV-2 spike dynamics. The International Journal of High Performance Computing Applications, 35(5), 432-451.
CASASNOVAS, R., LIMONGELLI, V., TIWARY, P., CARLONI, P. & PARRINELLO, M. (2017). Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. Journal of the American Chemical Society, 139(13), 4780-4788.
CHATTERJEE, P., SENGUL, M. Y., KUMAR, A. & MACKERELL, A. D. (2022). Harnessing Deep Learning for Optimization of Lennard-Jones Parameters for the Polarizable Classical Drude Oscillator Force Field. Journal of Chemical Theory and Computation, 18(4), 2388-2407.
CHEN, H., LIU, H., FENG, H., FU, H., CAI, W., SHAO, X. & CHIPOT, C. (2021). MLCV: Bridging Machine-Learning-Based Dimensionality Reduction and Free-Energy Calculation. Journal of Chemical Information and Modeling.
CHUANG, C. H., CHIOU, S. J., CHENG, T. L. & WANG, Y. T. (2018). A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. Sci Rep, 8(1), 6336.
CIEMNY, M., KURCINSKI, M., KAMEL, K., KOLINSKI, A., ALAM, N., SCHUELER-FURMAN, O. & KMIECIK, S. (2018). Protein-peptide docking: opportunities and challenges. Drug Discov. Today, 23(8), 1530-1537.
D.A. CASE, D. S. C., T.E. CHEATHAM, III, T.A. DARDEN, R.E. DUKE, T.J. GIESE, H. GOHLKE, A.W. GOETZ, D. GREENE, N. HOMEYER, S. IZADI, A. KOVALENKO, T.S. LEE, S. LEGRAND, P. LI, C. LIN, J. LIU, T. LUCHKO, R. LUO, D. MERMELSTEIN, K.M. MERZ, G. MONARD, H. NGUYEN, I. OMELYAN, A. ONUFRIEV, F. PAN, R. QI, D.R. ROE, A. ROITBERG, C. SAGUI, C.L. SIMMERLING, W.M. BOTELLO-SMITH, J. SWAILS, R.C. WALKER, J. WANG, J. WANG, R.M. WOLF, X. WU, L. XIAO, D.M. YORK AND P.A. KOLLMAN (2022), AMBER 2022, UNIVERSITY OF CALIFORNIA, SAN FRANCISCO.

DARVE, E. & POHORILLE, A. (2001). Calculating free energies using average force. The Journal of Chemical Physics, 115(20), 9169-9183.
DARVE, E., RODRÍGUEZ-GÓMEZ, D. & POHORILLE, A. (2008). Adaptive biasing force method for scalar and vector free energy calculations. J Chem Phys, 128(14), 144120.
DEB, I. & FRANK, A. T. (2019). Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. Journal of Chemical Theory and Computation, 15(11), 5817-5828.
DO, H. N., WANG, J., BHATTARAI, A. & MIAO, Y. (2022). GLOW: A Workflow Integrating Gaussian-Accelerated Molecular Dynamics and Deep Learning for Free Energy Profiling. Journal of Chemical Theory and Computation, 18(3), 1423-1436.
DRAPER-JOYCE, C. J., BHOLA, R., WANG, J., BHATTARAI, A., NGUYEN, A. T., O'SULLIVAN, K., CHIA, L. Y., VENUGOPAL, H., VALANT, C. & THAL, D. M. (2021). Positive allosteric mechanisms of adenosine A1 receptor-mediated analgesia. Nature, 597(7877), 571-576.
DROR, R. O., GREEN, H. F., VALANT, C., BORHANI, D. W., VALCOURT, J. R., PAN, A. C., ARLOW, D. H., CANALS, M., LANE, J. R., RAHMANI, R., BAELL, J. B., SEXTON, P. M., CHRISTOPOULOS, A. & SHAW, D. E. (2013). Structural basis for modulation of a G-protein-coupled receptor by allosteric drugs. Nature, 503, 295.
ELBER, R. (2020). Milestoning: An Efficient Approach for Atomically Detailed Simulations of Kinetics in Biophysics. Annual Review of Biophysics, 49(1), 69-85.
ERMAK, D. L. & MCCAMMON, J. A. (1978). Brownian dynamics with hydrodynamic interactions. The Journal of Chemical Physics, 69(4), 1352-1360.
EVANS, R., O'NEILL, M., PRITZEL, A., ANTROPOVA, N., SENIOR, A., GREEN, T., ŽÍDEK, A., BATES, R., BLACKWELL, S., YIM, J., RONNEBERGER, O., BODENSTEIN, S., ZIELINSKI, M., BRIDGLAND, A., POTAPENKO, A., COWIE, A., TUNYASUVUNAKOOL, K., JAIN, R., CLANCY, E., KOHLI, P., JUMPER, J. & HASSABIS, D. (2022). Protein complex prediction with AlphaFold-Multimer. bioRxiv, 2021.2010.2004.463034.
FERREIRA, L. G., OLIVA, G. & ANDRICOPULO, A. D. (2016). Protein-protein interaction inhibitors: advances in anticancer drug design. Expert opinion on drug discovery, 11(10), 957-968.
GABDOULLINE, R. R. & WADE, R. C. (2001). Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. Journal of molecular biology, 306(5), 1139-1155.
GLIELMO, A., HUSIC, B. E., RODRIGUEZ, A., CLEMENTI, C., NOÉ, F. & LAIO, A. (2021). Unsupervised Learning Methods for Molecular Simulation Data. Chemical Reviews.
GUILLAIN, F. & THUSIUS, D. (1970). Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. Journal of the American Chemical Society, 92(18), 5534-5536.
GUMBART, J. C., ROUX, B. & CHIPOT, C. (2013a). Efficient determination of protein-protein standard binding free energies from first principles. J. Chem. Theory Comput., 9(8), 3789-3798.
GUMBART, J. C., ROUX, B. & CHIPOT, C. (2013b). Efficient determination of protein–protein standard binding free energies from first principles. Journal of Chemical Theory and Computation, 9(8), 3789-3798.

HAMELBERG, D., DE OLIVEIRA, C. A. F. & MCCAMMON, J. A. (2007). Sampling of slow diffusive conformational transitions with accelerated molecular dynamics. J Chem Phys, 127(15), 10B614.
HAMELBERG, D., MONGAN, J. & MCCAMMON, J. A. (2004). Accelerated molecular dynamics: a promising and efficient simulation method for biomolecules. J Chem Phys, 120(24), 11919-11929.
HAN, W. & SCHULTEN, K. (2014). Fibril Elongation by Aβ17–42: Kinetic Network Analysis of Hybrid-Resolution Molecular Dynamics Simulations. Journal of the American Chemical Society, 136(35), 12450-12460.
HARVEY, M. J., GIUPPONI, G. & FABRITIIS, G. D. (2009). ACEMD: Accelerating Biomolecular Dynamics in the Microsecond Time Scale. J Chem Theory Comput, 5(6), 1632-1639.
HE, Z., PAUL, F. & ROUX, B. (2021). A critical perspective on Markov state model treatments of protein–protein association using coarse-grained simulations. The Journal of Chemical Physics, 154(8), 084101.
HOLLINGSWORTH, S. A. & DROR, R. O. (2018). Molecular Dynamics Simulation for All. Neuron, 99(6), 1129-1143.
HUANG, Y.-M. M. (2021). Multiscale computational study of ligand binding pathways: Case of p38 MAP kinase and its inhibitors. Biophysical journal, 120(18), 3881-3892.
JAGGER, B. R., OJHA, A. A. & AMARO, R. E. (2020). Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. Journal of Chemical Theory and Computation, 16(8), 5348-5357.
JENSEN, M. Ø., JOGINI, V., BORHANI, D. W., LEFFLER, A. E., DROR, R. O. & SHAW, D. E. (2012). Mechanism of Voltage Gating in Potassium Channels. Science, 336(6078), 229-233.
JIANG, D., HSIEH, C.-Y., WU, Z., KANG, Y., WANG, J., WANG, E., LIAO, B., SHEN, C., XU, L., WU, J., CAO, D. & HOU, T. (2021). InteractionGraphNet: A Novel and Efficient Deep Graph Representation Learning Framework for Accurate Protein–Ligand Interaction Predictions. Journal of Medicinal Chemistry.
JOHNSTON, J. M. & FILIZOLA, M. (2011). Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. Curr Opin Struct Biol, 21(4), 552-558.
JONES, D., KIM, H., ZHANG, X., ZEMLA, A., STEVENSON, G., BENNETT, W. F. D., KIRSHNER, D., WONG, S. E., LIGHTSTONE, F. C. & ALLEN, J. E. (2021). Improved Protein–Ligand Binding Affinity Prediction with Structure-Based Deep Fusion Inference. Journal of Chemical Information and Modeling, 61(4), 1583-1592.
JOSHI, D. C. & LIN, J. H. (2019a). Delineating Protein-Protein Curvilinear Dissociation Pathways and Energetics with Naive Multiple-Walker Umbrella Sampling Simulations. J. Comput. Chem., 40(17), 1652-1663.
JOSHI, D. C. & LIN, J. H. (2019b). Delineating Protein–Protein Curvilinear Dissociation Pathways and Energetics with Naïve Multiple-Walker Umbrella Sampling Simulations. Journal of Computational Chemistry, 40(17), 1652-1663.
JUMPER, J., EVANS, R., PRITZEL, A., GREEN, T., FIGURNOV, M., RONNEBERGER, O., TUNYASUVUNAKOOL, K., BATES, R., ŽÍDEK, A. & POTAPENKO, A. (2021). Highly accurate protein structure prediction with AlphaFold. Nature, 596(7873), 583-589.

KAMENIK, A. S., LINKER, S. M. & RINIKER, S. (2022). Enhanced sampling without borders: on global biasing functions and how to reweight them. Physical Chemistry Chemical Physics.
KAMIYA, M. & SUGITA, Y. (2018). Flexible selection of the solute region in replica exchange with solute tempering: Application to protein-folding simulations. The Journal of Chemical Physics, 149(7), 072304.
KAMP, F., WINKLER, E., TRAMBAUER, J., EBKE, A., FLUHRER, R. & STEINER, H. (2015). Intramembrane proteolysis of β-amyloid precursor protein by γ-secretase is an unusually slow process. Biophysical journal, 108(5), 1229-1237.
KAPPEL, K., MIAO, Y. L. & MCCAMMON, J. A. (2015). Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. Q Rev Biophys, 48(4), 479-487.
KARPLUS, M. & MCCAMMON, J. A. (2002). Molecular dynamics simulations of biomolecules. Nat Struct Biol, 9(9), 646-652.
KHAMIS, M. A., GOMAA, W. & AHMED, W. F. (2015). Machine learning in computational docking. Artificial intelligence in medicine, 63(3), 135-152.
KINGSLEY, L. J., ESQUIVEL-RODRÍGUEZ, J., YANG, Y., KIHARA, D. & LILL, M. A. (2016). Ranking protein–protein docking results using steered molecular dynamics and potential of mean force calculations. Journal of Computational Chemistry, 37(20), 1861-1865.
KINNINGS, S. L., LIU, N., TONGE, P. J., JACKSON, R. M., XIE, L. & BOURNE, P. E. (2011). A machine learning-based method to improve docking scoring functions and its application to drug repurposing. Journal of Chemical Information and Modeling, 51(2), 408-419.
KOKH, D. B. & WADE, R. C. (2021). G Protein-Coupled Receptor–Ligand Dissociation Rates and Mechanisms from τRAMD Simulations. Journal of Chemical Theory and Computation, 17(10), 6610-6623.
LAMPRAKIS, C., ANDREADELIS, I., MANCHESTER, J., VELEZ-VEGA, C., DUCA, J. S. & COURNIA, Z. (2021). Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. Journal of Chemical Theory and Computation, 17(5), 3088-3102.
LANE, T. J., SHUKLA, D., BEAUCHAMP, K. A. & PANDE, V. S. (2013). To milliseconds and beyond: challenges in the simulation of protein folding. Curr Opin Struct Biol, 23(1), 58-65.
LIAO, J. M. & WANG, Y. T. (2019). In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. J Biomol Struct Dyn, 37(1), 166-177.
LIMONGELLI, V., BONOMI, M. & PARRINELLO, M. (2013). Funnel metadynamics as accurate binding free-energy method. Proceedings of the National Academy of Sciences, 110(16), 6358-6363.
LINDORFF-LARSEN, K., PIANA, S., DROR, R. O. & SHAW, D. E. (2011). How Fast-Folding Proteins Fold. Science, 334(6055), 517-520.
MIAO, Y., BHATTARAI, A., NGUYEN, A. T. N., CHRISTOPOULOS, A. & MAY, L. T. (2018a). Structural Basis for Binding of Allosteric Drug Leads in the Adenosine A1 Receptor. Sci Rep, 8(1), 16836.

MIAO, Y., BHATTARAI, A. & WANG, J. (2020). Ligand Gaussian Accelerated Molecular Dynamics (LiGaMD): Characterization of Ligand Binding Thermodynamics and Kinetics. J Chem Theory Comput, 16(9), 5526-5547.
MIAO, Y., CALIMAN, ALISHA D. & MCCAMMON, J. A. (2015a). Allosteric Effects of Sodium Ion Binding on Activation of the M3 Muscarinic G-Protein-Coupled Receptor. Biophys J, 108(7), 1796-1806.
MIAO, Y., FEHER, V. A. & MCCAMMON, J. A. (2015b). Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. Journal of Chemical Theory and Computation, 11(8), 3584-3595.
MIAO, Y., HUANG, Y.-M. M., WALKER, R. C., MCCAMMON, J. A. & CHANG, C.-E. A. (2018b). Ligand Binding Pathways and Conformational Transitions of the HIV Protease. Biochemistry, 57(9), 1533-1541.
MIAO, Y. & MCCAMMON, J. A. (2016). Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. Proc Natl Acad Sci U S A, 113(43), 12162-12167.
MIAO, Y. & MCCAMMON, J. A. (2018). Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. Proc Natl Acad Sci U S A, 115(12), 3036-3041.
MIURA, K. (2018). An overview of current methods to confirm protein-protein interactions. Protein and peptide letters, 25(8), 728-733.
MLÝNSKÝ, V., JANEČEK, M., KÜHROVÁ, P., FRÖHLKING, T., OTYEPKA, M., BUSSI, G., BANÁŠ, P. & ŠPONER, J. (2022). Toward Convergence in Folding Simulations of RNA Tetraloops: Comparison of Enhanced Sampling Techniques and Effects of Force Field Modifications. Journal of Chemical Theory and Computation, 18(4), 2642-2656.
MORRIS, G. M., HUEY, R., LINDSTROM, W., SANNER, M. F., BELEW, R. K., GOODSELL, D. S. & OLSON, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. J Comput Chem, 30(16), 2785-2791.
MOTTA, S., CALLEA, L., BONATI, L. & PANDINI, A. (2022). PathDetect-SOM: A Neural Network Approach for the Identification of Pathways in Ligand Binding Simulations. Journal of Chemical Theory and Computation.
NGUYEN, A. T., VECCHIO, E. A., THOMAS, T., NGUYEN, T. D., AURELIO, L., SCAMMELLS, P. J., WHITE, P. J., SEXTON, P. M., GREGORY, K. J., MAY, L. T. & CHRISTOPOULOS, A. (2016). Role of the Second Extracellular Loop of the Adenosine A1 Receptor on Allosteric Modulator Binding, Signaling, and Cooperativity. Mol Pharmacol, 90(6), 715-725.
NOÉ, F. (2020). Machine learning for molecular dynamics on long timescales. In Machine learning meets quantum physics, pp. 331-372: Springer.
NOOREN, I. M. & THORNTON, J. M. (2003). Diversity of protein-protein interactions. EMBO J, 22(14), 3486-3492.
NUNES-ALVES, A., KOKH, D. B. & WADE, R. C. (2021). Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τRAMD simulations. Current Research in Structural Biology, 3, 106-111.
OSHIMA, H., RE, S. & SUGITA, Y. (2019a). Replica-Exchange Umbrella Sampling Combined with Gaussian Accelerated Molecular Dynamics for Free-Energy Calculation of Biomolecules. J Chem Theory Comput, 15(10), 5199-5208.

OSHIMA, H., RE, S. & SUGITA, Y. (2019b). 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.
PAN, A. C., JACOBSON, D., YATSENKO, K., SRITHARAN, D., WEINREICH, T. M. & SHAW, D. E. (2019). Atomic-level characterization of protein-protein association. Proc Natl Acad Sci U S A, 116(10), 4244-4249.
PAN, A. C., XU, H., PALPANT, T. & SHAW, D. E. (2017). Quantitative Characterization of the Binding and Unbinding of Millimolar Drug Fragments with Molecular Dynamics Simulations. Journal of Chemical Theory and Computation, 13(7), 3372-3377.
PANG, Y. T., MIAO, Y., WANG, Y. & MCCAMMON, J. A. (2017). Gaussian Accelerated Molecular Dynamics in NAMD. J Chem Theory Comput, 13(1), 9-19.
PAUL, F., WEHMEYER, C., ABUALROUS, E. T., WU, H., CRABTREE, M. D., SCHÖNEBERG, J., CLARKE, J., FREUND, C., WEIKL, T. R. & NOÉ, F. (2017). Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. Nature Communications, 8(1), 1095.
PEETERS, M. C., WISSE, L. E., DINAJ, A., VROLING, B., VRIEND, G. & IJZERMAN, A. P. (2012). The role of the second and third extracellular loops of the adenosine A1 receptor in activation and allosteric modulation. Biochem Pharmacol, 84(1), 76-87.
PLATTNER, N., DOERR, S., DE FABRITIIS, G. & NOÉ, F. (2017). Complete protein–protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. Nature Chemistry, 9(10), 1005-1011.
PLATTNER, N. & NOE, F. (2015). Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. Nat Commun, 6, 7653.
POLTAVSKY, I. & TKATCHENKO, A. (2021). Machine Learning Force Fields: Recent Advances and Remaining Challenges. The journal of physical chemistry letters, 6551-6564.
PORTER, K. A., XIA, B., BEGLOV, D., BOHNUUD, T., ALAM, N., SCHUELER-FURMAN, O. & KOZAKOV, D. (2017). ClusPro PeptiDock: efficient global docking of peptide recognition motifs using FFT. Bioinformatics, 33(20), 3299-3301.
RAMANATHAN, A., SAVOL, A., BURGER, V., QUINN, S., AGARWAL, P. K. & CHENNUBHOTLA, C. (2012). Statistical inference for big data problems in molecular biophysics. In Neural Information Processing Systems: Workshop on Big Learning. Citeseer.
RANIOLO, S. & LIMONGELLI, V. (2020). Ligand binding free-energy calculations with funnel metadynamics. Nature Protocols.
RE, S., OSHIMA, H., KASAHARA, K., KAMIYA, M. & SUGITA, Y. (2019). Encounter complexes and hidden poses of kinase-inhibitor binding on the free-energy landscape. Proceedings of the National Academy of Sciences, 116(37), 18404-18409.
ROBUSTELLI, P., PIANA, S. & SHAW, D. E. (2020). Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. J Am Chem Soc, 142(25), 11092-11101.
SAGLAM, A. S. & CHONG, L. T. (2019). Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. Chemical science, 10(8), 2360-2372.

SALAWU, E. O. (2018). The Impairment of TorsinA's Binding to and Interactions With Its Activator: An Atomistic Molecular Dynamics Study of Primary Dystonia. Front Mol Biosci, 5, 64.

SALEH, N., IBRAHIM, P., SALADINO, G., GERVASIO, F. L. & CLARK, T. (2017). An Efficient Metadynamics-Based Protocol To Model the Binding Affinity and the Transition State Ensemble of G-Protein-Coupled Receptor Ligands. Journal of Chemical Information and Modeling, 57(5), 1210-1217.

SCHREIBER, G. & FERSHT, A. R. (1993). Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. Biochemistry, 32(19), 5145-5150.

SCHUETZ, D. A., BERNETTI, M., BERTAZZO, M., MUSIL, D., EGGENWEILER, H. M., RECANATINI, M., MASETTI, M., ECKER, G. F. & CAVALLI, A. (2018). Predicting Residence Time And Drug Unbinding Pathway Through Scaled Molecular Dynamics. J Chem Inf Model.

SCHUETZ, D. A., DE WITTE, W. E. A., WONG, Y. C., KNASMUELLER, B., RICHTER, L., KOKH, D. B., SADIQ, S. K., BOSMA, R., NEDERPELT, I., HEITMAN, L. H., SEGALA, E., AMARAL, M., GUO, D., ANDRES, D., GEORGI, V., STODDART, L. A., HILL, S., COOKE, R. M., DE GRAAF, C., LEURS, R., FRECH, M., WADE, R. C., DE LANGE, E. C. M., AP, I. J., MULLER-FAHRNOW, A. & ECKER, G. F. (2017). Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. Drug Discov Today, 22(6), 896-911.

SCOTT, D. E., BAYLY, A. R., ABELL, C. & SKIDMORE, J. (2016). Small molecules, big targets: drug discovery faces the protein–protein interaction challenge. Nature Reviews Drug Discovery, 15(8), 533-550.

SHAN, Y., KIM, E. T., EASTWOOD, M. P., DROR, R. O., SEELIGER, M. A. & SHAW, D. E. (2011). How Does a Drug Molecule Find Its Target Binding Site? J Am Chem Soc, 133(24), 9181-9183.

SHAO, Q. & ZHU, W. (2019). Exploring the Ligand Binding/Unbinding Pathway by Selectively Enhanced Sampling of Ligand in a Protein–Ligand Complex. The Journal of Physical Chemistry B, 123(38), 7974-7983.

SHAW, D. E., ADAMS, P. J., AZARIA, A., BANK, J. A., BATSON, B., BELL, A., BERGDORF, M., BHATT, J., BUTTS, J. A., CORREIA, T., DIRKS, R. M., DROR, R. O., EASTWOOD, M. P., EDWARDS, B., EVEN, A., FELDMANN, P., FENN, M., FENTON, C. H., FORTE, A., GAGLIARDO, J., GILL, G., GORLATOVA, M., GRESKAMP, B., GROSSMAN, J. P., GULLINGSRUD, J., HARPER, A., HASENPLAUGH, W., HEILY, M., HESHMAT, B. C., HUNT, J., IERARDI, D. J., ISEROVICH, L., JACKSON, B. L., JOHNSON, N. P., KIRK, M. M., KLEPEIS, J. L., KUSKIN, J. S., MACKENZIE, K. M., MADER, R. J., MCGOWEN, R., MCLAUGHLIN, A., MORAES, M. A., NASR, M. H., NOCIOLO, L. J., O'DONNELL, L., PARKER, A., PETICOLAS, J. L., POCINA, G., PREDESCU, C., QUAN, T., SALMON, J. K., SCHWINK, C., SHIM, K. S., SIDDIQUE, N., SPENGLER, J., SZALAY, T., TABLADILLO, R., TARTLER, R., TAUBE, A. G., THEOBALD, M., TOWLES, B., VICK, W., WANG, S. C., WAZLOWSKI, M., WEINGARTEN, M. J., WILLIAMS, J. M. & YUH, K. A. (2021). Anton 3: twenty microseconds of molecular dynamics simulation before lunch. In Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, pp. Article 1. Association for Computing Machinery, St. Louis, Missouri.

SHAW, D. E., MARAGAKIS, P., LINDORFF-LARSEN, K., PIANA, S., DROR, R. O., EASTWOOD, M. P., BANK, J. A., JUMPER, J. M., SALMON, J. K., SHAN, Y. & WRIGGERS, W. (2010). Atomic-level characterization of the structural dynamics of proteins. Science, 330(6002), 341-346.

SHEN, T. & HAMELBERG, D. (2008). A statistical analysis of the precision of reweighting-based simulations. J Chem Phys, 129(3), 034103.

SIEBENMORGEN, T. & ZACHARIAS, M. (2020). Efficient Refinement and Free Energy Scoring of Predicted Protein–Protein Complexes Using Replica Exchange with Repulsive Scaling. Journal of Chemical Information and Modeling, 60(11), 5552-5562.

SIEKER, F., STRAATSMA, T. P., SPRINGER, S. & ZACHARIAS, M. (2008). Differential tapasin dependence of MHC class I molecules correlates with conformational changes upon peptide dissociation: A molecular dynamics simulation study. Molecular Immunology, 45(14), 3714-3722.

SINKO, W., MIAO, Y., DE OLIVEIRA, C. S. A. F. & MCCAMMON, J. A. (2013). Population based reweighting of scaled molecular dynamics. The Journal of Physical Chemistry B, 117(42), 12759-12768.

SOUZA, P. C. T., ALESSANDRI, R., BARNOUD, J., THALLMAIR, S., FAUSTINO, I., GRÜNEWALD, F., PATMANIDIS, I., ABDIZADEH, H., BRUININKS, B. M. H., WASSENAAR, T. A., KROON, P. C., MELCR, J., NIETO, V., CORRADI, V., KHAN, H. M., DOMAŃSKI, J., JAVANAINEN, M., MARTINEZ-SEARA, H., REUTER, N., BEST, R. B., VATTULAINEN, I., MONTICELLI, L., PERIOLE, X., TIELEMAN, D. P., DE VRIES, A. H. & MARRINK, S. J. (2021). Martini 3: a general purpose force field for coarse-grained molecular dynamics. Nature Methods, 18(4), 382-388.

SOUZA, P. C. T., THALLMAIR, S., CONFLITTI, P., RAMIREZ-PALACIOS, C., ALESSANDRI, R., RANIOLO, S., LIMONGELLI, V. & MARRINK, S. J. (2020). Protein-ligand binding with the coarse-grained Martini model. Nat Commun, 11(1), 3714.

SPAAR, A., DAMMER, C., GABDOULLINE, R. R., WADE, R. C. & HELMS, V. (2006). Diffusional encounter of barnase and barstar. Biophysical journal, 90(6), 1913-1924.

SUGITA, Y., KAMIYA, M., OSHIMA, H. & RE, S. (2019). Replica-exchange methods for biomolecular simulations. In Biomolecular Simulations, pp. 155-177: Springer.

SUGITA, Y. & OKAMOTO, Y. (1999). Replica-exchange molecular dynamics method for protein folding. Chemical physics letters, 314(1-2), 141-151.

SUN, H., LI, Y., SHEN, M., LI, D., KANG, Y. & HOU, T. (2017). Characterizing Drug–Target Residence Time with Metadynamics: How To Achieve Dissociation Rate Efficiently without Losing Accuracy against Time-Consuming Approaches. Journal of Chemical Information and Modeling, 57(8), 1895-1906.

SUN, L., VANDERMAUSE, J., BATZNER, S., XIE, Y., CLARK, D., CHEN, W. & KOZINSKY, B. (2022). Multitask Machine Learning of Collective Variables for Enhanced Sampling of Rare Events. Journal of Chemical Theory and Computation, 18(4), 2341-2353.

SUSSMAN, J. L., LIN, D., JIANG, J., MANNING, N. O., PRILUSKY, J., RITTER, O. & ABOLA, E. E. (1998). Protein Data Bank (PDB): database of three-dimensional structural information of biological macromolecules. Acta Crystallographica Section D: Biological Crystallography, 54(6), 1078-1084.

TIWARY, P. & PARRINELLO, M. (2013). From metadynamics to dynamics. Physical review letters, 111(23), 230602.

UNKE, O. T., CHMIELA, S., SAUCEDA, H. E., GASTEGGER, M., POLTAVSKY, I., SCHÜTT, K. T., TKATCHENKO, A. & MÜLLER, K.-R. (2021). Machine Learning Force Fields. Chemical Reviews.
VAKSER, I. A. (2020). Challenges in protein docking. Current Opinion in Structural Biology, 64, 160-165.
VOTAPKA, L. W. & AMARO, R. E. (2015). Multiscale Estimation of Binding Kinetics Using Brownian Dynamics, Molecular Dynamics and Milestoning. PLOS Computational Biology, 11(10), e1004381.
VOTER, A. F. (1997). Hyperdynamics: Accelerated molecular dynamics of infrequent events. Phys Rev Lett, 78(20), 3908.
WANG, G. & ZHU, W. (2016). Molecular docking for drug discovery and development: a widely used approach but far from perfect. Future Medicinal Chemistry, 8(14), 1707-1710.
WANG, J., ARANTES, P. R., BHATTARAI, A., HSU, R. V., PAWNIKAR, S., HUANG, Y. M., PALERMO, G. & MIAO, Y. (2021). Gaussian accelerated molecular dynamics (GaMD): principles and applications. Wiley Interdiscip Rev Comput Mol Sci, 11(5), e1521.
WANG, J., ISHCHENKO, A., ZHANG, W., RAZAVI, A. & LANGLEY, D. (2022a). A highly accurate metadynamics-based Dissociation Free Energy method to calculate protein–protein and protein–ligand binding potencies. Scientific Reports, 12(1), 2024.
WANG, J., LAN, L., WU, X., XU, L. & MIAO, Y. (2022b). Mechanism of RNA recognition by a Musashi RNA-binding protein. Current Research in Structural Biology, 4, 10-20.
WANG, J. & MIAO, Y. (2020). Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. J Chem Phys, 153(15), 154109.
WANG, J. & MIAO, Y. (2022). Protein–Protein Interaction-Gaussian Accelerated Molecular Dynamics (PPI-GaMD): Characterization of Protein Binding Thermodynamics and Kinetics. Journal of Chemical Theory and Computation.
WANG, Y.-T. & CHAN, Y.-H. (2017). Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. Sci Rep, 7(1), 7828.
WANG, Y., RIBEIRO, J. M. L. & TIWARY, P. (2019). Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. Nature Communications, 10(1), 3573.
WANG, Y., RIBEIRO, J. M. L. & TIWARY, P. (2020). Machine learning approaches for analyzing and enhancing molecular dynamics simulations. Current Opinion in Structural Biology, 61, 139-145.
WIECZOREK, G. & ZIELENKIEWICZ, P. (2008). Influence of Macromolecular Crowding on Protein-Protein Association Rates—a Brownian Dynamics Study. Biophysical journal, 95(11), 5030-5036.
WU, X., KNUDSEN, B., FELLER, S. M., ZHENG, J., SALI, A., COWBURN, D., HANAFUSA, H. & KURIYAN, J. (1995). Structural basis for the specific interaction of lysine-containing proline-rich peptides with the N-terminal SH3 domain of c-Crk. Structure, 3(2), 215-226.

XIE, L., SHEN, L., CHEN, Z.-N. & YANG, M. (2017). Efficient free energy calculations by combining two complementary tempering sampling methods. The Journal of Chemical Physics, 146(2), 024103.
XUE, Y., YUWEN, T., ZHU, F. & SKRYNNIKOV, N. R. (2014). Role of electrostatic interactions in binding of peptides and intrinsically disordered proteins to their folded targets. 1. NMR and MD characterization of the complex between the c-Crk N-SH3 domain and the peptide Sos. Biochemistry, 53(41), 6473-6495.
YANG, L., LIU, C.-W., SHAO, Q., ZHANG, J. & GAO, Y. Q. (2015). From Thermodynamics to Kinetics: Enhanced Sampling of Rare Events. Accounts of Chemical Research, 48(4), 947-955.
YANG, L. & QIN GAO, Y. (2009). A selective integrated tempering method. The Journal of Chemical Physics, 131(21), 12B606.
YOU, W., TANG, Z. & CHANG, C.-E. A. (2019). Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? Journal of Chemical Theory and Computation, 15(4), 2433-2443.
ZHANG, J., WANG, N., MIAO, Y., HAUSER, F., MCCAMMON, J. A., RAPPEL, W.-J. & SCHROEDER, J. I. (2018). Identification of SLAC1 anion channel residues required for CO2 bicarbonate sensing and regulation of stomatal movements. Proc Natl Acad Sci U S A.
ZOU, R., ZHOU, Y., WANG, Y., KUANG, G., ÅGREN, H., WU, J. & TU, Y. (2020). Free Energy Profile and Kinetics of Coupled Folding and Binding of the Intrinsically Disordered Protein p53 with MDM2. J Chem Inf Model, 60(3), 1551-1558.
ZSOLDOS, Z., REID, D., SIMON, A., SADJAD, S. B. & JOHNSON, A. P. (2007). eHiTS: a new fast, exhaustive flexible ligand docking system. Journal of Molecular Graphics and Modelling, 26(1), 198-212.
ZUCKERMAN, D. M. (2011). Equilibrium Sampling in Biomolecular Simulations. Annual Review of Biophysics, 40(1), 41-62.
ZWIER, M. C., PRATT, A. J., ADELMAN, J. L., KAUS, J. W., ZUCKERMAN, D. M. & CHONG, L. T. (2016). Efficient atomistic simulation of pathways and calculation of rate constants for a protein–peptide binding process: Application to the MDM2 protein and an intrinsically disordered p53 peptide. The journal of physical chemistry letters, 7(17), 3440-3445.

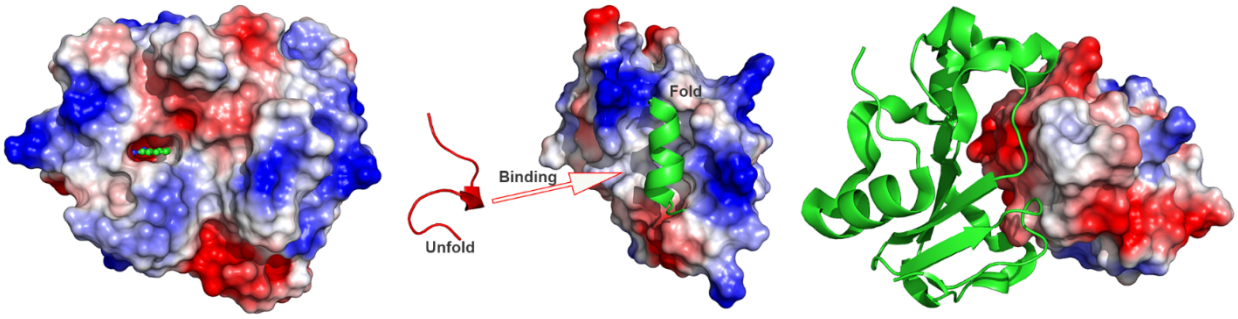

Figure 1. Schematic illustration of biomolecular recognition: (A) Small-molecule ligand binding, (B) peptide binding and (C) protein-protein interactions (PPIs).

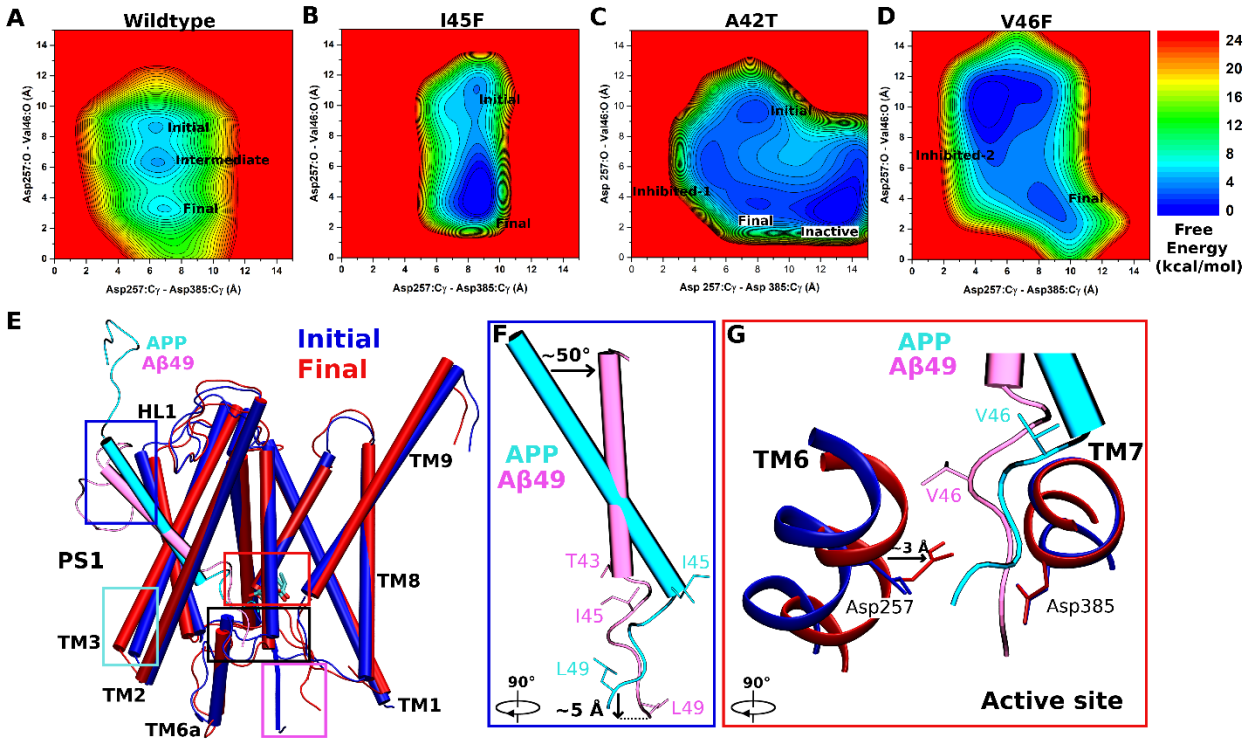

Figure 2: Mechanism of Tripeptide Trimming of Amyloid β-Peptide 49 by γ-Secretase. 2D free energy profiles calculated regarding Asp257 - Asp 385 distance and Asp257 – Aβ49 Val46 distance calculated from Pep-GaMD simulations of (A) wildtype Aβ49 bound γ-secretase, (B) I45F mutant Aβ49 bound γ-secretase, (C) A42T mutant Aβ49 bound γ-secretase, and (D) V46F mutant Aβ49 bound γ-secretase systems. (E) Structures of catalytic subunit PS1 bound to APP and Aβ49 substrates representing the "Initial" and "Final" conformational states, respectively. (F) Conformational changes in (F) Aβ49 and (G) active site of the enzyme during transition from Initial to Final activated state for ζ cleavage. Adapted with permission from Bhattari A, Devkota S., Do H.N., Wang J., Bhattarai S., Wolfe M.S. and Miao Y. Journal of the American Chemical Society. 10.1021/jacs.1c10533.Copyright 2022 American Chemical Society.

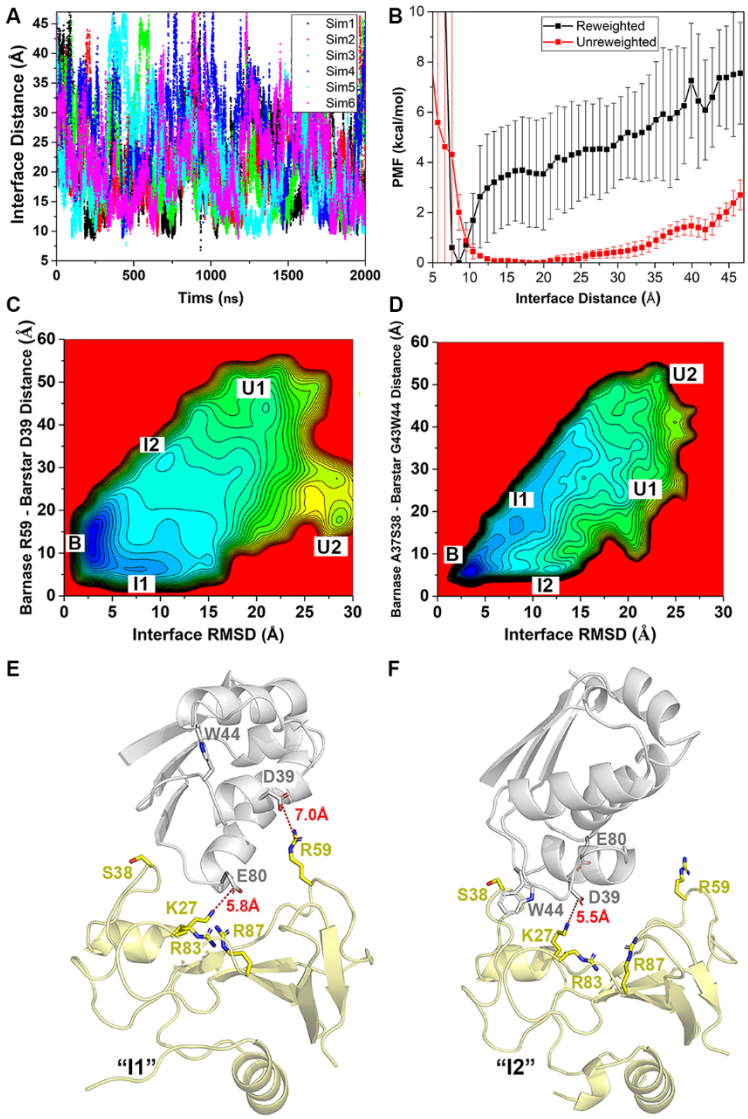

Figure 3. PPI-GaMD simulations of barnase binding/dissociation to barstar. (**A**) Time courses of protein-protein interface distance calculated from six independent 2 μs PPI-GaMD simulations. (**B**) Original (reweighted) and modified (no reweighting) PMF profiles of the protein interface distance averaged over six PPI-GaMD simulations. Error bars are standard deviations of the free energy values calculated from six PPI-GaMD simulations. (**C**) 2D PMF profiles regarding the interface RMSD and the distance between the CZ atom of barnase Arg59 and CG atom of barstar Asp39. (**D**) 2D PMF profiles regarding the interface RMSD and the distance between the center of masses (COMs) of barnase residues Ala37-Ser38 and barstar residues Gly43-Trp44. (**E**–**F**) Low-energy conformations as identified from the 2D PMF profiles of the (**E**) intermediate "I1", (**F**) intermediate "I2". Strong electrostatic interactions are shown in red dash lines with their corresponding distance values labeled in the intermediate "I1" (**E**) and "I2" (**F**). Adapted with permission from Wang J., Miao Y. Journal of Chemical Theory and Computation. 10.1021/acs.jctc.1c00974. Copyright 2022 American Chemical Society.

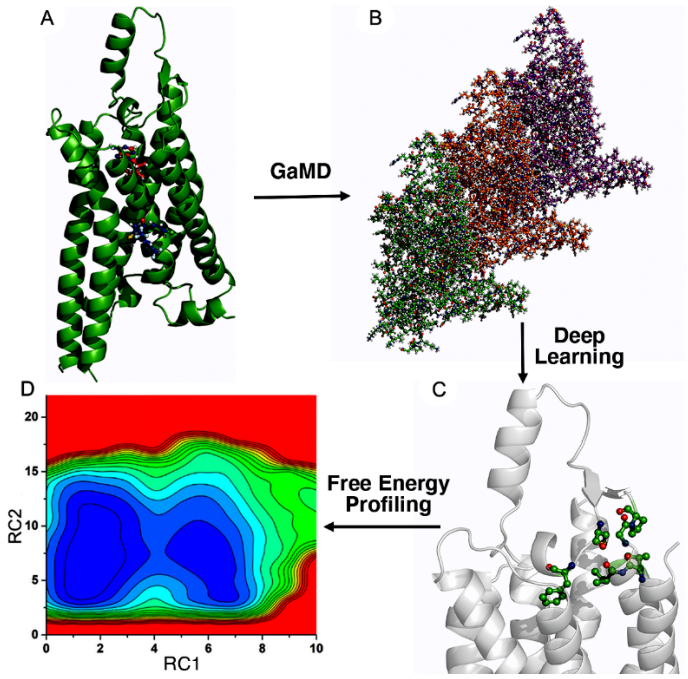

Figure 4. Overview of the Gaussian accelerated molecular dynamics (GaMD), Deep Learning (DL) and Free Energy PrOfiling Workflow (GLOW). (A) With structures of our interest, GaMD simulations are applied for enhanced sampling of the system dynamics. **(B)** DL models are then built with GaMD trajectories of residue contact maps transformed into image representations. **(C)** The DL analysis allows us to identify important residue contacts and system reaction coordinates (RCs). **(D)** Free energy profiles of the RCs are finally calculated through reweighting of GaMD simulations to characterize the system dynamics. Adapted with permission from Do H.N. Wang J. Bhattari A. and Miao Y. Journal of Chemical Theory and Computation. 10.1021/acs.jctc.1c01055. Copyright 2022 American Chemical Society.