

Predicting biomolecular binding kinetics: A review

Jinan Wang¹, Hung N. Do¹, Kushal Koirala¹, Yinglong Miao^{1,*}

¹Center for Computational Biology and Department of Molecular Biosciences, University of Kansas,

Lawrence, Kansas 66047

*To whom correspondence should be addressed: miao@ku.edu

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, 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 have 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 including the protein-small molecule, protein-peptide and protein-protein/antibody structures. Biomolecular binding plays a key role in many fundamental biological processes¹. Accurate characterization of biomolecular binding thermodynamics and kinetics is key for therapeutic design². Recently, drug residence time or dissociation rate appears to correlate with drug efficacy better than the binding free energy³. With remarkable theoretical and technical developments, increasing numbers of experimental and computational methods are available for calculating the biomolecular binding kinetic rates^{3a, 3e, 3h, 4}. 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, with focus on the Molecular Dynamics (MD) and enhanced sampling methods, which have emerged as rapidly evolving techniques for studying biomolecular binding kinetics.

2. Available experimental techniques to measure binding kinetics

Most experimental techniques⁵ 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^{4b}. 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, ligands are tagged to follow the time course of their binding to targets, thus allowing for the spontaneous measurement of binding kinetic rates⁶. 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 processes⁷. The fluorescent resonance energy transfer (FRET) is one popular spectroscopy based approach⁷. 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^{2a}.

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)⁸, BindingDB⁹, kinetics of featured interactions (KOFFI)¹⁰, PDBbind¹¹, structural database of kinetics and energetics of mutant protein interactions (SKEMPI)¹², kinetic and thermodynamic database of mutant protein interactions (dbMPIKT)¹³ and so on^{3d, 14}.

KDBI⁸ 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 distinguishing biomolecular pathways. The binding kinetic data includes protein-protein/DNA/RNA/ligand and ligand-DNA/RNA interactions. BindingDB⁹ is one widely used database for exploring protein-small molecule interactions, containing ~1.1 million compounds

and 8.9 thousand 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 literatures and other databases such as PubChem, CheEMBL, PDSP Ki, and CSAR. Additionally, BindingDB provides an option for experimentalists to directly deposit their data. KOFFI¹⁰ is developed to provide binding kinetic rates along with experimental protocol. It includes 1705 individual entries. Notably, it contains a rating system to assess 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. PDBbind¹¹ was initially developed for collecting binding affinity data and complex structures for developing docking score. In 2022, it released a sub-database (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. SKEMPI¹² and dbMPIKT¹³ mainly focus on protein-protein interaction (PPI). SKEMPI¹² contains 713 protein-protein binding kinetic rates upon mutation. dbMPIKT¹³ 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)¹⁵. While these methods are often based on experimental structures, many of them consider each biomolecular complex with only one single structure¹⁵. Nunes-Alves et al.¹⁵ 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, 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. 22 and 11 inhibitors were used for training and testing in the COMBINE analysis, respectively. The first step in the COMBINE analysis involved modelling 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 non-bonded interactions. Weights to scale the protein-small molecule interaction energies were obtained using partial least square 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 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 docking¹⁵. Although 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 structures 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 another study, Schuetz et al.¹⁶ performed matched molecular pair (MMP) analysis on datasets 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 dataset (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 dataset included various chemical modifications, with the top 20 representing less than 65% of the entire dataset. 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 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 charged moiety, leading to 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 2018, Ganatra and Wade applied COMBINE analysis to derive QSKRs for the dissociation rates (k_{off}) of inhibitors of HSP90 and HIV-1 protease¹⁷. 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 were 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 were 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 dataset 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 HIV-1 protease inhibitors. The resulting COMBINE models for k_{off} rates had very good predictive power ($Q^2_{\text{LOO}} = 0.69$ for HSP90, and $Q^2_{\text{LOO}} = 0.70$ for HIV-1 protease), which could also identify contributing protein-ligand interactions for binding kinetics.

In order to explore molecular details of biomolecular binding processes on a large scale, Chiu et al.¹⁸ recently integrated coarse-grained normal mode analysis (NMA) with multi-target 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 RMLR as the dot product of ligand displacement vector after normalization and the residue displacement vector, and 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:

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

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 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 not only found to recapitulate the results from MD simulations but also 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.

Engel et al. designed novel and irreversible epidermal growth factor receptor (EGFR) inhibitors using a structure-based approach rationalized by subsequent computational analysis of conformational ligand ensembles in solution¹⁹. The approach was based on a screening hit that was identified in a phenotype screen of ~1,500 compounds in 80 non-small cell lung cancer (NSCLC) cell lines. With X-ray crystallography, the binding mode in engineered cSrc (T338M/S345C), a validated model system for EGFR-T790M, was deciphered. Chemical synthesis revealed further compound collections that increased biochemical potency and selectivity toward mutated (L858R

and L858R/T790M) vs. wildtype EGFR. Kinetic studies were performed to investigate the rate and efficiency of covalent bond formation for the most effective inhibitors, **5b** and **6a**. The corresponding time-dependent affinity (K_i) and reactivity (k_{inact}) parameters with respect to the mutants L858R and T790M/L858R of EGFR were determined by an activity-based assay. The respective IC₅₀ values of **5b** and **6a** were monitored after treatment of the respective proteins in a time-dependent manner. These values were correlated to the respective incubation times, from which K_i and k_{inact} parameters could be determined⁸. Inhibitors **5b** and **6a** were found to exhibit extraordinarily high affinity toward EGFR L858R/T790M, with respective K_i values of 0.64 and 0.32 nM, and specific, moderate reactivity, with k_{inact} values of 0.116 and 0.137 min⁻¹. On the contrary, both the binding affinity and specific reactivity of **5b** and **6a** toward EGFR L858R were significantly impaired (K_i = 70.2 and 833 nM, and k_{inact} = 0.017 and 0.055 min⁻¹). 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 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 dynamics²⁰. The accessible timescale of conventional MD (cMD) has reached hundreds of microseconds thanks to remarkable advances in computing hardware (e.g., the Anton supercomputer and GPUs) and software developments²¹. Notably, the latest Anton3^{21f} has enabled hundreds-of-microseconds cMD simulations per day. The cMD simulations have been widely applied to investigate biomolecular binding processes²². However, it is still challenging for cMD to simulate repetitive biomolecular dissociation and rebinding processes^{21a, 23}. In this regards, enhanced sampling methods²⁴ have been

developed to simulate biomolecular binding and dissociation processes, and predict the associated binding kinetic rates, including the widely used Weighted Ensemble²⁵, mile-stoning method²⁶, Gaussian accelerated MD (GaMD)²⁷, Metadynamics²⁸, Markov State Modeling (MSM)²⁹, Random Acceleration Molecular Dynamics (RAMD)³⁰, scaled MD³¹ and so on. Recent years have seen a significant increasing numbers of studies on predicting biomolecular binding kinetic rates using MD simulations (**Fig. 1A**). To evaluate the accuracy of simulation predicted kinetic rates, we define the prediction errors of binding and dissociation kinetic rates as:

$$\begin{aligned}\Delta \log k_{on} &= \log k_{on}^{sim} - \log k_{on}^{exp}, \\ \Delta \log k_{off} &= \log k_{off}^{sim} - \log k_{off}^{exp},\end{aligned}\quad (5)$$

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 (**Fig.1B**), suggesting good prediction accuracy obtained from MD simulations. 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.

Protein-small molecule binding kinetics

Compared with slower ligand dissociation process, ligand binding is much quicker, which allows cMD to capture 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 μ s cMD simulations performed by Shan et al.^{22a}. The estimated binding association rate (k_{on}) was $0.19 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, being highly consistent with the experimental value of $0.5 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$. The same system was used to test a novel approach-unaggregated unbiased MD (UUMD) developed by

Sohraby et al.³². In contrast to the repulsion added to special atom in the ligand by Shan et al.^{22a}, the repulsion in the UUMD was 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 $0.75 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, 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 timescale³³. Based on Martini coarse-grained model, Dandekar et al.³⁴ 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 ($36.8 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, $6.9 \times 10^5 \text{ s}^{-1}$). The corresponding experimental values were ($2.9 \times 10^7 \text{ M}^{-1}\text{s}^{-1}$, 600 s^{-1}). Therefore, the predicted k_{on} value was ~ 13 folds higher than the experimental data. However, 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 calculations³⁵. For example, simulation enabled estimation of kinetic rates (SEEKR)^{35b, 36} is a multiscale simulation approach combining MD, Brownian dynamics, and mile-stoning for predicting protein–ligand binding association and dissociation rates. SEEKR has been shown to estimate accurate binding kinetic rates with simulation time reduced by a factor of 10^{35b} . Using the trypsin-benzamidine model system as example, the SEEKR and its latest version SEEKR2 predicted the binding kinetic rates of (k_{on} , k_{off})

at $(12 \pm 0.5 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 174 \pm 9 \text{s}^{-1})$ and $(2.4 \pm 0.2 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 990 \pm 130 \text{s}^{-1})$, respectively, being highly consistent with the corresponding experimental data of $(2.9 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 600 \text{s}^{-1})$.

Mile-stoning method²⁶ has been applied to predict the dissociation rate of the Imatinib drug to Abl kinase. Based on the total of $1.043 \mu\text{s}$ simulations, the value of k_{off} was predicted as 18 s^{-1} , being highly consistent with the experimental value of $25 \pm 6 \text{s}^{-1}$. Weighted Ensemble³⁷ and MSM^{29a} 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 $2,660 \text{ s}^{-1}$ was predicted with one weighted ensemble³⁷ of a total amount of $8.75 \mu\text{s}$ cMD simulations, being ~ 4.43 times faster than the experimental value. Another weighted ensemble³⁸ of a total of $0.48 \mu\text{s}$ cMD was able to predict the T4 lysozome (T4L)-benzene binding kinetic rates of ($k_{\text{on}}, k_{\text{off}}$) at $(0.53 \pm 0.08 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 791 \pm 197 \text{s}^{-1})$, being highly consistent with the corresponding experimental value of $(0.08-0.1 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 950 \pm 200 \text{s}^{-1})$. MSM was able to simultaneously predict the ligand association and dissociation rates through longer aggregated cMD simulations. For example, one MSM built with $59 \mu\text{s}$ cMD simulation data was able to accurately predict T4L-benzene binding kinetic rates. The predicted binding kinetic rate values of ($k_{\text{on}}, k_{\text{off}}$) were $(0.21 \pm 0.09 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 310 \pm 130 \text{s}^{-1})$, being highly consistent with the experimental data of $(0.08-0.1 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 950 \pm 200 \text{s}^{-1})$. MSM built with $50 \mu\text{s}$ cMD simulation data was used to predict the binding kinetic rates of the trypsin-benzamidine system^{29b}. The predicted values of ($k_{\text{on}}, k_{\text{off}}$) were $(15.0 \pm 2.0 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 9.5 \pm 3.3 \times 10^4 \text{s}^{-1})$, being in line with the experimental values of $(2.9 \times 10^7 \text{M}^{-1}\text{s}^{-1}, 600 \text{s}^{-1})$. However, these calculations required very expensive computational resources.

Metadynamics³⁹ has been widely applied to investigate the ligand binding kinetics. Multiple Infrequent Metadynamics (InMetaD) simulations with a total of $5 \mu\text{s}$ 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 ($1.18 \pm 1.0 \times 10^7 \text{ M}^{-1} \text{s}^{-1}$, $9.1 \pm 2.5 \text{ s}^{-1}$), being smaller than the experimental values of ($2.9 \times 10^7 \text{ M}^{-1} \text{s}^{-1}$, 600 s^{-1}). Similar smaller predicted values of (k_{on} , k_{off}) at ($0.0035 \pm 0.002 \times 10^7 \text{ M}^{-1} \text{s}^{-1}$, $7 \pm 2 \text{ s}^{-1}$) were observed in another 12 μs InMetaD simulations of benzene binding to T4L^{28a}. For the Src-Dasatinib system, one study with 7 μs InMetaD simulations⁴⁰ was able to predict the k_{off} value of $0.048 \pm 0.024 \text{ s}^{-1}$, being highly consistent with the experimental value of 0.06 s^{-1} . For the p38 α -compound I system, 6.8 μs InMetaD simulations^{28f} predicted the k_{off} value of $0.020 \pm 0.011 \text{ s}^{-1}$, 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.⁴¹ applied two approaches to obtain the RESP charges for 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. While simulations with RESP-B3LYP charges could predict the k_{off} value of $3.7 \pm 0.7 \times 10^{-4} \text{ s}^{-1}$, being in line with the experimental data of $1.0 \pm 0.2 \times 10^{-2} \text{ s}^{-1}$. For the Src-Imatinib system, Haldar et al.⁴² 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 to the experimental value of $0.11 \pm 0.08 \text{ s}^{-1}$. Although Metadynamics simulations have shown remarkable improvements in capturing ligand binding and dissociation processes that occur over exceedingly long timescales, users often face a challenge for defining collective variables (CVs), which requires expert knowledge of the studied systems⁴³. The simulations may suffer from a “hidden energy barrier” problem if important CVs were missed during the simulation setup⁴⁴. To facilitate the choice of CVs, machine learning (ML) has been incorporated into Metadynamics simulations.

Wang et al. developed a predictive information bottleneck (PIB) approach to identify CVs and predict biomolecular dissociation rates⁴⁵. The PIB was tested on the system of benzene binding to T4L and the predicted k_{off} value was $3.3 \pm 0.8 \text{ s}^{-1}$, being consistent with other InMetaD simulations but needing much shorter simulations^{29c}. In another study, Filizola et al.⁴⁶ 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 $\sim 6 \text{ }\mu\text{s}$ InMetaD simulations, the predicted k_{off} for the morphine and bruprenorphine were $0.057 \pm 0.005 \text{ s}^{-1}$ and $0.021 \pm 0.003 \text{ s}^{-1}$, respectively, being within one order of magnitude difference from experimental values of $0.0023 \pm 0.001 \text{ s}^{-1}$ and $0.0018 \pm 0.03 \text{ s}^{-1}$. Very recently, Narjes et al.⁴⁷ 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 $\sim 2.74 \text{ }\mu\text{s}$ 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 MD⁴⁸ has been mainly used for the prediction of k_{off} as a scale factor ranging from 0 to 1 is introduced in the simulations to reduce energy barrier to facilitate ligand dissociation. For example, Schuetz et al.^{31b} performed scaled MD simulations to accurately predict the residence time and drug dissociation pathways of different inhibitors in Hsp90. In a recent study⁴⁹, 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. 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.^{30a} 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.

GaMD²⁷ 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 order²⁷. GaMD has been successfully applied to simulate important biomolecular processes, including protein/RNA folding^{27b, 50}, ligand/protein/RNA binding^{50a, 51}, and protein conformational changes^{51f, 52}. However, it remained challenging to accurately predict ligand binding kinetic rates through normal GaMD^{27a, 53}. Recently, a “selective GaMD” algorithm, called Ligand GaMD (LiGaMD)⁵⁴, has been developed to allow for more efficiently 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 non-bonded potential energies in protein P, ligand 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. Ligand binding mainly involves the non-bonded 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 systems⁵⁴. 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 $1.15 \pm 0.79 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $3.53 \pm 1.41 \text{ s}^{-1}$, respectively. These data were comparable to the experimental values⁵⁵ of $2.9 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$ 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 domain⁵⁶. The Nirmatrelvir binding and dissociation rates were predicted to be $3.20 \pm 0.21 \times 10^5 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $2.92 \pm 0.37 \times 10^3 \text{ s}^{-1}$, respectively. As no available experimentally determined binding kinetic rates, the authors predicted the dissociation constant (k_D) from the predicted binding kinetic rates by equation $k_D = k_{\text{off}}/k_{\text{on}}$. Notably, the predicted k_D was $9.10 \pm 0.29 \text{ nM}$, being highly consistent with the available experimental value of $7 \pm 3 \text{ nM}$ ⁵⁷, demonstrating high accuracy of the predicted binding kinetic rates from LiGaMD simulations.

Protein-peptide binding kinetics

In comparison with the extensively studied protein-small molecule binding, protein-peptide binding studies are much less although increasing number of peptide-based drugs are being licensed to market in recent years⁵⁸. Large conformational changes of peptides often occur during binding to target proteins, bringing huge challenges for modeling^{22b, 59}. For example, coupled

folding-upon binding mechanism has been observed in several systems of peptide binding to proteins^{22b, 59}. Only a few number of computational approaches have been implemented to predict peptide binding kinetic rates, including the InMetaD⁵⁹, Weighted Ensemble⁶⁰, MSM⁶¹, and Peptide GaMD (Pep-GaMD)⁶² (**Table 3**).

InMetaD simulations with three CVs have successfully predicted the peptide binding and dissociation rates for the system of p53-MDM2⁵⁹. Based on 27 μ s InMetaD simulations⁵⁹, the predicted values of (k_{on} , k_{off}) were ($0.43 \pm 0.22 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, $0.7 \pm 0.4 \text{ s}^{-1}$), being comparable to the corresponding experimental values of ($0.92 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, 2.06 s^{-1}). Weighted Ensemble of a total amount of $\sim 120 \mu$ s cMD simulations in implicit solvent was performed on the same p53-MDM2 system⁶⁰. The predicted p53 binding kinetic rate (k_{on}) was 7 s^{-1} , being highly consistent with the experimental data of 2.06 s^{-1} . Built on a total of 831 μ s cMD simulations of p53 binding to the MDM2, the MSM⁶¹ predicted accurate values of k_{on} and k_{off} at $0.019 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$ 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 MDM2⁶³. The PMI dissociation process is rather slow with the residence time at the timescale of second. Therefore, $\sim 50 \mu$ s Hamiltonian replica exchange simulations were performed to predict the dissociate rate. The predicted values of (k_{on} , k_{off}) were ($300 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, $0.125-1.13 \text{ s}^{-1}$), being comparable to the corresponding experimental values of ($52.7 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, 0.037 s^{-1}).

Based on GaMD, we recently developed an algorithm called peptide GaMD or “Pep-GaMD” that enhances sampling of protein-peptide interactions⁶². As above mentioned, large conformational change involved in the process of peptide binding to target proteins^{22b, 59}. Therefore,

peptide binding involves in both the bonded and non-bonded interaction energies of the peptide. Thus, the essential peptide 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 essential peptide 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-GaMD⁶² 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 domains⁶⁴, including “PPPVPPRR” (PDB: 1CKB), “PPPALPPKK” (PDB: 1CKA) and “PAMPAR” (PDB: 1SSH). Repetitive peptide binding and unbinding events were captured in independent 1 μ s 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 was ($4060 \pm 2260 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$, $1450 \pm 1170 \text{ s}^{-1}$), being within 1 order of the experimental data of ($150 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$, 8900 s^{-1}).

Protein-protein binding kinetics

Protein-protein interactions (PPIs) play key roles in many fundamental biological processes, including cellular signal transduction, immune responses and so on¹. Moreover, PPIs are implicated in the development of numerous human diseases and served as important drug targets.⁶⁵ 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 protein-small molecule and protein-peptide interactions. Protein-protein binding and unbinding processes often occurred in significantly longer timescale. Particularly, 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^{20d}. Based on 28 successfully binding events captured in a total of ~213 μ s Anton cMD simulations with TIP4P2005 water model^{20d}, the predicted barnase binding rate (k_{on}) was $6 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$, being in line with the experimental value of $60 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$. Less barnase binding events (24) with slower predicted binding rate ($2.3 \times 10^7 \text{ M}^{-1} \text{ s}^{-1}$) were observed with the TIP3P water model. Additionally, Pan et al.^{20d} successfully predicted the binding kinetic association rates of another two systems of insulin dimerization and Ras binding to Ras-binding domain of c-RAF-1 (Ras-Raf-RBD). Based on 6 successful binding events among the total of 294.8 μ s cMD simulations, the predicted association rate (k_{on}) of the insulin dimerization was $0.41 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$, being comparable to the experimental value of $11.4 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$. For the Ras-Raf-RBD system, 117 μ s cMD simulations successfully captured 7 binding events and predicted k_{on} value of $2.6 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$, being highly consistent with the experimental data of $4.5 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$. However, it remains challenging to simulate the protein dissociation with cMD^{20d}.

Weighted Ensemble^{24d} of a total of ~18 μ s cMD simulations were able to capture 203 barnase binding events and accurately predict the barnase-barstar binding rate constant (k_{on}) of $23 \pm 10 \times 10^7 \text{ M}^{-1} \cdot \text{s}^{-1}$. Plattner et al.⁶⁶ performed high throughput MD simulations of the barnase binding to barstar to build MSM. A total of 1700 μ s cMD simulations with 1,892 independent replicas starting from unbound state captured 74 barnase binding events. Another set of 300 μ s adaptive MD simulations captured 16 and 10 times of barnase binding and dissociation events, respectively. Based on the total of 2,000 μ s simulation data, the obtained MSM was able to predict

intermediate structures, binding energies and kinetic rates that were consistent with experimental data⁶⁶.

Recently, we developed a selective PPI-GaMD method⁶⁷ 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 non-bonded 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 rebinding⁶⁷. PPI-GaMD⁶² has been demonstrated on the model system of barnase binding to the barstar. Six independent 2 μ s 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 simulations. 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^{65c}. Additionally, the PPI-GaMD simulations allowed us to calculate the protein binding kinetics. The average k_{on} and k_{off} were predicted as $21.7 \pm 13.8 \times 10^8 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $7.32 \pm 4.95 \times 10^{-6} \text{ s}^{-1}$, being consistent with the corresponding experimental values of $6.0 \times 10^8 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $8.0 \times 10^{-6} \text{ s}^{-1}$, respectively.

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 throughput¹⁵. For MD simulations, accuracy of binding free energy calculations could be within 1.0 kcal/mol with the modern techniques⁶⁸. 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 derive orders of magnitude from the experimental data (**Tables 2-4 & Fig 1B**). Nevertheless, MD simulations have enabled characterization of biomolecular binding pathways and kinetics, attracting increasing attentions in recent years. With advances in computer hardware and accuracy of force fields, long timescale cMD simulations with all-atom and/or coarse-grained models have successfully captured biomolecular binding process and predicted accurate binding associate rates^{20d}, although slower dissociation processes are still difficult to simulate.

Enhanced sampling methods have greatly reduced the computational cost for calculations of biomolecular kinetics. Among various enhanced sampling methods, the MSM, InMetady and GaMD appear to be the most used techniques that allow for simultaneous predictions of biomolecular binding association and dissociation rates (**Fig. 2**). Another trend is the incorporation of ML into enhanced sampling methods to further improve sampling efficiency and prediction accuracy of biomolecular binding kinetic rates^{46, 69}.

Overall, current computational methods have been tested mostly on 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. This would suggest

a need for community blind challenges on 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 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) and National Science Foundation (2121063).

6. References

1. Nooren, I. M.; Thornton, J. M., Diversity of protein–protein interactions. *The EMBO journal* **2003**, *22* (14), 3486-3492.
2. (a) 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; (b) Hajduk, P. J.; Greer, J., A decade of fragment-based drug design: strategic advances and lessons learned. *Nature reviews Drug discovery* **2007**, *6* (3), 211-219; (c) Klebe, G., Applying thermodynamic profiling in lead finding and optimization. *Nature Reviews Drug Discovery* **2015**, *14* (2), 95-110.
3. (a) 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., Kinetics for Drug Discovery: an industry-driven effort to target drug residence time. *Drug Discovery Today* **2017**, *22* (6), 896-911; (b) Tonge, P. J., Drug–target kinetics in drug discovery. *ACS chemical neuroscience* **2018**, *9* (1), 29-39; (c) 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. *Frontiers in Molecular Biosciences* **2022**, *9*; (d) Ijzerman, A. P.; Guo, D., Drug–Target Association Kinetics in Drug Discovery. *Trends in Biochemical Sciences* **2019**, *44* (10), 861-871; (e) Ferruz, N.; De Fabritiis, G., Binding Kinetics in Drug Discovery. *Molecular Informatics* **2016**, *35* (6-7), 216-226; (f) 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. *Journal of Medicinal Chemistry* **2021**, *64* (4), 1844-1855; (g) Holdgate, G. A.; Gill, A. L., Kinetic efficiency: the missing metric for enhancing compound quality? *Drug Discovery Today* **2011**, *16* (21), 910-913; (h) Copeland, R. A.; Pompliano, D. L.; Meek, T. D., Drug-target residence time and its implications for lead optimization. *Nat Rev Drug Discov* **2006**, *5* (9), 730-739.
4. (a) Sohraby, F.; Nunes-Alves, A. In *Recent advances in computational methods for studying ligand binding kinetics*, 2022; (b) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A., Kinetics of Drug Binding and Residence Time. *Annual Review of Physical Chemistry* **2019**, *70* (1), 143-171; (c) Lu, H.; Tonge, P. J., Drug–target residence time: critical information for lead optimization. *Current Opinion in Chemical Biology* **2010**, *14* (4), 467-474; (d) Bruce, N. J.; Ganatra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C., New approaches for computing ligand–receptor binding kinetics. *Current Opinion in Structural Biology* **2018**, *49*, 1-10; (e) Dahl, G.; Akerud, T., Pharmacokinetics and the drug–target residence time concept. *Drug Discovery Today* **2013**, *18* (15), 697-707; (f) Copeland, R. A., Drug–target interaction kinetics: underutilized in drug optimization? *Future Medicinal Chemistry* **2016**, *8* (18), 2173-2175; (g) Copeland, R. A., The drug–target residence time model: a 10-year retrospective. *Nature Reviews Drug Discovery* **2016**, *15* (2), 87-95.
5. (a) Gleitsman, K. R.; Sengupta, R. N.; Herschlag, D., Slow molecular recognition by RNA. *Rna* **2017**, *23* (12), 1745-1753; (b) 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.
6. (a) Hulme, E. C.; Trevethick, M. A., Ligand binding assays at equilibrium: validation and interpretation. *British journal of pharmacology* **2010**, *161* (6), 1219-1237; (b) Guo, D.; Hillger, J. M.; Ijzerman, A. P.; Heitman, L. H., Drug-target residence time—a case for G protein-coupled receptors. *Medicinal research reviews* **2014**, *34* (4), 856-892.

7. 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.
8. 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 Research* **2003**, *31* (1), 255-257.
9. 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 Research* **2015**, *44* (D1), D1045-D1053.
10. 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*.
11. 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.
12. 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.
13. Liu, Q.; Chen, P.; Wang, B.; Zhang, J.; Li, J., dbMPIKT: a database of kinetic and thermodynamic mutant protein interactions. *BMC Bioinformatics* **2018**, *19* (1), 455.
14. 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.
15. Nunes-Alves, A.; Ormersbach, F.; Wade, R. C., Prediction of the Drug–Target Binding Kinetics for Flexible Proteins by Comparative Binding Energy Analysis. *Journal of Chemical Information and Modeling* **2021**, *61* (7), 3708-3721.
16. Schuetz, D. A.; Richter, L.; Martini, R.; Ecker, G. F., A structure–kinetic relationship study using matched molecular pair analysis. *RSC Medicinal Chemistry* **2020**, *11* (11), 1285-1294.
17. Ganatra, G. K.; Wade, R. C., Prediction of Drug–Target Binding Kinetics by Comparative Binding Energy Analysis. *ACS Medicinal Chemistry Letters* **2018**, *9* (11), 1134-1139.
18. Chiu, S. H.; Xie, L., Toward High-Throughput Predictive Modeling of Protein Binding/Unbinding Kinetics. *Journal of Chemical Information and Modeling* **2016**, *56* (6), 1164-1174.
19. Engel, J.; Richters, A.; Getlik, M.; Tomassi, S.; Keul, M.; Termathe, M.; Lategahn, J.; Becker, C.; Mayer-Wrangowski, S.; Grüter, C.; Uhlenbrock, N.; Krüll, J.; Schaumann, N.; Eppmann, S.; Kibies, P.; Hoffgaard, F.; Heil, J.; Menninger, S.; Ortiz-Cuaran, S.; Heuckmann, J. M.; Tinnefeld, V.; Zahedi, R. P.; Sos, M. L.; Schultz-Fademrecht, C.; Thomas, R. K.; Kast, S. M.; Rauh, D., Targeting Drug Resistance in EGFR with Covalent Inhibitors: A Structure-Based Design Approach. *Journal of Medicinal Chemistry* **2015**, *58* (17), 6844-6863.
20. (a) 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. *Journal of Chemical Theory and Computation* **2021**, *17* (5), 3088-3102; (b) He, Z.; Paul, F.; Roux, B., A critical perspective on Markov state model treatments of protein–protein association using coarse-grained simulations. *The Journal of Chemical Physics* **2021**, *154* (8), 084101; (c) Basdevant, N.; Borgis, D.; Ha-Duong, T., Modeling protein–protein recognition in solution using the coarse-grained force field SCORPION. *Journal of Chemical Theory and Computation* **2013**, *9* (1), 803-813; (d) Pan, A. C.; Jacobson, D.; Yatsenko, K.; Sritharan, D.; Weinreich, T. M.; Shaw, D. E., Atomic-level characterization of protein–protein association. *Proceedings of the National Academy of Sciences* **2019**, 201815431; (e) Karplus, M.; McCammon, J. A., Molecular dynamics simulations of biomolecules. *Nature structural biology* **2002**, *9* (9), 646-652; (f) Tang, Z.; Roberts, C. C.; Chang, C.-e. A., Understanding ligand-receptor non-covalent binding kinetics using molecular modeling. *FBL* **2017**, *22* (6), 960-981.

21. (a) Hollingsworth, S. A.; Dror, R. O., Molecular Dynamics Simulation for All. *Neuron* **2018**, *99* (6), 1129-1143; (b) Harvey, M. J.; Giupponi, G.; Fabritiis, G. D., ACEMD: accelerating biomolecular dynamics in the microsecond time scale. *Journal of chemical theory and computation* **2009**, *5* (6), 1632-1639; (c) Johnston, J. M.; Filizola, M., Showcasing modern molecular dynamics simulations of membrane proteins through G protein-coupled receptors. *Current opinion in structural biology* **2011**, *21* (4), 552-558; (d) 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., Atomic-level characterization of the structural dynamics of proteins. *Science* **2010**, *330* (6002), 341-346; (e) Lane, T. J.; Shukla, D.; Beauchamp, K. A.; Pande, V. S., To milliseconds and beyond: challenges in the simulation of protein folding. *Current opinion in structural biology* **2013**, *23* (1), 58-65; (f) 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*, Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis, 2021; pp 1-11.

22. (a) 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? *Journal of the American Chemical Society* **2011**, *133* (24), 9181-9183; (b) Robustelli, P.; Piana, S.; Shaw, D. E., Mechanism of Coupled Folding-upon-Binding of an Intrinsically Disordered Protein. *Journal of the American Chemical Society* **2020**, *142* (25), 11092-11101; (c) 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.

23. Wang, J.; Bhattacharai, A.; Do, H. N.; Miao, Y., Challenges and frontiers of computational modelling of biomolecular recognition. *QRB Discovery* **2022**, *3*, e13.

24. (a) Kamenik, A. S.; Linker, S. M.; Riniker, S., Enhanced sampling without borders: on global biasing functions and how to reweight them. *Physical Chemistry Chemical Physics* **2022**, *24* (3), 1225-1236; (b) Wong, C. F., Molecular simulation of drug-binding kinetics. *Molecular Simulation* **2014**, *40* (10-11), 889-903; (c) 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; (d) Saglam, A. S.; Chong, L. T., Protein–protein binding pathways and calculations of rate constants using fully-continuous, explicit-solvent simulations. *Chemical Science* **2019**, *10*, 2360-2372.

25. (a) 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; (b) Nunes-

Alves, A.; Zuckerman, D. M.; Arantes, G. M., Escape of a Small Molecule from Inside T4 Lysozyme by Multiple Pathways. *Biophysical Journal* **2018**, *114* (5), 1058-1066.

26. Narayan, B.; Buchete, N.-V.; Elber, R., Computer Simulations of the Dissociation Mechanism of Gleevec from Abl Kinase with Milestoning. *The Journal of Physical Chemistry B* **2021**, *125* (22), 5706-5715.

27. (a) Wang, J.; Arantes, P. R.; Bhattacharai, 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; (b) Miao, Y.; Feher, V. A.; McCammon, J. A., Gaussian Accelerated Molecular Dynamics: Unconstrained Enhanced Sampling and Free Energy Calculation. *Journal of Chemical Theory and Computation* **2015**, *11* (8), 3584-3595.

28. (a) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K., Frequency adaptive metadynamics for the calculation of rare-event kinetics. *The Journal of Chemical Physics* **2018**, *149* (7), 072309; (b) Wang, Y.; Martins, J. M.; Lindorff-Larsen, K., Biomolecular conformational changes and ligand binding: from kinetics to thermodynamics. *Chemical Science* **2017**, *8* (9), 6466-6473; (c) Banerjee, P.; Bagchi, B., Dynamical control by water at a molecular level in protein dimer association and dissociation. *Proceedings of the National Academy of Sciences* **2020**, *117* (5), 2302-2308; (d) 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. *The Journal of Physical Chemistry B* **2020**, *124* (27), 5571-5587; (e) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M., Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences* **2015**, *112* (5), E386-E391; (f) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M., Unbinding kinetics of a p38 MAP kinase type II inhibitor from metadynamics simulations. *Journal of the American Chemical Society* **2017**.

29. (a) Buch, I.; Giorgino, T.; De Fabritiis, G., Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proceedings of the National Academy of Sciences* **2011**, *108* (25), 10184-10189; (b) Plattner, N.; Noé, F., Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nature communications* **2015**, *6* (1), 1-10; (c) 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 Computational Biology* **2018**, *14* (5), e1006180.

30. (a) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C., Ligand unbinding mechanisms and kinetics for T4 lysozyme mutants from τ RAMD simulations. *Current Research in Structural Biology* **2021**, *3*, 106-111; (b) 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. *Journal of Chemical Theory and Computation* **2018**, *14* (7), 3859-3869; (c) Wang, T.; Duan, Y., Chromophore channeling in the G-protein coupled receptor rhodopsin. *Journal of the American Chemical Society* **2007**, *129* (22), 6970-6971.

31. (a) Deb, I.; Frank, A. T., Accelerating Rare Dissociative Processes in Biomolecules Using Selectively Scaled MD Simulations. *Journal of Chemical Theory and Computation* **2019**, *15* (11), 5817-5828; (b) 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* **2018**.

32. 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.

33. (a) Souza, P. C.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M.; Wassenaar, T. A., Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nature methods* **2021**, *18* (4), 382-388; (b) 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. *Nature Communications* **2020**, *11* (1), 3714.

34. Dandekar, B. R.; Mondal, J., Capturing Protein–Ligand Recognition Pathways in Coarse-Grained Simulation. *The Journal of Physical Chemistry Letters* **2020**, *11* (13), 5302-5311.

35. (a) Huang, Y.-m. M., Multiscale computational study of ligand binding pathways: Case of p38 MAP kinase and its inhibitors. *Biophysical Journal* **2021**, *120* (18), 3881-3892; (b) Jagger, B. R.; Ojha, A. A.; Amaro, R. E., Predicting Ligand Binding Kinetics Using a Markovian Milestoning with Voronoi Tessellations Multiscale Approach. *Journal of Chemical Theory and Computation* **2020**, *16* (8), 5348-5357; (c) Elber, R., Milestoning: An efficient approach for atomically detailed simulations of kinetics in biophysics. *Annual review of biophysics* **2020**, *49*, 69-85.

36. Votapka, L. W.; Amaro, R. E., Multiscale estimation of binding kinetics using Brownian dynamics, molecular dynamics and milestoning. *PLoS computational biology* **2015**, *11* (10), e1004381.

37. Donyapour, N.; Roussey, N. M.; Dickson, A., REVO: Resampling of ensembles by variation optimization. *The Journal of Chemical Physics* **2019**, *150* (24), 244112.

38. Ray, D.; Stone, S. E.; Andricioaei, I., Markovian Weighted Ensemble Milestoning (M-WEM): Long-Time Kinetics from Short Trajectories. *Journal of Chemical Theory and Computation* **2022**, *18* (1), 79-95.

39. (a) Limongelli, V.; Bonomi, M.; Parrinello, M., Funnel metadynamics as accurate binding free-energy method. *Proceedings of the National Academy of Sciences* **2013**, *110* (16), 6358-6363; (b) Tiwary, P.; Parrinello, M., From metadynamics to dynamics. *Physical review letters* **2013**, *111* (23), 230602.

40. Tiwary, P.; Mondal, J.; Berne, B. J., How and when does an anticancer drug leave its binding site? *Science Advances* **2017**, *3* (5).

41. 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.

42. 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.

43. (a) Abrams, C.; Bussi, G., Enhanced sampling in molecular dynamics using metadynamics, replica-exchange, and temperature-acceleration. *Entropy* **2013**, *16* (1), 163-199; (b) Zuckerman, D. M., Equilibrium sampling in biomolecular simulation. *Annual review of biophysics* **2011**, *40*, 41.

44. 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.

45. Wang, Y.; Ribeiro, J. M. L.; Tiwary, P., Past–future information bottleneck for sampling molecular reaction coordinate simultaneously with thermodynamics and kinetics. *Nature Communications* **2019**, *10* (1), 3573.

46. 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. *The Journal of Chemical Physics* **2020**, *153* (12), 124105.

47. Ansari, N.; Rizzi, V.; Parrinello, M., Water regulates the residence time of Benzamidine in Trypsin. *Nat Commun* **2022**, *13* (1), 5438.

48. Sinko, W.; Miao, Y.; de Oliveira, C. s. A. F.; McCammon, J. A., Population based reweighting of scaled molecular dynamics. *The Journal of Physical Chemistry B* **2013**, *117* (42), 12759-12768.

49. 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. *Journal of Chemical Theory and Computation* **2021**, *17* (10), 6522-6535.

50. (a) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A., Gaussian accelerated molecular dynamics in NAMD. *Journal of chemical theory and computation* **2017**, *13* (1), 9-19; (b) Copeland, M. M.; Do, H. N.; Votapka, L.; Joshi, K.; Wang, J.; Amaro, R. E.; Miao, Y., Gaussian Accelerated Molecular Dynamics in OpenMM. *The Journal of Physical Chemistry B* **2022**.

51. (a) 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. *Scientific reports* **2018**, *8* (1), 1-9; (b) 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; (c) Wang, J.; Lan, L.; Wu, X.; Xu, L.; Miao, Y., Mechanism of RNA recognition by a Musashi RNA-binding protein. *Current Research in Structural Biology* **2022**, *4*, 10-20; (d) 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. *Scientific reports* **2017**, *7* (1), 1-11; (e) Liao, J.-M.; Wang, Y.-T., In silico studies of conformational dynamics of Mu opioid receptor performed using gaussian accelerated molecular dynamics. *Journal of Biomolecular Structure and Dynamics* **2019**, *37* (1), 166-177; (f) Miao, Y.; McCammon, J. A., Graded activation and free energy landscapes of a muscarinic G-protein–coupled receptor. *Proceedings of the National Academy of Sciences* **2016**, *113* (43), 12162-12167; (g) Miao, Y.; Caliman, A. D.; McCammon, J. A., Allosteric effects of sodium ion binding on activation of the m3 muscarinic g-protein-coupled receptor. *Biophysical journal* **2015**, *108* (7), 1796-1806.

52. (a) Salawu, E. O., The impairment of torsinA's binding to and interactions with its activator: an atomistic molecular dynamics study of primary dystonia. *Frontiers in molecular biosciences* **2018**, *5*, 64; (b) 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. *Proceedings of the National Academy of Sciences* **2018**, *115* (44), 11129-11137.

53. Miao, Y.; McCammon, J. A., Mechanism of the G-protein mimetic nanobody binding to a muscarinic G-protein-coupled receptor. *Proceedings of the National Academy of Sciences* **2018**, *115* (12), 3036-3041.

54. Miao, Y.; Bhattacharai, A.; Wang, J., Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *Journal of Chemical Theory and Computation* **2020**, *16* (9), 5526-5547.

55. Guillain, F.; Thusius, D., 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* **1970**, *92* (18), 5534-5536.

56. 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. *Physical Chemistry Chemical Physics* **2022**, *24* (37), 22898-22904.

57. 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. *Nature Communications* **2022**, *13* (1), 2268.

58. (a) Fosgerau, K.; Hoffmann, T., Peptide therapeutics: current status and future directions. *Drug Discovery Today* **2015**, *20* (1), 122-128; (b) Ahrens, V. M.; Bellmann-Sickert, K.; Beck-Sickinger, A. G., Peptides and peptide conjugates: therapeutics on the upward path. *Future Medicinal Chemistry* **2012**, *4* (12), 1567-1586; (c) 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. *International Journal of Molecular Sciences* **2019**, *20* (10), 2383.

59. 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. *Journal of Chemical Information and Modeling* **2020**, *60* (3), 1551-1558.

60. Zwier, M. C.; Pratt, A. J.; Adelman, J. L.; Kaus, J. W.; Zuckerman, D. M.; Chong, L. T., Efficient Atomistic Simulation of Pathways and Calculation of Rate Constants for a Protein–Peptide Binding Process: Application to the MDM2 Protein and an Intrinsically Disordered p53 Peptide. *The Journal of Physical Chemistry Letters* **2016**, *7* (17), 3440-3445.

61. Zhou, G.; Pantelopoulos, G. A.; Mukherjee, S.; Voelz, V. A., Bridging Microscopic and Macroscopic Mechanisms of p53-MDM2 Binding with Kinetic Network Models. *Biophysical Journal* **2017**, *113* (4), 785-793.

62. 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.

63. 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. *Nature Communications* **2017**, *8* (1), 1095.

64. (a) Ahmad, M.; Helms, V., How do proteins associate? A lesson from SH3 domain. *Chemistry Central Journal* **2009**, *3* (1), 1-1; (b) Ball, L. J.; Kühne, R.; Schneider-Mergener, J.; Oschkinat, H., Recognition of proline-rich motifs by protein–protein-interaction domains. *Angewandte Chemie International Edition* **2005**, *44* (19), 2852-2869.

65. (a) Ferreira, L. G.; Oliva, G.; Andricopulo, A. D., Protein-protein interaction inhibitors: advances in anticancer drug design. *Expert opinion on drug discovery* **2016**, *11* (10), 957-968; (b) Scott, D. E.; Bayly, A. R.; Abell, C.; Skidmore, J., Small molecules, big targets: drug

discovery faces the protein–protein interaction challenge. *Nature Reviews Drug Discovery* **2016**, *15* (8), 533-550; (c) Schreiber, G.; Fersht, A. R., Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry* **1993**, *32* (19), 5145-5150.

66. 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**, *advance online publication*.

67. 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.

68. 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. *Journal of the American Chemical Society* **2015**, *137* (7), 2695-2703.

69. (a) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C., Machine Learning Analysis of τ RAMD Trajectories to Decipher Molecular Determinants of Drug-Target Residence Times. *Frontiers in Molecular Biosciences* **2019**, *6*; (b) Lamim Ribeiro, J. M.; Tiwary, P., Toward Achieving Efficient and Accurate Ligand-Protein Unbinding with Deep Learning and Molecular Dynamics through RAVE. *Journal of Chemical Theory and Computation* **2019**, *15* (1), 708-719.

70. Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E., SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. *Journal of Chemical Information and Modeling* **2022**, *62* (13), 3253-3262.

71. Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E., SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin–Benzamidine Binding. *The Journal of Physical Chemistry B* **2017**, *121* (15), 3597-3606.

72. Brotzakis, Z. F.; Limongelli, V.; Parrinello, M., Accelerating the Calculation of Protein–Ligand Binding Free Energy and Residence Times Using Dynamically Optimized Collective Variables. *Journal of Chemical Theory and Computation* **2019**, *15* (1), 743-750.

73. Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M., Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proceedings of the National Academy of Sciences* **2015**, *112* (5), E386-E391.

Table 1 Databases of biomolecular binding kinetics.

Database	Description	Website
KDBI	It includes 19,263 entries, which provides 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 interaction, including ~1.1 million compounds and 8.9 thousand 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 <i>k_{off}</i> dataset includes 680 entries with protein-small molecule complex structure.	http://www.pdbbind.org.cn/
SKEMPI	It focuses on protein-protein interaction, which records 713 binding association and dissociation rates upon mutation.	http://life.bsc.es/pid/mutation_database/
dbMPIKT	It focuses on protein-protein interaction, which contains 5291 protein binding association and dissociation rates upon mutation.	http://deeplearn.ahu.edu.cn/web/dbMPIKT/

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} ($10^7 \text{M}^{-1} \text{s}^{-1}$)	k_{off}^{exp} (s^{-1})	k_{on}^{sim} ($10^7 \text{M}^{-1} \text{s}^{-1}$)	k_{off}^{sim} (s^{-1})	Sim. time (μs)	$\Delta \log k_{on}$	$\Delta \log k_{off}$	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 ³⁸
Trypsin-Benzamidine	SEEKR2	2.9	600	2.4±0.2	990±130	5	-0.082	0.22	-	2022 ⁷⁰
Trypsin-Benzamidine	InMetaD+ML	2.9	600	-	1560	2.75	-	0.41	AMBER ff14SB force field and GAFF	2022 ⁴⁷
Abl kinase-imatinib	Mile-stoning	0.15±0.01	25±6	--	18	1.043	-	-0.14	CHARMM 36 and CGenFF	2021 ²⁶
Trypsin-Benzamidine	LiGaMD	2.9	600	1.15±0.79	3.53±1.41	5	-0.40	-2.23	AMBER ff14SB and GAFF	2020 ⁵⁴
Trypsin-Benzamidine	SEEKR	2.9	600	12±0.5	174±9	4.4	0.62	-0.54	-	2020 ⁷¹
M2-Iperoxo	Frequency-adaptive MetaD	-	0.01±0.002	-	3.7±0.7x10 ⁻⁴	8	-	-1.43	AMBER14SB and GAFF	2020 ⁴¹
Trypsin-Benzamidine	CGMD	2.9	600	36.8	6.9x10 ⁵	428	1.10	3.06	MARTINI	2020 ³⁴
μ OR-morphine	InMetaD+ML	0.29±0.001	0.023±0.001	-	0.057±0.005	6	-	0.39	CHARMM 36 and CGenFF	2020 ⁴⁶
μ OR-bruprenorphine	InMetaD+ML	1.33±0.01	0.0018±0.003	-	0.021±0.003	19	-	1.07	CHARMM 36 and CGenFF	2020 ⁴⁶
Src-Dasatinib	cMD	0.5	0.06	0.76	-	6.6	0.18	-	OPLS	2020 ³²
Src-Dasatinib	CGMD	0.5	0.06	4	-	300	0.90	-	MARTINI	2020 ^{33b}
Trypsin-Benzamidine	InMetaD	2.9	600	-	4176±324	~1.00	-	0.84	CHARMM 36 and CGenFF	2019 ⁷²
Trypsin-Benzamidine	WE	2.9	600	-	2660	8.75	-	0.65	CHARMM and CGenFF	2019 ³⁷
T4L-BEN	ML	0.08-0.1	950±200	-	3.3±0.8	-	-	-2.46	CHARMM22*	2019 ⁴⁵
T4L-BEN	InMetaD	0.08-0.1	950±200	0.0035±0.002	7±2	12	-1.36	-2.13	CHARMM22* and CGenFF	2018 ^{28a}
T4L-BEN	MSM	0.08-0.1	950±200	0.21±0.09	310±130	59	0.42	-0.49	CHARMM36	2018 ^{29c}
T4L-BEN	WE	0.08-0.1	950±200	-	1000	29	-	0.022	CHARMM36	2018 ^{25b}
Src-Imatinib	MetaD	-	0.11±0.08	-	0.026	-	-	-0.63	AMBER FF99SB-ILDN and GAFF	2018 ⁴²
Src-Dasatinib	InMetaD	0.5	0.06	-	0.048±0.024	7	-	-0.096	OPLS	2017 ⁴⁰
P38 α -compound I	InMetaD	0.0118	0.14	-	0.02±0.01	6.8	-	-0.84	AMBER FF99SB-ILDN and GAFF	2017 ^{28f}
Trypsin-Benzamidine	InMetaD	2.9	600	1.18±1.0	9.1±2.5	-	-0.39	-1.82	AMBER ff99SB-ILDN and GAFF	2015 ⁷³
Trypsin-Benzamidine	MSM	2.9	600	15±2	9.5±3.3x10 ⁴	50	0.71	2.20	AMBER ff99SB and GAFF	2011 ^{29a}
Src-Dasatinib	cMD	0.5	0.06	0.19	-	35	-0.42	-	AMBER ff99SB and GAFF	2011 ^{22a}

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} ($10^7 \text{M}^{-1} \text{s}^{-1}$)	k_{off}^{exp} (s^{-1})	k_{on}^{sim} ($10^7 \text{M}^{-1} \text{s}^{-1}$)	k_{off}^{sim} (s^{-1})	Sim. time (μs)	$\Delta \log k_{on}$	$\Delta \log k_{off}$	Force field	Year ^{Ref}
SH3-1CKB	Pep-GaMD	150	8900	4060 ± 2260	1450 ± 1170	3	1.43	-0.79	AMBER ff14SB	2020 ⁶²
MDM2/P53	InMetaD	0.92	2.06	0.43 ± 0.22	0.7 ± 0.4	27	0.88	-0.47	AMBER ff99SB-ILDN	2020 ⁵⁹
MDM2/PMI	MSM	52.7	0.037	330	0.125-1.13	500	0.80	0.53	AMBER ff99SB-ILDN	2017 ⁶³
MDM2/P53	MSM	0.92	2.06	0.019	2.5	831	0.88	0.08	AMBER ff99SB-ILDN-nmr	2017 ⁶¹
MDM2/P53	WE	0.92	2.06	7	-	120	0.88	-	AMBER ff99SB-ILDN	2016 ⁶⁰

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)	$\Delta log k_{on}$	$\Delta log k_{off}$	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 ⁶⁷
Barnase-Barstar	WE	60	8×10^{-6}	230 ± 100	-	18	0.58	-	AMBER ff03*	2019 ^{24d}
Barnase-Barstar	cMD	60	8×10^{-6}	2.3	-	440	-1.42	-	AMBER ff99SB-ILDN	2019 ^{20d}
Insulin Dimer	cMD	11.4	14800	0.41	-	294.8	-1.44	-	AMBER ff99SB-ILDN	2019 ^{20d}
Ras-Raf-RBD	cMD	4.5	7.4	2.6	-	117	-0.24	-	AMBER ff99SB-ILDN	2019 ^{20d}
Barnase-Barstar	MSM	60	8×10^{-6}	26.3-26.5	3×10^{-6}	1700	-0.36	-0.42	AMBER ff99SB	2017 ⁶⁶

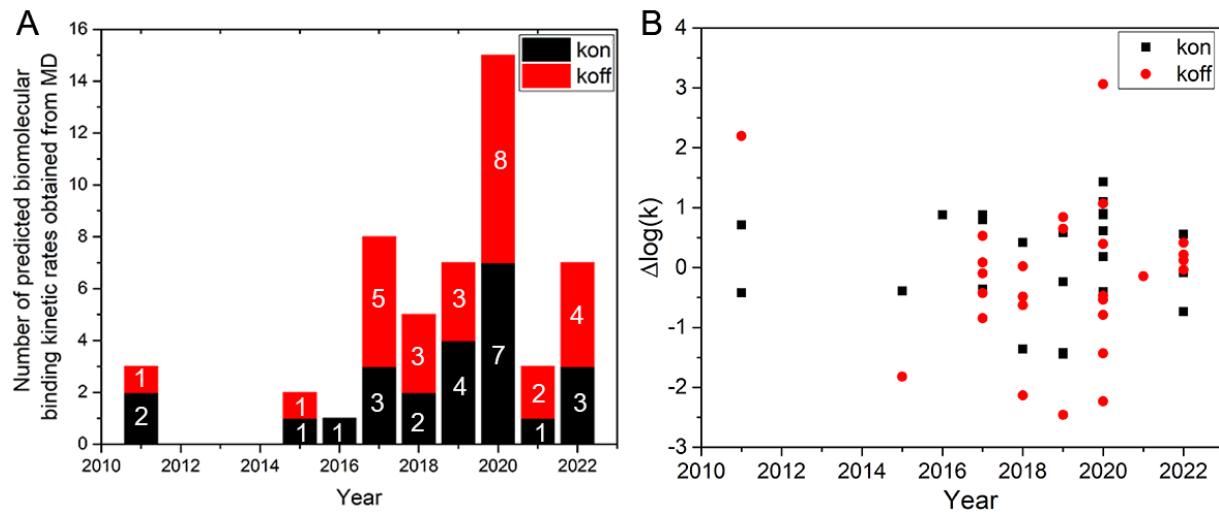


Figure 1. The number (A) and accuracy (B) of predictions of biomolecular binding kinetic rates obtained from MD simulations plotted over the years.

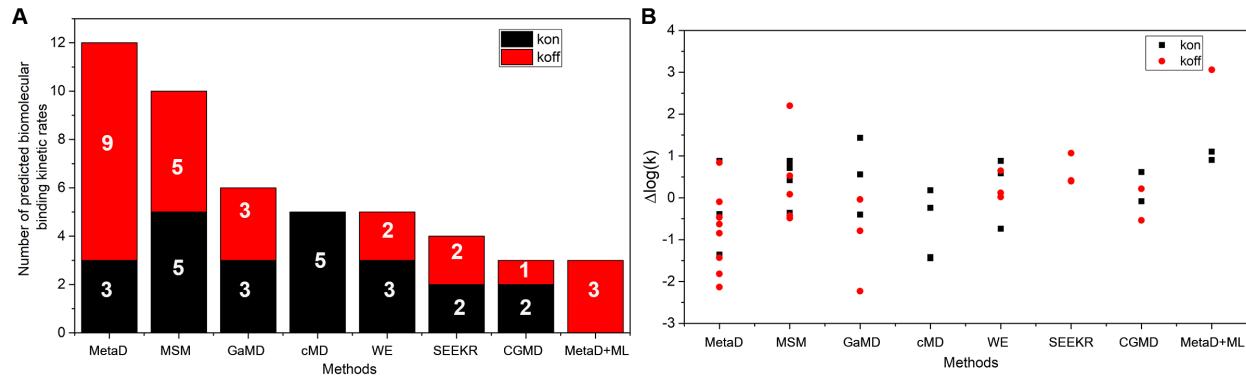


Figure 2. The number (A) and accuracy (B) of predicted biomolecular binding kinetic rates using different MD techniques, including Metadynamics (MetaD), Markov State Models (MSM), Gaussian accelerated MD (GaMD), conventional MD (cMD), Weighted Ensemble (WE), simulation enabled estimation of kinetic rates (SEEKR), coarse-grained MD (CGMD) and combination of Metadynamics and Machine Learning (MetaD+ML).

