JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Review

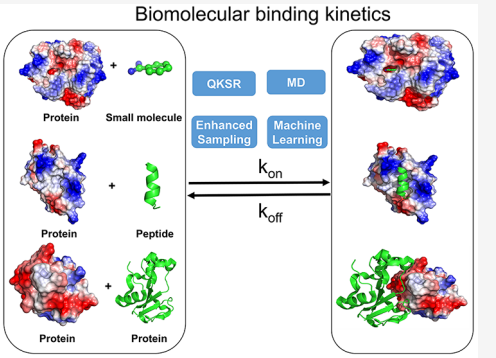
Predicting Biomolecular Binding Kinetics: A Review

Jinan Wang, Hung N. Do, Kushal Koirala, and Yinglong Miao*

Cite This: *J. Chem. Theory Comput.* 2023, 19, 2135−2148 Read Online

ACCESS | Metrics & More | Article Recommendations

ABSTRACT: Biomolecular binding kinetics including the association (k_{on}) and dissociation (k_{off}) rates are critical parameters for therapeutic design of small-molecule drugs, peptides, and antibodies. Notably, the drug molecule residence time or dissociation rate has been shown to correlate with their efficacies better than binding affinities. A wide range of modeling approaches including quantitative structure-kinetic relationship models, Molecular Dynamics simulations, enhanced sampling, and Machine Learning has been developed to explore biomolecular binding and dissociation mechanisms and predict binding kinetic rates. Here, we review recent advances in computational modeling of biomolecular binding kinetics, with an outlook for future improvements.



1. INTRODUCTION

Life processes are critically dependent on the formation of biomolecular complexes, particularly the protein-small molecule, protein−peptide, and protein−protein/antibody structures. Biomolecular binding plays a key role in many fundamental biological processes.[1] Accurate characterization of biomolecular binding thermodynamics and kinetics is key for therapeutic design.[2−4] The ligand free energy and kinetics are related as $\Delta G_{binding} = -RT \ln k_d$ with $k_d = (k_{off}/k_{on})$. It is possible for ligands with similar binding free energy to exhibit distinct binding and dissociation kinetic rates. Particularly, drug residence time or dissociation rate appears to correlate with *in vivo* drug efficacy better than the binding free energy.[5−12] Therefore, understanding the receptor−ligand binding and unbinding process and accurate predictions of ligand binding kinetic rates could be valuable for drug discovery and development. It is desirable to decrease the ligand dissociation rate or increase its residence time to improve its efficacy.[10] For example, Li et al.[10] optimized the donepezil drug to compound 12 through adding two F atoms to decrease the dissociation rate from its target acetylcholinesterase, which demonstrated significantly improved efficacy and a lower effective dose than that of donepezil. With remarkable theoretical and technical developments, increasing numbers of experimental and computational methods are available for calculating the biomolecular binding kinetic rates.[5,9,12−20] However, it remains challenging for both experimental and computational approaches to accurately predict biomolecular binding kinetic rates with high throughput.

In this review, we will first briefly describe available experimental techniques for determining biomolecular binding kinetic rates. We will then discuss computational approaches to predict the biomolecular binding kinetics published during 2010−2022, with a focus on the Molecular Dynamics (MD) and enhanced sampling methods, including Weighted Ensemble,[21−23] milestoning,[24] simulation enabled estimation of kinetic rates (SEEKR),[25−27] Gaussian accelerated Molecular Dynamics (GaMD),[28,29] Metadynamics[30−37] and its combination with Machine Learning (ML),[38] Markov State Modeling (MSM),[39−41] Random Acceleration Molecular Dynamics (RAMD),[42−44] scaled MD,[45−47] and so on. These computational approaches have emerged as rapidly evolving techniques for studying biomolecular binding kinetics.

2. AVAILABLE EXPERIMENTAL TECHNIQUES TO MEASURE BINDING KINETICS

Most experimental techniques[14,48,49] for determining biomolecular binding kinetic rates are mainly relying on monitoring a specific signal over time during the binding and dissociation processes. According to signal source, experimental methods could be generally divided into two classes: assays with and without a label for detection.[14] Radio and spectroscopic labeling are the main choices for labeling assays. A radiolabel essentially comes from the presence of radioactive isotopes in the molecule, which could emit special radiation when they decay to more stable states. In radiometric binding assays,

Received: October 31, 2022
Published: March 29, 2023

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

Table 1. Databases of Biomolecular Binding Kinetics

database	description	web site
KDBI	It includes 19,263 entries, which provide experimentally verified kinetic rates for protein−protein/DNA/RNA/ligand and ligand-DNA/RNA interactions.	http://xin.cz3.nus.edu.sg/group/kdbi/kdbi.asp
BindingDB	It focuses on protein−ligand interactions, including ∼1.1 million compounds and 8900 targets.	https://www.bindingdb.org/rwd/bind/index.jsp The webpage of binding kinetic rates: https://bindingdb.org/rwd/bind/ByKI.jsp?specified=Kn
KOFFI	It includes 1705 entries and a rating system to measure the quality of experimental data.	http://koffidb.org/
PDBbind	The k_{off} data set includes 680 entries with protein-small molecule complex structures.	http://www.pdbbind.org.cn/
SKEMPI	It focuses on protein−protein interactions, which record 713 binding association and dissociation rates upon mutation.	http://life.bsc.es/pid/mutation_database/
dbMPIKT	It focuses on protein−protein interactions, which contain 5291 protein binding association and dissociation rates upon mutation.	http://deeplearner.ahu.edu.cn/web/dbMPIKT/

ligands are tagged to follow the time course of their binding to targets, thus allowing for the spontaneous measurement of binding kinetic rates.[50,51] In the spectroscopy-based assays, ligands are labeled with fluorophore groups. After absorbing a certain wavelength's light, fluorophore groups could emit characteristic light, allowing for detecting the binding and dissociation processes.[52] The fluorescent resonance energy transfer (FRET) is one popular spectroscopy based approach.[52] The labeling approach is often less efficient as it requires the labeling of the ligand, which is labor intensive and could be challenging for certain drugs. For the label free approaches, surface plasmon resonance (SPR) is one of the most widely used methods, especially in characterizing the biomolecular binding kinetics of pharmaceutical interest.[2] High throughput is one important advantage of SPR, which is often used to analyze a series of ligands.

With developments of experimental techniques, recent years have seen significantly increasing numbers of biomolecular binding kinetic data, including the protein-small molecule, protein−peptide, and protein−protein binding kinetic rate constants. Many experimental binding kinetic rates have been collected in different publicly accessible databases. A number of databases as listed in Table 1 are useful for exploring biomolecular binding kinetics, including the kinetic data of biomolecular interactions (KDBI),[53] BindingDB,[54] kinetics of featured interactions (KOFFI),[55] PDBbind,[56] structural database of kinetics and energetics of mutant protein interactions (SKEMPI),[57] kinetic and thermodynamic database of mutant protein interactions (dbMPIKT),[58] and so on.[8,59]

KDBI[53] is developed to provide experimentally verified binding kinetic rates for interactions involving proteins and nucleic acids (RNA and DNA). It includes 19,263 entries of 10,532 distinctive biomolecular interactions. The binding kinetic data includes protein−protein/DNA/RNA/ligand and ligand-DNA/RNA interactions. BindingDB[54] is one widely used database for exploring protein-small molecule interactions, containing ∼1.1 million compounds and 8900 targets with clearly defined quantitative measurement for binding affinities and kinetic rates. BindingDB provides a special kinetic database via link https://bindingdb.org/rwd/bind/ByKI.jsp?specified=Kn. The data of BindingDB are extracted from published literature and other databases such as PubChem, CheEMBL, PDSP K_i, and CSAR. Additionally, BindingDB provides an option for experimentalists to directly deposit their data. KOFFI[55] is developed to provide binding kinetic rates along with experimental protocol. It includes 1705 individual entries. Notably, it contains a rating system to assess the quality of experimental data. A user can perform a direct search within the Anabel's KOFFI database and evaluate the quality of their binding data. PDBbind[56] was initially developed for collecting binding affinity data and complex structures for developing docking score. In 2022, it released a subdatabase (k_{off} set) containing 169 entries of protein-small molecule dissociation rates. One advantage of PDBbind is the availability of the protein-small molecule complex structures, which could be convenient for molecular modeling. SKEMPI[57] and dbMPIKT[58] mainly focus on protein−protein interaction (PPI). SKEMPI[57] contains 713 protein−protein binding kinetic rates upon mutation. dbMPIKT[58] contains 5291 entries of protein−protein binding kinetic rates involving mutation. In summary, developments of experimental techniques and increasing biomolecular binding kinetic data collected in the databases will greatly facilitate modeling of biomolecular binding kinetics and therapeutic design.

3. QUANTITATIVE STRUCTURE-KINETIC RELATIONSHIPS

Optimal kinetic parameters for biomolecular binding could significantly improve drug efficacy. For that reason, several molecular modeling techniques have been developed to predict biomolecular binding kinetic rates and derive quantitative structure-kinetic relationships (QSKRs).[60] While these methods are often based on experimental structures, many of them consider each biomolecular complex with only one single structure.[60] Nunes-Alves et al.[60] modified the COMparative BINding Energy (COMBINE) analysis, which uses *holo* structure to predict binding parameters, to include extra options of using multiple protein-small molecule complex structures. They did so by docking small molecules to a protein conformational ensemble obtained from MD simulations. Specifically, the full data set for COMBINE analysis consisted of 33 inhibitors of p38 MAP kinase, which were chosen given availability of experimental k_{off} values and experimental structures of the inhibitor bound to p38 MAP kinase or to other kinases in the DFG-out conformation state. Twenty-two and eleven inhibitors were used for training and testing in the COMBINE analysis, respectively. The first step in the COMBINE analysis involved modeling of the two sets of structures and derivation of COMBINE analysis models. After energy minimization of the complex structures, interaction energy components were obtained with the AMBER ff14SB force field to describe bonded and nonbonded interactions. Weights to scale the protein-small molecule interaction energies were obtained using partial least-squares regression. To account for multiple structures, the COMBINE was modified to retrieve an average response using N structures for each protein-small molecule complex, in which each structure was treated independently during regression to

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

obtain weights for interaction energies. Here, exponential or arithmetic averages could be used

$$(\log I)_{exp}^{comp} = -\log \frac{1}{N}\sum_{j=1}^{N} e^{-\log I^j} \quad (1)$$

$$(\log I)_{arit}^{comp} = \frac{1}{N}\sum_{j=1}^{N} \log I^j \quad (2)$$

where $(\log I)_{exp}^{comp}$ and $(\log I)_{arit}^{comp}$ were the predictions for the response variable using exponential and arithmetic averages, j was the index of the structure used, $\log I^j$ was the prediction made using the j^{th} structure, and N was the number of structures to describe one protein-small molecule complex. In one of the two structure sets used for the COMBINE analysis, each complex was represented using one experimental crystal structure. In another set, each complex was represented using 10 structures from ensemble docking.[60] Although the COMBINE model obtained with multiple structures from ensemble docking took protein−ligand flexibility into consideration, the predictive power was lower than the model from a single, energy-minimized crystal structure for each protein−ligand complex. Nevertheless, the incorporation of protein−ligand flexibility highlighted additional important protein−ligand interactions that led to longer residence time.

In 2018, Ganotra and Wade applied COMBINE analysis to derive QSKRs for the dissociation rates (k_{off}) of inhibitors of HSP90 and HIV-1 protease.[61] Protein-specific scoring functions were derived by correlating k_{off} with a subset of weighted interaction energy components determined from energy minimized biomolecular complex structures. A set of 3D structures of protein−ligand complexes was modeled and energy minimized. Protein−ligand interaction energies were first calculated, then partitioned, and subjected to partial least-squares projection to latent structures (PLS) regression. A statistical model was derived to correlate the activity of interest to weighted selected components of the protein−ligand interaction energy decomposed on a per residue basis, based on the following equation

$$\log(k) = \sum_{i=1}^{n} w_i \Delta u_i + C \quad (3)$$

where k was the rate constant of interest, and Δu_i was per residue terms of the ligand−receptor interaction energy, calculated for n residues. The coefficients w_i and constant C could be determined from PLS regression. The data set used for the COMBINE analysis of HSP90 and HIV-1 protease inhibitors consisted of 70 and 36 compounds, respectively. Experimental k_{off} values ranged from 0.0001 to 0.83 s^{-1} for the HSP90 inhibitors and 0.00022 to 83.3 s^{-1} for the HIV-1 protease inhibitors. For the COMBINE analysis, 207 Coulombic and 207 Lennard-Jones (LJ) interaction energy terms were calculated for the HSP90 inhibitors, and 198 Coulombic and 198 LJ energies were calculated for the HIV-1 protease inhibitors. The resulting COMBINE models for k_{off} rates had very good predictive power (R^2 = 0.80 and Q^2_{LOO} = 0.69 for HSP90, and R^2 = 0.94 and Q^2_{LOO} = 0.70 for HIV-1 protease). They could also identify contributing protein−ligand interactions for the binding kinetics.

In another study, Schuetz et al.[62] performed matched molecular pair (MMP) analysis on data sets assembled from the Kinetic for Drug Discovery database, which included 3812 small molecules annotated to 78 different targets from five diverse protein classes, including G-protein-coupled receptors (GPCRs), kinases and other enzymes, heat shock proteins (HSPs), and ion channels. The kinetic data set (KIND) contains complex structures along with their respective binding kinetic data (k_{on}, k_{off}, K_D). To elucidate the impact of small structural changes on the binding kinetic behavior, a total of 395 MMPs extracted from KIND were performed. The pairs were made of two molecules possessing identical scaffolds and showing minor chemical modifications. This data set included various chemical modifications, with the top 20 representing less than 65% of the entire data set. The most common modification, which was replacement of a hydrogen atom by a methyl group, comprised around 15%. To demonstrate that changes in a molecule's polarity are the major factor for the alteration of binding association rate k_{on}, the authors focused on analyzing the top 20 MMPs with highest differences in k_{on} values. For 16 out of 20 MMPs, a substitution that increases polarity was observed. The largest differences in k_{on} were found with the introduction of a charged moiety, leading to a decrease of 0.5−2.0 orders of magnitude. The decrease in k_{on} might come from electrostatic repulsion and/or desolvation penalties. Conversely, an improvement in binding affinity was observed if modifications established additional interactions in the final bound complexes. The dissociation rate k_{off} was also analyzed following the same protocol for k_{on}. In contrast to k_{on}, the change of molecular polarity in the MMPs did not produce a consistent shift in k_{off}.

In order to explore molecular details of biomolecular binding processes on a large scale, Chiu et al.[63] recently integrated coarse-grained normal-mode analysis (NMA) with multitarget machine learning (MTML) to address the above challenge and tested their method using the HIV-1 protease as a model system. The workflow included four phases. In phase 1, the 3D complex structure of the ligand-bound HIV-1 protease was built. Ligands without experimental structure were docked into the HIV-1 protease using the eHiTS software. In phase 2, residues in the ligand-binding site were identified. Coarse-grained NMA was performed for both *apo* and *holo* structures. The authors defined relative movement of ligand−residue (RMLR) as the dot product of the ligand displacement vector after normalization and the residue displacement vector and relative movement of residue−residue (RMRR) as the dot product of the displacement vectors of a residue for the *apo* and *holo* structures. Therefore, RMLR and RMRR could be derived from the NMA and describe the conformational dynamics impact of ligand binding on the residues in the binding site. In phase 3, five principal data sets were constructed. Pairwise decomposition of the residue interaction energy was computed by minimizing 39 ligand-bound HIV-1 complexes with NAMD simulations using the generalized Born implicit solvent (GBIS) method. The final simulation conformations were used to compute the residue-decomposed pairwise interaction energy (PIE), the van der Waals energy (VDWE), and the electrostatic energy (EE) between the ligand and protein residues. The energetic features (PIE, VDWE, and EE) and conformational dynamics features (RMRR and RMLR), along with experimentally determined k_{on} and k_{off} data, were used to train MTML models in phase 4 of the workflow. The model was evaluated regarding the accuracies in the predictions of binding kinetic rate constants k_{on} and k_{off} using the following formula

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

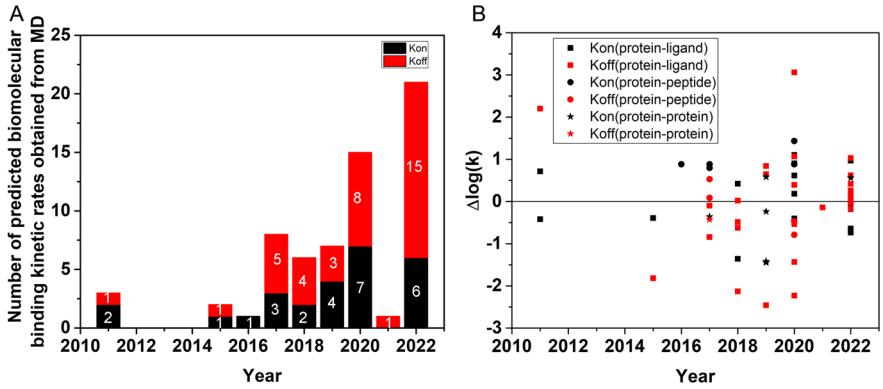


Figure 1. Number (A) and accuracy (B) of predictions of biomolecular binding kinetic rates obtained from MD simulations plotted over the years. The protein−ligand, protein−peptide, and protein−protein complexes were plotted in squares, circles, and asterisks, respectively.

$$\text{accuracy} = \sum_{i=1}^{n} \frac{A_i}{N} \tag{4}$$

where A_i was the prediction accuracy for each case, and N was the total number of cases. A_i = 100% when both k_{on} and k_{off} were accurately predicted, and A_i = 0% when neither was correctly predicted. The ML algorithms are generally sensitive to imbalances in the training data. In order to avoid the bias of ML algorithms in handling an imbalanced small training data set, the author discretized the $\log_{10}(k_{off})$ and $\log_{10}(k_{on})$ values of 39 HIV-1 protease inhibitors into four different binary classes, with the labels (0,0), (0,1), (1,0), and (1,1) on the two-dimensional space of $\log_{10}(k_{on})$ and $\log_{10}(k_{off})$, using the criteria of $\log_{10}(k_{off}) = -2$ and $\log_{10}(k_{on}) = 5.6$, where class (0,1) is enriched with five FDA-approved drugs. Then, the author evaluated the prediction accuracy of the combined four-class $\log_{10}(k_{on})/\log_{10}(k_{off})$. They found that the prediction accuracy in their models was higher than that of the random guess. The model was further evaluated in high-throughput screening of molecules with *in vivo* drug activity on the basis of k_{on} and k_{off} using the receiver operating characteristic (ROC) curve and the area under the ROC curve (AUC). The computational models were found not only to recapitulate the results from MD simulations but also to accurately predict protein−ligand binding kinetic rates, with an accuracy of 74.35% when combined with energy features. In addition, the integrated models showed that the coherent coupling of conformational dynamics and thermodynamic interactions between the receptor and ligand played a critical role in determining protein−ligand binding kinetic rates.

In summary, with increasing numbers of available experimental binding kinetic data and advances in the modeling approaches, the built QSKR will become more accurate and allow for high-throughput screening, which is very helpful at the early stage of drug design.

4. MOLECULAR DYNAMICS AND ENHANCED SAMPLING METHODS FOR PREDICTING BINDING KINETICS

MD is a powerful technique for simulations of biomolecular structural dynamics.[64−69] The accessible time scale of conventional MD (cMD) has reached hundreds of microseconds thanks to remarkable advances in computing hardware (e.g., the Anton supercomputer and GPUs) and algorithm developments.[70−75] Notably, the latest Anton3[75] has enabled hundreds-of-microseconds cMD simulations per day for ATPase and Satellite Tobacco Mosaic Virus (STMV) with a total number of atoms ranging from 328 K to 1,067 K. The cMD simulations have been widely applied to investigate biomolecular binding processes.[76−78] However, it is still challenging for cMD to simulate repetitive biomolecular dissociation and rebinding processes.[70,79] In this regards, enhanced sampling methods[80−83] have been developed to simulate biomolecular binding and dissociation processes and predict the associated binding kinetic rates. Recent years have seen a significant increasing numbers of studies on predicting biomolecular binding kinetic rates using MD simulations (Figure 1A). To evaluate the accuracy of simulation predicted kinetic rates, we define the prediction errors of binding and dissociation kinetic rates as

$$\Delta \log k_{on} = \log k_{on}^{sim} - \log k_{on}^{exp},$$
$$\Delta \log k_{off} = \log k_{off}^{sim} - \log k_{off}^{exp} \tag{5}$$

where simulation predicted binding (k_{on}^{sim}) and dissociation (k_{off}^{sim}) rates are compared with experimentally determined binding (k_{on}^{exp}) and dissociation (k_{off}^{exp}) rates. Most values of the $\Delta \log k$ are in the range of −1 to 1 (Figure 1B), suggesting good prediction accuracy obtained from MD simulations. Due to the difficulty of prediction in ligand binding kinetic rates, $\Delta \log k$ in a range of −2 to 2 is acceptable for comparing computational predictions with experimental values. In the next sections, we will discuss recent applications of the above-mentioned methods in exploring biomolecular binding kinetics for distinct protein-small molecule, protein−peptide, and protein−protein binding systems, while very few simulation studies have been carried out about binding kinetics of nucleic acids.

Protein-Small Molecule Binding Kinetics. Compared with slower ligand dissociation process, ligand binding is much quicker, which allows cMD to capture the ligand binding process and predict the binding association rate (k_{on}). For example, spontaneous binding of the Dasatinib drug to its target Src kinase was observed in a total of ∼35.0 μs cMD simulations performed by Shan et al.[76] The estimated binding association rate (k_{on}) was 0.19 × 10^7 M^{-1} s^{-1}, being highly consistent with the experimental value of 0.5 × 10^7 M^{-1} s^{-1}. The same system was used to test a novel approach-unaggregated unbiased MD (UUMD) developed by Sohraby et al.[84] In contrast to the repulsion added to the special atom in the ligand by Shan et al.,[76] the repulsion in the UUMD was

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

Table 2. Summary of Computer Simulation Predicted Protein-Ligand Binding (k_{on}^{sim}) and Dissociation (k_{off}^{sim}) Rates Compared with Experimentally Determined Binding (k_{on}^{exp}) and Dissociation (k_{off}^{exp}) Rates

system	method	k_{on}^{exp} (10^7 M^{-1} s^{-1})	k_{off}^{exp} (s^{-1})	k_{on}^{sim} (10^7 M^{-1} s^{-1})	k_{off}^{sim} (s^{-1})	sim. time (μs)	Δ log k_{on}	Δ log k_{off}	force field	year[ref]
trypsin-benzamidine	M-WEM	2.9	600	0.53 ± 0.08	791 ± 197	0.48	−0.74	0.12	AMBER ff14SB and GAFF	2022[85]
trypsin-benzamidine	SEEKR2	2.9	600	2.4 ± 0.2	990 ± 130	5	−0.082	0.22		2022[86]
trypsin-benzamidine	InMetaD+ML	2.9	600		1560	2.75		0.41	AMBER ff14SB and GAFF	2022[87]
trypsin-benzamidine	LiGaMD	2.9	600	1.15 ± 0.79	3.53 ± 1.41	5	−0.40	−2.23	AMBER ff14SB and GAFF	2020[88]
trypsin-benzamidine	SEEKR	2.9	600	12 ± 0.5	174 ± 9	4.4	0.62	−0.54		2020[89]
trypsin-benzamidine	CGMD	2.9	600	36.8	6.9×10^5	428	1.10	3.06	MARTINI	2020[90]
trypsin-benzamidine	InMetaD	2.9	600		4176 ± 324	∼1.00		0.84	CHARMM 36 and CGenFF	2019[91]
trypsin-benzamidine	WE	2.9	600		2660	8.75		0.65	CHARMM and CGenFF	2019[21]
trypsin-benzamidine	InMetaD	2.9	600	1.18 ± 1.0	9.1 ± 2.5		−0.39	−1.82	AMBER ff99SB-ILDN and GAFF	2015[36]
trypsin-benzamidine	MSM	2.9	600	15 ± 2	$9.5 \pm 3.3 \times 10^4$	50	0.71	2.20	AMBER ff99SB and GAFF	2011[39]
T4L-BEN	LiGaMD2	0.08−0.1	950 ± 200	0.742 ± 0.481	1440 ± 880	3	0.97	0.18	AMBER ff14SB and GAFF	2022[92]
T4L-INO	LiGaMD2	0.07−0.1	325	0.299 ± 0.287	3490 ± 560	3	−0.64	1.03	AMBER ff14SB and GAFF	2022[92]
M102A T4L-BEN	LiGaMD2	0.30−0.50	3000	0.957 ± 0.629	2010 ± 1610	3	−0.18	−0.17	AMBER ff14SB and GAFF	2022[92]
T4L-BEN	ML	0.08−0.1	950 ± 200		3.3 ± 0.8			−2.46	CHARMM22*	2019[93]
T4L-BEN	InMetaD	0.08−0.1	950 ± 200	0.0035 ± 0.002	7 ± 2	12	−1.36	−2.13	CHARMM22* and CGenFF	2018[32]
T4L-BEN	MSM	0.08−0.1	950 ± 200	0.21 ± 0.09	310 ± 130	59	0.42	−0.49	CHARMM36	2018[41]
T4L-BEN	WE	0.08−0.1	950 ± 200		1000	29		0.022	CHARMM36	2018[23]
Src-Dasatinib	cMD	0.5	0.06	0.76		6.6	0.18		OPLS	2020[84]
Src-Dasatinib	CGMD	0.5	0.06	4		300	0.90		MARTINI	2020[94]
Src-Imatinib	MetaD		0.11 ± 0.08		0.026			−0.63	AMBER ff99SB-ILDN and GAFF	2018[95]
Src-Dasatinib	InMetaD	0.5	0.06		0.048 ± 0.024	7		−0.096	OPLS	2017[96]
Src-Dasatinib	cMD	0.5	0.06	0.19		35	−0.42		AMBER ff99SB and GAFF	2011[76]
P38α-compound I	InMetaD	0.0118	0.14		0.02 ± 0.01	6.8		−0.84	AMBER ff99SB-ILDN and GAFF	2017[37]
JAK2-Inhibitor 9	SEEKR2		0.090		0.091	25.2		0.002754	AMBER ff14SB and GAFF	2022[27]
JAK2-Inhibitor 7	SEEKR2		0.15		0.15	25.2		−0.00905	AMBER ff14SB and GAFF	2022[27]
JAK2-Inhibitor 6	SEEKR2		0.31		0.43	25.2		0.14229	AMBER ff14SB and GAFF	2022[27]
JAK3-inhibitor 5	SEEKR2		1.28		5.35	24		0.620253	AMBER ff14SB and GAFF	2022[27]
JAK2-Inhibitor 9	SEEKR2		0.090		0.091	25.2		0.002754	AMBER ff14SB and GAFF	2022[27]
JAK2-inhibitor 5 JAK3-Inhibitor 9	SEEKR2		0.32		0.39	24		0.09519	AMBER ff14SB and GAFF	2022[27]
JAK3-Inhibitor 7	SEEKR2		0.77		0.90	24		0.066279	AMBER ff14SB and GAFF	2022[27]
JAK3-inhibitor 5	SEEKR2		1.28		5.35	24		0.620253	AMBER ff14SB and GAFF	2022[27]
M2-Iperoxo	Frequency-adaptive MetaD		0.01 ± 0.002		$3.7 \pm 0.0.7 \times 10^{-4}$	8		−1.43	AMBER ff14SB and GAFF	2020[97]
μOR-morphine	InMetaD+ML	0.29 ± 0.001	0.023 ± 0.001		0.057 ± 0.005	6		0.39	CHARMM 36 and CGenFF	2020[38]
μOR-bruprenorphine	InMetaD+ML	1.33 ± 0.01	0.0018 ± 0.003		0.021 ± 0.003	19		1.07	CHARMM 36 and CGenFF	2020[38]

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

added to a virtual interaction site in the ligand to avoid aggregation. Notably, the UUMD could capture multiple independent Dasatinib binding events within nanosecond simulations. The predicted binding association rate (k_{on}) was 0.75×10^7 M^{-1} s^{-1}, being highly consistent with the experimental data (Table 2). It is worth noting that no dissociation event was observed in the cMD simulations, prohibiting calculation of ligand dissociation rate (k_{off}).

Coarse-grained models were developed for MD simulations to reduce the demands for computational resources and extend the simulation time scale.[94,98] Based on the Martini coarse-grained model, Dandekar et al.[90] captured spontaneous binding of benzamidine to the trypsin binding pocket from bulk solvent. Based on 426 μs MD simulation data, they predicted the binding kinetic rates of (k_{on}, k_{off}) at (36.8×10^7 M^{-1} s^{-1}, 6.9×10^5 s^{-1}). The corresponding experimental values were (2.9×10^7 M^{-1} s^{-1}, 600 s^{-1}). Therefore, the predicted k_{on} value was ~13-fold higher than the experimental data. However, a large derivation was observed between the predicted and experimentally determined k_{off}.

Multiscale computational approaches have been developed to improve the efficiency and accuracy of ligand binding thermodynamics and kinetics calculations.[25,99,100] For example, SEEKR[25−27] is a multiscale simulation approach combining MD, Brownian dynamics, and milestoning for predicting protein−ligand binding association and dissociation rates. The recently developed version of SEEKR with Markovian milestoning with the Voronoi tessellations approach has been shown to estimate accurate binding kinetic rates with the simulation time reduced by a factor of ~10 in comparison to the original SEEKR.[25] Using the trypsin-benzamidine model system as an example, the SEEKR[89] and its latest version SEEKR2[86] predicted the binding kinetic rates of (k_{on}, k_{off}) at ($12 \pm 0.5 \times 10^7$ M^{-1} s^{-1}, 174 ± 9 s^{-1}) and ($2.4 \pm 0.2 \times 10^7$ M^{-1} s^{-1}, 990 ± 130 s^{-1}), respectively, being highly consistent with the corresponding experimental data of (2.9×10^7 M^{-1} s^{-1}, 600 s^{-1}). Particularly, SEEKR2[27] was recently applied to predict the dissociation rates of a number of inhibitors for the Janus Kinase (JAK) system. The predicted values k_{off} agreed very well with the experimental data with Δ log k less than 1 (Table 2).[27]

The milestoning method[24] has been applied to predict the dissociation rate of the Imatinib drug to Abl kinase. Based on the total of 1.043 μs simulations, the value of k_{off} was predicted as 18 s^{-1}, being highly consistent with the experimental value of 25 ± 6 s^{-1}. Weighted Ensemble[21] and MSM[39] have been developed to improve prediction of ligand binding kinetic rates based on a large number of short cMD trajectories. In the trypsin-benzamidine system, the dissociation rate (k_{off}) of 2,660 s^{-1} was predicted with one weighted ensemble[21] of a total amount of 8.75 μs cMD simulations, being ~4.43 times faster than the experimental value. Another weighted ensemble combined with milestoning[85] of a total of 0.48 μs cMD simulations was able to predict the T4 lysozome (T4L)-benzene binding kinetic rates of (k_{on}, k_{off}) at ($0.53 \pm 0.08 \times 10^7$ M^{-1} s^{-1}, 791 ± 197 s^{-1}), being highly consistent with the corresponding experimental value of ($0.08-0.1 \times 10^7$ M^{-1} s^{-1}, 950 ± 200 s^{-1}). MSM was able to simultaneously predict the ligand association and dissociation rates through longer aggregated cMD simulations. For example, one MSM built with 59 μs cMD simulation data was able to accurately predict T4L-benzene binding kinetic rates.[41] The predicted binding kinetic rate values of (k_{on}, k_{off}) were ($0.21 \pm 0.09 \times 10^7$ M^{-1} s^{-1}, 310 ± 130 s^{-1}), being highly consistent with the experimental data of ($0.08-0.1 \times 10^7$ M^{-1} s^{-1}, 950 ± 200 s^{-1}). MSM built with 50 μs cMD simulation data was used to predict the binding kinetic rates of the trypsin-benzamidine system.[40] The predicted values of (k_{on}, k_{off}) were ($15.0 \pm 2.0 \times 10^7$ M^{-1} s^{-1}, 9.5 ± 3.310^4 s^{-1}), being in line with the experimental values of (2.9×10^7 M^{-1} s^{-1}, 600 s^{-1}). However, these calculations required very expensive computational resources.

Metadynamics[30,31] has been widely applied to investigate the ligand binding kinetics. Multiple Infrequent Metadynamics (InMetaD) simulations with a total of 5 μs trajectories were performed to predict the pathways of benzamidine binding to the trypsin and the binding kinetic rates. The predicted values of (k_{on}, k_{off}) were ($1.18 \pm 1.0 \times 10^7$ M^{-1} s^{-1}, 9.1 ± 2.5 s^{-1}), being smaller than the experimental values of (2.9×10^7 M^{-1} s^{-1}, 600 s^{-1}). Similar smaller predicted values of (k_{on}, k_{off}) at ($0.0035 \pm 0.002 \times 10^7$ M^{-1} s^{-1}, 7 ± 2 s^{-1}) were observed in another 12 μs InMetaD simulations of benzene binding to T4L.[32] For the Src-Dasatinib system, one study with 7 μs InMetaD simulations[96] was able to predict the k_{off} value of 0.048 ± 0.024 s^{-1}, being highly consistent with the experimental value of 0.06 s^{-1}. For the p38α-compound I system, 6.8 μs InMetaD simulations[37] predicted the k_{off} value of 0.020 ± 0.011 s^{-1}, being in line with the experimental value of 0.14 s^{-1}. Besides, accuracy of force field also plays a critical role in predicting biomolecular binding kinetic rates. For example, Capelli et al.[97] applied two approaches to obtain the RESP charges for the drug Iperoxo to predict its dissociation rate in the M2 receptor. The two approaches included the one with Amber standard methodology based on HF/6-31G* (RESP-HF) calculations and another one based on DFT/B3LYP (RESP-B3LYP) calculations. The simulations based on RESP-HF charges failed to predict the k_{off} rate due to the unreasonable obtained transition state free energy. Simulations with RESP-B3LYP charges enabled prediction of the k_{off} value at $3.7 \pm 0.7 \times 10^{-4}$ s^{-1}, with an ~2 order of magnitude deviation from the experimental value of $1.0 \pm 0.2 \times 10^{-2}$ s^{-1}. For the Src-Imatinib system, Haldar et al.[95] showed that accounting for changes in charge distribution with QM/MM calculations improved the Imatinib dissociation rate from 0.0114 s^{-1} to 0.026 s^{-1}, being more consistent with the experimental value of 0.11 ± 0.08 s^{-1}. Although Metadynamics simulations have shown remarkable improvements in capturing ligand binding and dissociation processes that occur over exceedingly long time scales, users often face a challenge for defining collective variables (CVs), which requires expert knowledge of the studied systems.[101,102] The simulations may suffer from a "hidden energy barrier" problem if important CVs were missed during the simulation setup.[103] To facilitate the choice of CVs, ML has been incorporated into Metadynamics simulations. Wang et al. developed a predictive information bottleneck (PIB) approach to identify CVs and predict biomolecular dissociation rates.[93] The PIB was tested on the system of benzene binding to T4L, and the predicted k_{off} value was 3.3 ± 0.8 s^{-1}, being consistent with other InMetaD simulations but needing much shorter simulations.[41] In another study, Filizola et al.[38] developed a novel approach, which combined InMetaD and ML methods including automatic mutual information noise omission and reweighted autoencoded variational Bayes to predict the dissociation kinetic rates of two drugs (morphine and bruprenorphine) in the μ-opioid receptor. Based on ~6 μs InMetaD simulations,

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

Table 3. Summary of Computer Simulation Predicted Peptide Binding (k_{on}^{sim}) and Dissociation (k_{off}^{sim}) Rates Compared with Experimentally Determined Binding (k_{on}^{exp}) and Dissociation (k_{off}^{exp}) Rates

system	method	k_{on}^{exp} (10^7 M^{-1} s^{-1})	k_{off}^{exp} (s^{-1})	k_{on}^{sim} (10^7 M^{-1} s^{-1})	k_{off}^{sim} (s^{-1})	sim. time (μs)	Δ log k_{on}	Δ log k_{off}	force field	year[ref]
MDM2/P53	InMetaD	0.92	2.06	0.43 ± 0.22	0.7 ± 0.4	27	0.88	−0.47	AMBER ff99SB-ILDN	2020[127]
MDM2/P53	MSM	0.92	2.06	0.019	2.5	831	0.88	0.08	AMBER ff99SB-ILDN-NMR	2017[129]
MDM2/P53	WE	0.92	2.06	7		120	0.88		AMBER ff99SB-ILDN	2016[128]
MDM2/PMI	MSM	52.7	0.037	330	0.125−1.13	500	0.80	0.53	AMBER ff99SB-ILDN	2017[131]
SH3-1CKB	Pep-GaMD	150	8900	4060 ± 2260	1450 ± 1170	3	1.43	−0.79	AMBER ff14SB	2020[130]

the predicted k_{off} for the morphine and bruprenorphine were 0.057 ± 0.005 s^{-1} and 0.021 ± 0.003 s^{-1}, respectively, being within 1 order of magnitude difference from experimental values of 0.0023 ± 0.001 s^{-1} and 0.0018 ± 0.03 s^{-1}. Very recently, Narjes et al.[87] combined ML and a novel Metadynamics approach, On-the-fly Probability Enhanced Sampling (OPES) flooding, to investigate the binding of benzamidine to trypsin. Based on a total of ∼2.74 μs OPES simulations, they captured 55 benzamidine unbinding events and predicted the k_{off} value of 1560 s^{-1}, being highly consistent with the experimental data.

Scaled MD[45,104,105] has been mainly used for the prediction of k_{off} as a scaling factor ranging from 0 to 1 is introduced in the simulations to reduce the energy barrier to facilitate ligand dissociation. For example, Schuetz et al.[47] performed scaled MD simulations to accurately predict the residence time and drug dissociation pathways of different inhibitors in Hsp90. In a recent study,[106] Bianciotto et al. applied scaled MD simulations to predict the residence time and ligand unbinding pathways for a set of 27 ligands of Hsp90, being highly consistent with experimental data. The same group reported another novel method based on adiabatic biased MD with an electrostatics-like collective variable (elABMD)[107] to explore the protein−ligand dissociation process. elABMD correctly ranked a series of ligands binding to glucokinase, being consistent with available experimental data. In the RAMD simulations, an additional random force is applied on the ligand to promote its movement. Similar to scaled MD, RAMD is mainly used in the ligand dissociation simulations to qualitatively predict dissociation rates. In one recent study, Nunes-Alves et al.[42] performed RAMD simulations to predict ligand dissociation rates of T4L. The predicted kinetic rates correlated well with experimental values for various systems with different ligands, temperatures, and protein mutations.

GaMD[28,29] is developed to apply a harmonic boost potential to enhance sampling with 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.[28,29] GaMD has been successfully applied to simulate important biomolecular processes, including protein/RNA folding,[29,108,109] ligand/protein/RNA binding,[108,110−116] and protein conformational changes.[115,117,118] However, it remained challenging to accurately predict ligand binding kinetic rates through normal GaMD.[28,119] Recently, a "selective GaMD" algorithm, called Ligand GaMD (LiGaMD),[88,92] has been developed to allow for more efficient sampling of ligand binding and dissociation processes, which thus allows to accurately predict the ligand binding kinetic rates. For the protein ligand binding system, the system contains ligand L, protein P, and the biological environment E. The system potential energy could be decomposed into the following terms

$$V(r) = V_{P,b}(r_P) + V_{L,b}(r_L) + V_{E,b}(r_E) + V_{PP,nb}(r_P) + V_{LL,nb}(r_L) + V_{EE,nb}(r_E) + V_{PL,nb}(r_{PL}) + V_{PE,nb}(r_{PE}) + V_{LE,nb}(r_{LE}) \quad (6)$$

where $V_{P,b}$, $V_{L,b}$, and $V_{E,b}$ are the bonded potential energies in protein P, ligand L, and environment E, respectively. $V_{PP,nb}$, $V_{LL,nb}$, and $V_{EE,nb}$ are the self-nonbonded potential energies in protein P, ligand L, and environment E, respectively. $V_{PL,nb}$, $V_{PE,nb}$, and $V_{LE,nb}$ are the nonbonded interaction energies between P-L, P-E, and L-E, respectively. Ligand binding mainly involves the nonbonded interaction energies of the ligand. Therefore, LiGaMD selectively boosts on the ligand essential energy term of $V_{ligand}(r) = V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. In order to facilitate ligand rebinding, another boost was added to the remaining potential interaction of the system. Repetitive binding and dissociation of small-molecule ligands were captured in the LiGaMD simulations of host−guest and protein−ligand binding model systems.[88] Repetitive guest binding and dissociation in the β-cyclodextrin host were observed in hundreds-of-nanoseconds LiGaMD simulations. Accelerations of ligand kinetic rates in LiGaMD simulations were properly estimated using Kramers' rate theory. Furthermore, microsecond LiGaMD simulations observed repetitive benzamidine binding and dissociation in trypsin. The benzamidine binding and dissociation rates were predicted to be 1.15 ± 0.79 × 10^7 $M^{-1}\cdot s^{-1}$ and 3.53 ± 1.41 s^{-1}, respectively. These data were comparable to the experimental values[120] of 2.9 × 10^7 $M^{-1}\cdot s^{-1}$ and 600 s^{-1}. Very recently, five replicas of 5 μs LiGaMD simulations successfully captured repetitive Nirmatrelvir drug binding and dissociation in the 3CLpro binding domain.[121] The Nirmatrelvir binding and dissociation rates were predicted to be 3.20 ± 0.21 × 10^5 $M^{-1}\cdot s^{-1}$ and 2.92 ± 0.37 × $10^3 s^{-1}$, respectively. As there were no available experimentally determined binding kinetic rates, the authors predicted the dissociation constant (k_D) from the predicted binding kinetic rates by equation $k_D = k_{off}/k_{on}$. Notably, the predicted k_D was 9.10 ± 0.29 nM, being highly consistent with the available experimental value of 7 ± 3 nM,[122] demonstrating high accuracy of the predicted binding kinetic rates from LiGaMD simulations. A newer version, LiGaMD2,[92] was recently developed, in which a selective boost potential was applied to both the ligand and protein residues in the binding pocket to improve sampling of ligand binding and dissociation. The predicted values of (k_{on}, k_{off})[123] in three complexes of BEN bound to the L99A T4L (T4L:L99A-BEN) and M102A T4L (T4L:M102A-BEN), and IND bound to the

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

Table 4. Summary of Computer Simulation Predicted Protein−Protein Binding (k_{on}^{sim}) and Dissociation (k_{off}^{sim}) Rates Compared with Experimentally Determined Binding (k_{on}^{exp}) and Dissociation (k_{off}^{exp}) Rates

system	method	k_{on}^{exp} (10^7 M^{-1} s^{-1})	k_{off}^{exp} (s^{-1})	k_{on}^{sim} (10^7 M^{-1} s^{-1})	k_{off}^{sim} (s^{-1})	sim. time (μs)	Δ log k_{on}	Δ log k_{off}	force field	year[ref]
barnase-barstar	PPI-GaMD	60	8×10^{-6}	217 ± 138	$7.32 \pm 4.95 \times 10^{-6}$	12	0.56	−0.038	AMBER ff14SB	2022[112]
barnase-barstar	WE	60	8×10^{-6}	230 ± 100		18	0.58		AMBER ff03*	2019[83]
barnase-barstar	cMD	60	8×10^{-6}	2.3		440	−1.42		AMBER ff99SB-ILDN	2019[67]
barnase-barstar	MSM	60	8×10^{-6}	26.3−26.5	3×10^{-6}	1700	−0.36	−0.42	AMBER ff99SB	2017[137]
insulin dimer	cMD	11.4	14800	0.41		294.8	−1.44		AMBER ff99SB-ILDN	2019[67]
Ras−Raf-RBD	cMD	4.5	7.4	2.6		117	−0.24		AMBER ff99SB-ILDN	2019[67]

L99A T4L (T4L:L99A-IND) were ($7.42 \pm 4.81 \times 10^6$ $M^{-1}\cdot s^{-1}$, 1441 ± 883 s^{-1}), ($9.57 \pm 6.29 \times 10^6$ $M^{-1}\cdot s^{-1}$, 2011 ± 1606 s^{-1}), and ($2.99 \pm 2.87 \times 10^6$ $M^{-1}\cdot s^{-1}$, 3494 ± 559 s^{-1}), being highly consistent with the corresponding experimental values of ($0.8-1.0 \times 10^6$ $M^{-1}\cdot s^{-1}$, 950 s^{-1}), ($3.0-5.0 \times 10^6$ $M^{-1}\cdot s^{-1}$, 3000), and ($0.7-1.0 \times 10^6$ $M^{-1}\cdot s^{-1}$, 325 s^{-1}), respectively.

Protein−Peptide Binding Kinetics. In comparison with the extensively studied protein-small molecule binding, protein−peptide binding studies are much less although an increasing number of peptide-based drugs are being licensed to market in recent years.[124−126] Large conformational changes of peptides often occur during binding to target proteins, bringing huge challenges for modeling.[77,127] For example, the coupled folding-upon binding mechanism has been observed in serval systems of peptide binding to proteins.[77,127] Only a few number of computational approaches have been implemented to predict peptide binding kinetic rates, including the InMetaD,[127] Weighted Ensemble,[128] MSM,[129] and Peptide GaMD (Pep-GaMD)[130] (Table 3).

InMetaD simulations with three CVs have successfully predicted the peptide binding and dissociation rates for the system of p53-MDM2.[127] Based on 27 μs InMetaD simulations,[127] the predicted values of (k_{on}, k_{off}) were ($0.43 \pm 0.22 \times 10^7$ M^{-1} s^{-1}, 0.7 ± 0.4 s^{-1}), being comparable to the corresponding experimental values of (0.92×10^7 M^{-1} s^{-1}, 2.06 s^{-1}). Weighted Ensemble of a total amount of ∼120 μs cMD simulations in implicit solvent was performed on the same p53-MDM2 system.[128] The predicted p53 binding kinetic rate (k_{on}) was 7 s^{-1}, being highly consistent with the experiential data of 2.06 s^{-1}. Built on a total of 831 μs cMD simulations of p53 binding to the MDM2, the MSM[129] predicted accurate values of k_{on} and k_{off} at 0.019×10^7 M^{-1} s^{-1} and 2.5 s^{-1}, respectively. However, the simulations needed for building MSM are much longer than the Weighted Ensemble and InMetaD simulations. Another MSM built on hundreds-of-microsecond cMD and Hamiltonian replica exchange simulations has been implemented to characterize binding and dissociation of the PMI peptide to the MDM2.[131] The PMI dissociation process is rather slow with the residence time at the time scale of second. Therefore, ∼50 μs Hamiltonian replica exchange simulations were performed to predict the dissociate rate. The predicted values of (k_{on}, k_{off}) were (300×10^7 M^{-1} s^{-1}, 0.125−1.13 s^{-1}), being comparable to the corresponding experimental values of (52.7×10^7 M^{-1} s^{-1}, 0.037 s^{-1}).

Based on GaMD, we recently developed an algorithm called peptide GaMD or "Pep-GaMD" that enhances sampling of protein−peptide interactions.[130] As above-mentioned, a large conformational change was involved in the process of peptide binding to target proteins.[77,127] Therefore, peptide binding involves both the bonded and nonbonded interaction energies of the peptide. Thus, the peptide essential potential energy is defined as $V_{peptide}(r) = V_{LL,b}(r_L) + V_{LL,nb}(r_L) + V_{PL,nb}(r_{PL}) + V_{LE,nb}(r_{LE})$. A selective boost was thus added to the peptide essential potential to facilitate the dissociation of peptides in the Pep-GaMD. In addition to selectively boosting the peptide, another boost potential is applied on the protein and solvent to enhance conformational sampling of the protein and facilitate peptide rebinding.

Pep-GaMD[130] has been developed to capture repetitive peptide binding and dissociation processes, which allows us to calculate the peptide binding free energies and kinetic rates. It has been demonstrated on binding of three model peptides to the SH3 domains,[132,133] including "PPPVPPRR" (PDB: 1CKB), "PPPALPPKK" (PDB: 1CKA), and "PAMPAR" (PDB: 1SSH). Repetitive peptide binding and unbinding events were captured in independent 1 μs Pep-GaMD simulations, allowing us to calculate peptide binding thermodynamics and kinetics. The predicted values of both binding free energies and kinetic rates from Pep-GaMD simulations were in good agreement with available experimental data. Particularly, the predicted peptide binding kinetic rates of 1CKB were ($4060 \pm 2260 \times 10^7$ $M^{-1}\cdot s^{-1}$, 1450 ± 1170 s^{-1}), being within 1 order of the experimental data of (150×10^7 $M^{-1}\cdot s^{-1}$, 8900 s^{-1}).

Protein−Protein Binding Kinetics. PPIs play key roles in many fundamental biological processes, including cellular signal transduction, immune responses, and so on.[1] Moreover, PPIs are implicated in the development of numerous human diseases and served as important drug targets.[134−136] PPIs exhibit unique features, being distinct from the protein-small molecule and protein−peptide interactions. The protein−protein binding affinity is often stronger than that of the protein-small molecule and protein−peptide interactions. Protein−protein binding and unbinding processes often occurred in significantly longer time scale. Particularly, the protein−protein dissociation process could take place in a much longer time scale, from seconds to even days. Tens of microseconds cMD simulations were able to capture barnase binding to barstar.[67] Based on 28 successfully binding events captured in a total of ∼213 μs Anton cMD simulations with the TIP4P2005 water model,[67] the predicted barnase binding

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

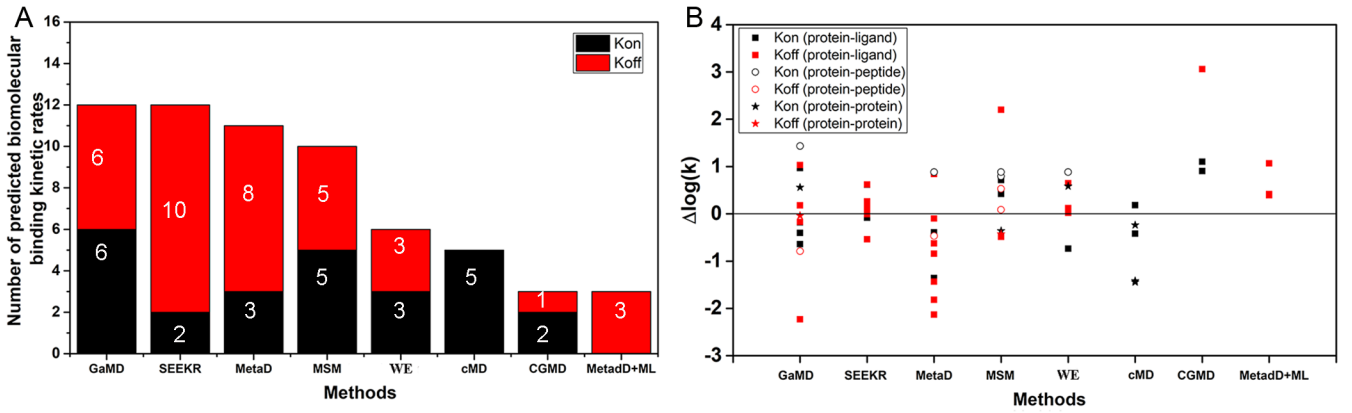


Figure 2. Number (A) and accuracy (B) of predicted biomolecular binding kinetic rates using different MD techniques, including Metadynamics (MetaD), Markov State Models (MSM), Gaussian accelerated MD (GaMD), conventional MD (cMD), Weighted Ensemble (WE), simulation enabled estimation of kinetic rates (SEEKR), coarse-grained MD (CGMD), and a combination of Metadynamics and Machine Learning (MetaD +ML). The protein−ligand, protein−peptide, and protein−protein complexes were plotted in squares, circles, and asterisks, respectively.

rate (k_{on}) was 6×10^7 M^{-1} s^{-1}, being in line with the experimental value of 60×10^7 M^{-1} s^{-1}. Fewer barnase binding events (24) with a slower predicted binding rate (2.3×10^7 M^{-1} s^{-1}) were observed with the the TIP3P water model (Table 4). Additionally, Pan et al.[67] successfully predicted the binding kinetic association rates of another two systems of insulin dimerization and Ras binding to the Ras-binding domain of c-RAF-1 (Ras-Raf-RBD). Based on 6 successful binding events among the total of 294.8 μs cMD simulations, the predicted association rate (k_{on}) of the insulin dimerization was 0.41×10^7 $M^{-1}\cdot s^{-1}$, being comparable to the experimental value of 11.4×10^7 $M^{-1}\cdot s^{-1}$ (Table 4). For the Ras-Raf-RBD system, 117 μs cMD simulations successfully captured 7 binding events and predicted the k_{on} value of 2.6×10^7 $M^{-1}\cdot s^{-1}$, being highly consistent with the experimental data of 4.5×10^7 $M^{-1}\cdot s^{-1}$ (Table 4). However, it remains challenging to simulate the protein dissociation with cMD.[67]

Weighted Ensemble[83] of a total of ~18 μs cMD simulations was able to capture 203 barnase binding events and accurately predict the barnase-barstar binding rate constant (k_{on}) of $23 \pm 10 \times 10^7$ $M^{-1}\cdot s^{-1}$ (Table 4). Plattner et al.[137] performed high throughput MD simulations of the barnase binding to barstar to build MSM. A total of 1700 μs cMD simulations with 1,892 independent replicas starting from an unbound state captured 74 barnase binding events. Another set of 300 μs adaptive MD simulations captured 16 and 10 times barnase binding and dissociation events, respectively. Based on the total of 2,000 μs simulation data, the obtained MSM was able to predict intermediate structures, binding energies, and kinetic rates that were consistent with experimental data[137] (Table 4).

Recently, we developed a selective PPI-GaMD method[112] to simulate repetitive protein binding and dissociation in order to calculate protein binding free energies and kinetics. The PPI simulation system consists of a ligand protein *L*, a target protein *P*, and a biological environment *E*. In PPI-GaMD, a selective boost potential is added to the nonbonded protein−protein interaction energy $V_{PL,nb}$. Another boost potential is applied on the remaining potential energy of the system to enhance conformational sampling of the proteins and facilitate protein diffusion and rebinding.[112] PPI-GaMD[112] has been demonstrated on the model system of barnase binding to the barstar. Six independent 2 μs PPI-GaMD simulations have successfully captured repetitive barstar dissociation and rebinding events. Three to six binding and dissociation events were observed in each individual PPI-GaMD simulation. The barnase 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.[136] Additionally, the PPI-GaMD simulations allowed us to calculate the protein binding kinetics. The average k_{on} and k_{off} were predicted as $21.7 \pm 13.8 \times 10^8$ $M^{-1}\cdot s^{-1}$ and $7.32 \pm 4.95 \times 10^{-6}$ s^{-1}, being consistent with the corresponding experimental values of 6.0×10^8 $M^{-1}\cdot s^{-1}$ and 8.0×10^{-6} s^{-1}, respectively (Table 4).

5. CONCLUSIONS AND OUTLOOK

Both experimental and computational techniques have achieved remarkable advances in characterizing biomolecular binding kinetics, including SPR, QSKR, MD, and enhanced sampling simulations. It is still very expensive and resource-consuming for experimental techniques to obtain biomolecular binding kinetic rates. Nevertheless, recent years have seen increasing numbers of experimental binding kinetic data, leading to a number of databases to collect such information.

Based on the experimental binding kinetic data, QSKRs have been developed to predict binding kinetic rate constants with high throughtput.[60] For MD simulations, accuracy of binding free energy calculations could be within 1.0 kcal/mol with the modern techniques.[138] Compared with extensively studied biomolecular binding thermodynamics, the accuracy and efficacy of modeling techniques for predicting biomolecular binding kinetics are still not very high. The predicted binding kinetic rate constants from MD simulations and related enhanced sampling methods could derivate orders of magnitude from the experimental data (Tables 2−4 and Figure 1B). Nevertheless, MD simulations have enabled characterization of biomolecular binding pathways and kinetics, attracting increasing attention in recent years. Enhanced sampling methods have greatly reduced the computational cost for calculations of biomolecular kinetics. Among various enhanced sampling methods, the MSM, Weighted Ensemble, Metadynamics, GaMD, and SEEKR appear to be the most used techniques that allow for simultaneous predictions of biomolecular binding association and dissociation rates (Figure 2). Higher sampling efficiency could be generally obtained using the Metadynamics with well

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148

predefined CVs than the using the CV-free methods including MSM, Weighted Ensemble, and GaMD. However, it is often challenging to predefine good CVs in Metadynamics simulations of complex biological systems especially for large biomolecular interactions. In case important CVs are missing during the simulation setup, Metadynamics simulations could suffer from the "hidden energy barrier" problem and still slow sampling convergence.[103] It is rather difficult to directly compare predictive accuracy among different methods as different force fields and systems were used. Nevertheless, the trypsin-benzamidine system is the most widely used system for benchmarking different methods (Table 1). The ranks of the accuracy in predicting k_{off} are Weighted Ensemble combined with milestone, SEEKR2, InMetaD combined with Machine Learning, SEEKR, Weighted Ensemble, InMetaD, MSM, LiGaMD, and CGMD. On the other hand, the methods providing accurate predictions of k_{on} are ranked as SEEKR2, InMetaD, LiGaMD, SEEKR, MSM, Weighted Ensemble combined with milestone, and CGMD. Apart from conformational sampling, the force field could also affect the prediction accuracy. For example, an ∼10-fold faster binding association rate (k_{on}) of the barnase to barstar was obtained using the TIP4P2015 water model than using the TIP3P water model.[67] The polarizable force field was able to generally improve the accuracy of the force field.[139] Another trend is the incorporation of ML into enhanced sampling methods to further improve sampling efficiency and prediction accuracy of biomolecular binding kinetic rates.[38,140,141] For example, the combination of InMetaD and ML decreased the prediction error $\Delta \log(k_{off})$ from 0.84[91] to 0.42[87] for the trypsin-benzamidine system using the same force field of AMBER ff14SB and GAFF. Overall, current computational methods have been tested mostly on very few model systems with published experimental kinetic data in the literature. The simulation protocols could be potentially calibrated to predict the kinetic rate constants against the experimental values. Future developments of enhanced sampling methods and force fields are still needed for more accurate predictions of biomolecular binding kinetics. This would suggest a need for the community to organize blind challenges of biomolecular binding kinetics predictions, in which participants predict the kinetic rates without knowing the experimental values and the predictions will be evaluated independently by the challenge organizers. Such challenges are expected to greatly facilitate improvements of the various techniques developed for predicting biomolecular binding kinetics in the field. In addition to protein−ligand binding, protein−peptide binding, and protein−protein interactions, interactions of nucleic acids (RNA and DNA) with small molecules and proteins remain largely underexplored and warrant more kinetics studies.

In summary, accurate calculations of biomolecular binding kinetics of large biomolecular complexes present grand challenges for computational modeling and enhanced sampling simulations. Further innovations in both computing hardware and method developments may help us to address these challenges in the future.

AUTHOR INFORMATION

Corresponding Author

Yinglong Miao − *Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States;* orcid.org/0000-0003-3714-1395; Email: miao@ku.edu

Authors

Jinan Wang − *Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States;* orcid.org/0000-0003-0162-212X

Hung N. Do − *Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States*

Kushal Koirala − *Center for Computational Biology and Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas 66047, United States*

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.2c01085

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

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 the National Science Foundation grant number ACI-1548562, the PSC Bridgs-2 GPU at the Pittsburgh Supercomputing Center through allocation [BIO220137] from the Advanced Cyberinfrastructure Coordination Ecosystem:Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296, 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) and the National Science Foundation (2121063).

REFERENCES

(1) Nooren, I. M.; Thornton, J. M. Diversity of protein−protein interactions. *EMBO J.* **2003**, *22* (14), 3486−3492.

(2) Núñez, S.; Venhorst, J.; Kruse, C. G. Target−drug interactions: first principles and their application to drug discovery. *Drug discovery today* **2012**, *17* (1−2), 10−22.

(3) Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery* **2007**, *6* (3), 211−219.

(4) Klebe, G. Applying thermodynamic profiling in lead finding and optimization. *Nat. Rev. Drug Discovery* **2015**, *14* (2), 95−110.

(5) 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.; et al. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today* **2017**, *22* (6), 896−911.

(6) Tonge, P. J. Drug−target kinetics in drug discovery. *ACS Chem. Neurosci.* **2018**, *9* (1), 29−39.

(7) Ahmad, K.; Rizzi, A.; Capelli, R.; Mandelli, D.; Lyu, W.; Carloni, P. Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective. *Front. Mol. Biosci.* **2022**, *9*, 899805.

(8) Ijzerman, A. P.; Guo, D. Drug−Target Association Kinetics in Drug Discovery. *Trends Biochem. Sci.* **2019**, *44* (10), 861−871.

(9) Ferruz, N.; De Fabritiis, G. Binding Kinetics in Drug Discovery. *Mol. Inf.* **2016**, *35* (6−7), 216−226.

(10) Zhou, Y.; Fu, Y.; Yin, W.; Li, J.; Wang, W.; Bai, F.; Xu, S.; Gong, Q.; Peng, T.; Hong, Y.; Zhang, D.; Zhang, D.; Liu, Q.; Xu, Y.;

Xu, H. E.; Zhang, H.; Jiang, H.; Liu, H. Kinetics-Driven Drug Design Strategy for Next-Generation Acetylcholinesterase Inhibitors to Clinical Candidate. *J. Med. Chem.* **2021**, *64* (4), 1844−1855.
(11) Holdgate, G. A.; Gill, A. L. Kinetic efficiency: the missing metric for enhancing compound quality? *Drug Discovery Today* **2011**, *16* (21), 910−913.
(12) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5* (9), 730−739.
(13) Sohraby, F.; Nunes-Alves, A. Advances in computational methods for ligand binding kinetics. *Trends Biochem. Sci.* **2022**, DOI: 10.1016/j.tibs.2022.11.003.
(14) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *Annu. Rev. Phys. Chem.* **2019**, *70* (1), 143−171.
(15) Lu, H.; Tonge, P. J. Drug−target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.* **2010**, *14* (4), 467−474.
(16) Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New approaches for computing ligand−receptor binding kinetics. *Curr. Opin. Struct. Biol.* **2018**, *49*, 1−10.
(17) Dahl, G.; Akerud, T. Pharmacokinetics and the drug−target residence time concept. *Drug Discovery Today* **2013**, *18* (15), 697−707.
(18) Copeland, R. A. Drug−target interaction kinetics: underutilized in drug optimization? *Future Med. Chem.* **2016**, *8* (18), 2173−2175.
(19) Copeland, R. A. The drug−target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discovery* **2016**, *15* (2), 87−95.
(20) Decherchi, S.; Cavalli, A. Thermodynamics and Kinetics of Drug-Target Binding by Molecular Simulation. *Chem. Rev.* **2020**, *120* (23), 12788−12833.
(21) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. *J. Chem. Phys.* **2019**, *150* (24), 244112.
(22) Lotz, S. D.; Dickson, A. Wepy: a flexible software framework for simulating rare events with weighted ensemble resampling. *ACS Omega* **2020**, *5* (49), 31608−31623.
(23) Nunes-Alves, A.; Zuckerman, D. M.; Arantes, G. M. Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophys. J.* **2018**, *114* (5), 1058−1066.
(24) Narayan, B.; Buchete, N.-V.; Elber, R. Computer Simulations of the Dissociation Mechanism of Gleevec from Abl Kinase with Milestoning. *J. Phys. Chem. B* **2021**, *125* (22), 5706−5715.
(25) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *J. Chem. Theory Comput.* **2020**, *16* (8), 5348−5357.
(26) Votapka, L. W.; Amaro, R. E. Multiscale estimation of binding kinetics using Brownian dynamics, molecular dynamics and milestoning. *PLoS Comput. Biol.* **2015**, *11* (10), e1004381.
(27) Ojha, A. A.; Srivastava, A.; Votapka, L. W.; Amaro, R. E. Selectivity and ranking of tight-binding JAK-STAT inhibitors using Markovian milestoning with Voronoi tessellations. *bioRxiv* **2022**, DOI: 10.1101/2022.11.10.516058.
(28) Wang, J.; Arantes, P. R.; Bhattarai, A.; Hsu, R. V.; Pawnikar, S.; Huang, Y. M.; Palermo, G.; Miao, Y. Gaussian accelerated molecular dynamics (GaMD): principles and applications. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2021**, *11* (5), e1521.
(29) Miao, Y.; Feher, V. A.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *J. Chem. Theory Comput.* **2015**, *11* (8), 3584−3595.
(30) Limongelli, V.; Bonomi, M.; Parrinello, M. Funnel metadynamics as accurate binding free-energy method. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110* (16), 6358−6363.
(31) Tiwary, P.; Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **2013**, *111* (23), 230602.
(32) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *J. Chem. Phys.* **2018**, *149* (7), 072309.
(33) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K. Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chem. Sci.* **2017**, *8* (9), 6466−6473.
(34) Banerjee, P.; Bagchi, B. Dynamical control by water at a molecular level in protein dimer association and dissociation. *Proc. Natl. Acad. Sci. U.S.A.* **2020**, *117* (5), 2302−2308.
(35) Antoszewski, A.; Feng, C.-J.; Vani, B. P.; Thiede, E. H.; Hong, L.; Weare, J.; Tokmakoff, A.; Dinner, A. R. Insulin dissociates by diverse mechanisms of coupled unfolding and unbinding. *J. Phys. Chem. B* **2020**, *124* (27), 5571−5587.
(36) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112* (5), E386−E391.
(37) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. *J. Am. Chem. Soc.* **2017**, *139*, 4780.
(38) Lamim Ribeiro, J. M.; Provasi, D.; Filizola, M. A combination of machine learning and infrequent metadynamics to efficiently predict kinetic rates, transition states, and molecular determinants of drug dissociation from G protein-coupled receptors. *J. Chem. Phys.* **2020**, *153* (12), 124105.
(39) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108* (25), 10184−10189.
(40) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6* (1), 7653.
(41) Mondal, J.; Ahalawat, N.; Pandit, S.; Kay, L. E.; Vallurupalli, P. Atomic resolution mechanism of ligand binding to a solvent inaccessible cavity in T4 lysozyme. *PLoS Comput. Biol.* **2018**, *14* (5), e1006180.
(42) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τRAMD simulations. *Curr. Res. Struct. Biol.* **2021**, *3*, 106−111.
(43) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of Drug-Target Residence Times by τ-Random Acceleration Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2018**, *14* (7), 3859−3869.
(44) Wang, T.; Duan, Y. Chromophore channeling in the G-protein coupled receptor rhodopsin. *J. Am. Chem. Soc.* **2007**, *129* (22), 6970−6971.
(45) Sinko, W.; Miao, Y.; de Oliveira, C. s. A. F.; McCammon, J. A. Population based reweighting of scaled molecular dynamics. *J. Phys. Chem. B* **2013**, *117* (42), 12759−12768.
(46) Deb, I.; Frank, A. T. Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. *J. Chem. Theory Comput.* **2019**, *15* (11), 5817−5828.
(47) Schuetz, D. A.; Bernetti, M.; Bertazzo, M.; Musil, D.; Eggenweiler, H. M.; Recanatini, M.; Masetti, M.; Ecker, G. F.; Cavalli, A. Predicting Residence Time And Drug Unbinding Pathway Through Scaled Molecular Dynamics. *J. Chem. Inf. Model* **2019**, *59*, 535.
(48) Gleitsman, K. R.; Sengupta, R. N.; Herschlag, D. Slow molecular recognition by RNA. *Rna* **2017**, *23* (12), 1745−1753.
(49) Kumar, M.; Lowery, R. G. A High-Throughput Method for Measuring Drug Residence Time Using the Transcreener ADP Assay. *SLAS DISCOVERY: Advancing Life Sciences R&D* **2017**, *22* (7), 915−922.
(50) Hulme, E. C.; Trevethick, M. A. Ligand binding assays at equilibrium: validation and interpretation. *Br. J. Pharmacol.* **2010**, *161* (6), 1219−1237.

(51) Guo, D.; Hillger, J. M.; IJzerman, A. P.; Heitman, L. H. Drug-target residence time—a case for G protein-coupled receptors. *Med. Res. Rev.* **2014**, *34* (4), 856−892.
(52) Calebiro, D.; Koszegi, Z.; Lanoiselee, Y.; Miljus, T.; O'Brien, S. G protein-coupled receptor-G protein interactions: a single-molecule perspective. *Physiol. Rev.* **2021**, *101* (3), 857−906.
(53) Ji, Z. L.; Chen, X.; Zhen, C. J.; Yao, L. X.; Han, L. Y.; Yeo, W. K.; Chung, P. C.; Puy, H. S.; Tay, Y. T.; Muhammad, A.; Chen, Y. Z. KDBI: Kinetic Data of Bio-molecular Interactions database. *Nucleic Acids Res.* **2003**, *31* (1), 255−257.
(54) Gilson, M. K.; Liu, T.; Baitaluk, M.; Nicola, G.; Hwang, L.; Chong, J. BindingDB in 2015: A public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.* **2016**, *44* (D1), D1045−D1053.
(55) Norval, L. W.; Krämer, S. D.; Gao, M.; Herz, T.; Li, J.; Rath, C.; Wöhrle, J.; Günther, S.; Roth, G. KOFFI and Anabel 2.0—a new binding kinetics database and its integration in an open-source binding analysis software. *Database* **2019**, *2019*, baz101.
(56) Liu, H.; Su, M.; Lin, H.-X.; Wang, R.; Li, Y. Public Data Set of Protein−Ligand Dissociation Kinetic Constants for Quantitative Structure−Kinetics Relationship Studies. *ACS Omega* **2022**, *7* (22), 18985−18996.
(57) Moal, I. H.; Fernández-Recio, J. SKEMPI: a Structural Kinetic and Energetic database of Mutant Protein Interactions and its use in empirical models. *Bioinformatics* **2012**, *28* (20), 2600−2607.
(58) Liu, Q.; Chen, P.; Wang, B.; Zhang, J.; Li, J. dbMPIKT: a database of kinetic and thermodynamic mutant protein interactions. *BMC Bioinf.* **2018**, *19* (1), 455.
(59) 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.; Ijzerman, A. P.; Müller-Fahrnow, A.; Ecker, G. F. Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today* **2017**, *22* (6), 896−911.
(60) Nunes-Alves, A.; Ormersbach, F.; Wade, R. C. Prediction of the Drug−Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis. *J. Chem. Inf. Model.* **2021**, *61* (7), 3708−3721.
(61) Ganotra, G. K.; Wade, R. C. Prediction of Drug−Target Binding Kinetics by Comparative Binding Energy Analysis. *ACS Med. Chem. Lett.* **2018**, *9* (11), 1134−1139.
(62) Schuetz, D. A.; Richter, L.; Martini, R.; Ecker, G. F. A structure−kinetic relationship study using matched molecular pair analysis. *RSC Med. Chem.* **2020**, *11* (11), 1285−1294.
(63) Chiu, S. H.; Xie, L. Toward High-Throughput Predictive Modeling of Protein Binding/Unbinding Kinetics. *J. Chem. Inf. Model.* **2016**, *56* (6), 1164−1174.
(64) Lamprakis, C.; Andreadelis, I.; Manchester, J.; Velez-Vega, C.; Duca, J. S.; Cournia, Z. Evaluating the efficiency of the Martini force field to study protein dimerization in aqueous and membrane environments. *J. Chem. Theory Comput.* **2021**, *17* (5), 3088−3102.
(65) He, Z.; Paul, F.; Roux, B. A critical perspective on Markov state model treatments of protein−protein association using coarse-grained simulations. *J. Chem. Phys.* **2021**, *154* (8), 084101.
(66) Basdevant, N.; Borgis, D.; Ha-Duong, T. Modeling protein−protein recognition in solution using the coarse-grained force field SCORPION. *J. Chem. Theory Comput.* **2013**, *9* (1), 803−813.
(67) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E. Atomic-level characterization of protein−protein association. *Proc. Natl. Acad. Sci. U.S.A.* **2019**, *116*, 4244.
(68) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9* (9), 646−652.
(69) Tang, Z.; Roberts, C. C.; Chang, C.-e. A. Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. *Front. Biosci.-Landmark* **2017**, *22* (6), 960−981.
(70) Hollingsworth, S. A.; Dror, R. O. Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129−1143.
(71) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D. ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *J. Chem. Theory Comput.* **2009**, *5* (6), 1632−1639.
(72) Johnston, J. M.; Filizola, M. Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. *Curr. Opin. Struct. Biol.* **2011**, *21* (4), 552−558.
(73) 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.; et al. Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330* (6002), 341−346.
(74) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S. To milliseconds and beyond: challenges in the simulation of protein folding. *Curr. Opin. Struct. Biol.* **2013**, *23* (1), 58−65.
(75) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T. In Anton 3: twenty microseconds of molecular dynamics simulation before lunch. *SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; New York, NY, 2021; pp 1−11, DOI: 10.1145/3458817.3487397.
(76) Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How Does a Drug Molecule Find Its Target Binding Site? *J. Am. Chem. Soc.* **2011**, *133* (24), 9181−9183.
(77) Robustelli, P.; Piana, S.; Shaw, D. E. Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *J. Am. Chem. Soc.* **2020**, *142* (25), 11092−11101.
(78) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug−receptor binding kinetics. *Drug Discovery Today* **2013**, *18* (13), 667−673.
(79) Wang, J.; Bhattarai, A.; Do, H. N.; Miao, Y. Challenges and frontiers of computational modelling of biomolecular recognition. *QRB Discovery* **2022**, *3*, e13.
(80) Kamenik, A. S.; Linker, S. M.; Riniker, S. Enhanced sampling without borders: on global biasing functions and how to reweight them. *Phys. Chem. Chem. Phys.* **2022**, *24* (3), 1225−1236.
(81) Wong, C. F. Molecular simulation of drug-binding kinetics. *Mol. Simul.* **2014**, *40* (10−11), 889−903.
(82) Kappel, K.; Miao, Y. L.; McCammon, J. A. Accelerated molecular dynamics simulations of ligand binding to a muscarinic G-protein-coupled receptor. *Q. Rev. Biophys.* **2015**, *48* (4), 479−487.
(83) Saglam, A. S.; Chong, L. T. Protein−protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chem. Sci.* **2019**, *10*, 2360−2372.
(84) Sohraby, F.; Javaheri Moghadam, M.; Aliyar, M.; Aryapour, H. A boosted unbiased molecular dynamics method for predicting ligands binding mechanisms: probing the binding pathway of dasatinib to Src-kinase. *Bioinformatics* **2020**, *36* (18), 4714−4720.
(85) Ray, D.; Stone, S. E.; Andricioaei, I. Markovian Weighted Ensemble Milestoning (M-WEM): Long-Time Kinetics from Short Trajectories. *J. Chem. Theory Comput.* **2022**, *18* (1), 79−95.
(86) Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. *J. Chem. Inf. Model.* **2022**, *62* (13), 3253−3262.
(87) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of Benzamidine in Trypsin. *Nat. Commun.* **2022**, *13* (1), 5438.
(88) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *J. Chem. Theory Comput.* **2020**, *16* (9), 5526−5547.
(89) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin−Benzamidine Binding. *J. Phys. Chem. B* **2017**, *121* (15), 3597−3606.

(90) Dandekar, B. R.; Mondal, J. Capturing Protein−Ligand Recognition Pathways in Coarse-Grained Simulation. *J. Phys. Chem. Lett.* **2020**, *11* (13), 5302−5311.
(91) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the Calculation of Protein−Ligand Binding Free Energy and Residence Times Using Dynamically Optimized Collective Variables. *J. Chem. Theory Comput.* **2019**, *15* (1), 743−750.
(92) Wang, J.; Miao, Y. Ligand Gaussian Accelerated Molecular Dynamics 2 (LiGaMD2): Improved Calculations of Ligand Binding Thermodynamics and Kinetics with Closed Protein Pocket. *J. Chem. Theory Comput.* **2023**, *19*, 733.
(93) Wang, Y.; Ribeiro, J. M. L.; Tiwary, P. Past−future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nat. Commun.* **2019**, *10* (1), 3573.
(94) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein−ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11* (1), 3714.
(95) Haldar, S.; Comitani, F.; Saladino, G.; Woods, C.; van der Kamp, M. W.; Mulholland, A. J.; Gervasio, F. L. A Multiscale Simulation Approach to Modeling Drug-Protein Binding Kinetics. *J. Chem. Theory Comput.* **2018**, *14* (11), 6093−6101.
(96) Tiwary, P.; Mondal, J.; Berne, B. J. How and when does an anticancer drug leave its binding site? *Science Advances* **2017**, *3* (5), e1700014.
(97) Capelli, R.; Lyu, W.; Bolnykh, V.; Meloni, S.; Olsen, J. M. H.; Rothlisberger, U.; Parrinello, M.; Carloni, P. Accuracy of Molecular Simulation-Based Predictions of koff Values: A Metadynamics Study. *J. Phys. Chem. Lett.* **2020**, *11* (15), 6373−6381.
(98) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A.; et al. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods* **2021**, *18* (4), 382−388.
(99) Huang, Y.-m. M. Multiscale computational study of ligand binding pathways: Case of p38 MAP kinase and its inhibitors. *Biophys. J.* **2021**, *120* (18), 3881−3892.
(100) Elber, R. Milestoning: An efficient approach for atomically detailed simulations of kinetics in biophysics. *Annu. Rev. Biophys.* **2020**, *49*, 69−85.
(101) Abrams, C.; Bussi, G. Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* **2014**, *16* (1), 163−199.
(102) Zuckerman, D. M. Equilibrium sampling in biomolecular simulation. *Annu. Rev. Biophys.* **2011**, *40*, 41.
(103) Bešker, N.; Gervasio, F. L. Using metadynamics and path collective variables to study ligand binding and induced conformational transitions. In *Computational drug discovery and design*; Springer: 2012; pp 501−513, DOI: 10.1007/978-1-61779-465-0_29.
(104) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.* **2015**, *5* (1), 11539.
(105) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. Molecular Dynamics Simulations and Kinetic Measurements to Estimate and Predict Protein−Ligand Residence Times. *J. Med. Chem.* **2016**, *59* (15), 7167−7176.
(106) Bianciotto, M.; Gkeka, P.; Kokh, D. B.; Wade, R. C.; Minoux, H. Contact Map Fingerprints of Protein−Ligand Unbinding Trajectories Reveal Mechanisms Determining Residence Times Computed from Scaled Molecular Dynamics. *J. Chem. Theory Comput.* **2021**, *17* (10), 6522−6535.
(107) Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Girotto, S.; Decherchi, S.; Cavalli, A. Investigating Drug−Target Residence Time in Kinases through Enhanced Sampling Simulations. *J. Am. Chem. Soc.* **2019**, *15* (8), 4646−4659.
(108) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian accelerated molecular dynamics in NAMD. *J. Chem. Theory Comput.* **2017**, *13* (1), 9−19.
(109) Copeland, M. M.; Do, H. N.; Votapka, L.; Joshi, K.; Wang, J.; Amaro, R. E.; Miao, Y. Gaussian Accelerated Molecular Dynamics in OpenMM. *J. Phys. Chem. B* **2022**, *126*, 5810.
(110) Chuang, C.-H.; Chiou, S.-j.; Cheng, T.-L.; Wang, Y.-T. A molecular dynamics simulation study decodes the Zika virus NS5 methyltransferase bound to SAH and RNA analogue. *Sci. Rep.* **2018**, *8* (1), 6336.
(111) Miao, Y.; Huang, Y.-m. M.; Walker, R. C.; McCammon, J. A.; Chang, C.-e. A. Ligand binding pathways and conformational transitions of the HIV protease. *Biochemistry* **2018**, *57* (9), 1533−1541.
(112) Wang, J.; Miao, Y. Protein-Protein Interaction-Gaussian Accelerated Molecular Dynamics (PPI-GaMD): Characterization of Protein Binding Thermodynamics and Kinetics. *J. Chem. Theory Comput.* **2022**, *18* (3), 1275−1285.
(113) Wang, Y.-T.; Chan, Y.-H. Understanding the molecular basis of agonist/antagonist mechanism of human mu opioid receptor through gaussian accelerated molecular dynamics method. *Sci. Rep.* **2017**, *7* (1), 7828.
(114) Liao, J.-M.; Wang, Y.-T. In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. *J. Biomol. Struct. Dyn.* **2019**, *37* (1), 166−177.
(115) Miao, Y.; McCammon, J. A. Graded activation and free energy landscapes of a muscarinic G-protein−coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113* (43), 12162−12167.
(116) Miao, Y.; Caliman, A. D.; McCammon, J. A. Allosteric effects of sodium ion binding on activation of the m3 muscarinic g-protein-coupled receptor. *Biophys. J.* **2015**, *108* (7), 1796−1806.
(117) Salawu, E. O. The impairment of torsinA's binding to and interactions with its activator: an atomistic molecular dynamics study of primary dystonia. *Front. Mol. Biosci.* **2018**, *5*, 64.
(118) Zhang, J.; Wang, N.; Miao, Y.; Hauser, F.; McCammon, J. A.; Rappel, W.-J.; Schroeder, J. I. Identification of SLAC1 anion channel residues required for CO2/bicarbonate sensing and regulation of stomatal movements. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (44), 11129−11137.
(119) Miao, Y.; McCammon, J. A. Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proc. Natl. Acad. Sci. U.S.A.* **2018**, *115* (12), 3036−3041.
(120) Guillain, F.; Thusius, D. Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J. Am. Chem. Soc.* **1970**, *92* (18), 5534−5536.
(121) Wang, Y.-T.; Liao, J.-M.; Lin, W.-W.; Li, C.-C.; Huang, B.-C.; Cheng, T.-L.; Chen, T.-C. Structural insights into Nirmatrelvir (PF-07321332)-3C-like SARS-CoV-2 protease complexation: a ligand Gaussian accelerated molecular dynamics study. *Phys. Chem. Chem. Phys.* **2022**, *24* (37), 22898−22904.
(122) Kneller, D. W.; Li, H.; Phillips, G.; Weiss, K. L.; Zhang, Q.; Arnould, M. A.; Jonsson, C. B.; Surendranathan, S.; Parvathareddy, J.; Blakeley, M. P.; Coates, L.; Louis, J. M.; Bonnesen, P. V.; Kovalevsky, A. Covalent narlaprevir- and boceprevir-derived hybrid inhibitors of SARS-CoV-2 main protease. *Nat. Commun.* **2022**, *13* (1), 2268.
(123) Feher, V. A.; Baldwin, E. P.; Dahlquist, F. W. Access of ligands to cavities within the core of a protein is rapid. *Nat. Struct. Biol.* **1996**, *3* (6), 516−521.
(124) Fosgerau, K.; Hoffmann, T. Peptide therapeutics: current status and future directions. *Drug Discovery Today* **2015**, *20* (1), 122−128.
(125) Ahrens, V. M.; Bellmann-Sickert, K.; Beck-Sickinger, A. G. Peptides and peptide conjugates: therapeutics on the upward path. *Future Med. Chem.* **2012**, *4* (12), 1567−1586.
(126) Lee, A. C.-L.; Harris, J. L.; Khanna, K. K.; Hong, J.-H. A Comprehensive Review on Current Advances in Peptide Drug Development and Design. *Int. J. Mol. Sci.* **2019**, *20* (10), 2383.
(127) Zou, R.; Zhou, Y.; Wang, Y.; Kuang, G.; Ågren, H.; Wu, J.; Tu, Y. Free Energy Profile and Kinetics of Coupled Folding and Binding

of the Intrinsically Disordered Protein p53 with MDM2. *J. Chem. Inf. Model.* **2020**, *60* (3), 1551−1558.
(128) Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T. Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein−Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *J. Phys. Chem. Lett.* **2016**, *7* (17), 3440−3445.
(129) Zhou, G.; Pantelopulos, G. A.; Mukherjee, S.; Voelz, V. A. Bridging Microscopic and Macroscopic Mechanisms of p53-MDM2 Binding with Kinetic Network Models. *Biophys. J.* **2017**, *113* (4), 785−793.
(130) Wang, J.; Miao, Y. Peptide Gaussian accelerated molecular dynamics (Pep-GaMD): Enhanced sampling and free energy and kinetics calculations of peptide binding. *J. Chem. Phys.* **2020**, *153* (15), 154109.
(131) Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F. Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* **2017**, *8* (1), 1095.
(132) Ahmad, M.; Helms, V. How do proteins associate? A lesson from SH3 domain. *Chem. Cent. J.* **2009**, *3* (S1), O22.
(133) Ball, L. J.; Kühne, R.; Schneider-Mergener, J.; Oschkinat, H. Recognition of proline-rich motifs by protein−protein-interaction domains. *Angew. Chem., Int. Ed.* **2005**, *44* (19), 2852−2869.
(134) Ferreira, L. G.; Oliva, G.; Andricopulo, A. D. Protein-protein interaction inhibitors: advances in anticancer drug design. *Expert Opin. Drug Discovery* **2016**, *11* (10), 957−968.
(135) Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J. Small molecules, big targets: drug discovery faces the protein−protein interaction challenge. *Nat. Rev. Drug Discovery* **2016**, *15* (8), 533−550.
(136) Schreiber, G.; Fersht, A. R. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993**, *32* (19), 5145−5150.
(137) Plattner, N.; Doerr, S.; De Fabritiis, G.; Noé, F. Complete protein−protein association kinetics in atomic detail revealed by molecular dynamics simulations and Markov modelling. *Nat. Chem.* **2017**, *9*, 1005.
(138) Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by Way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137* (7), 2695−2703.
(139) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable Force Fields for Biomolecular Simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48* (1), 371−394.
(140) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C. Machine Learning Analysis of τRAMD Trajectories to Decipher Molecular Determinants of Drug-Target Residence Times. *Front. Mol. Biosci.* **2019**, *6*, 36.
(141) Lamim Ribeiro, J. M.; Tiwary, P. Toward Achieving Efficient and Accurate Ligand-Protein Unbinding with Deep Learning and Molecular Dynamics through RAVE. *J. Chem. Theory Comput.* **2019**, *15* (1), 708−719.

https://doi.org/10.1021/acs.jctc.2c01085
J. Chem. Theory Comput. 2023, 19, 2135−2148