Enhanced-Sampling Simulations for the Estimation of Ligand Binding Kinetics: Current Status and Perspective

Katya Ahmad1, Andrea Rizzi1,2, Riccardo Capelli3, Davide Mandelli1, Wenping Lyu4,5 and Paolo Carloni1,6*

1Computational Biomedicine (IAS-5/INM-9), Forschungszentrum Jülich, Jülich, Germany, 2Atomistic Simulations, Istituto Italiano di Tecnologia, Genova, Italy, 3Department of Applied Science and Technology (DISAT), Politecnico di Torino, Torino, Italy, 4Warshel Institute for Computational Biology, School of Life and Health Sciences, The Chinese University of Hong Kong (Shenzhen), Shenzhen, China, 5School of Chemistry and Materials Science, University of Science and Technology of China, Hefei, China, 6Molecular Neuroscience and Neuroimaging (INM-11), Forschungszentrum Jülich, Jülich, Germany

The dissociation rate ($k_{\text{off}}$) associated with ligand unbinding events from proteins is a parameter of fundamental importance in drug design. Here we review recent major advancements in molecular simulation methodologies for the prediction of $k_{\text{off}}$. Next, we discuss the impact of the potential energy function models on the accuracy of calculated $k_{\text{off}}$ values. Finally, we provide a perspective from high-performance computing and machine learning which might help improve such predictions.

Keywords: kinetics, drug discovery, QM/MM, parallel computing, machine learning, enhanced sampling, molecular dynamics

1 INTRODUCTION

The kinetics of drugs unbinding from proteins is an important parameter for the drugs' efficacy. Pan et al., 2013; Copeland, 2021. Indeed, the drug-target residence time (Copeland, Pompliano and Meek, 2006) defined as the inverse of the dissociation rate $k_{\text{off}}$, has emerged as an effective surrogate measure of in vivo target occupancy, and it has been shown to correlate with clinical efficacy (Guo et al., 2012; Lee et al., 2019; Van Der Velden et al., 2020) along with other factors (e.g., association rates (Folmer, 2018; Lee et al., 2019) and target saturation (de Witte et al., 2018)). Residence time has been related not only to long-lasting pharmacodynamics but also to the reduced toxicity of specific inhibitors (Vauquelin et al., 2012).

Experimental approaches (most often combined with computations) measure ligand affinities and provide ligand binding poses for structure-based drug design campaigns (Durrant and McCammon, 2011; De Vivo et al., 2016; Proudfoot et al., 2017; Emwas et al., 2020; Mazzorana et al., 2020). They routinely also measure $k_{\text{off}}$ values (Pollard, 2010). However, they cannot usually access the structural determinants of the transition states associated with ligand unbinding. This information would be crucial to eventually design ligands with longer residence times. In contrast, all-atom molecular simulations (in particular molecular dynamics (MD)) can provide a detailed map of protein-ligand interactions and the atomic rearrangements that drive ligand unbinding. However, the residence time of tight binders can be as long as several hours (Li et al., 2014), much longer than the timescales reached by plain MD (milliseconds on dedicated, specialized machines) (Pan et al., 2019; Shaw et al., 2021). Thus, $k_{\text{off}}$ predictions based on such a straightforward approach so far have been few in number (Pan et al., 2017) or limited to model systems (Tang and Chang, 2018).
Enhanced sampling is a more general approach to the estimation of \( k_{\text{off}} \) regardless of the timescale of the unbinding event. One group of methods (including metadynamics, Gaussian Accelerated MD, scaled MD, and dissipation-corrected targeted MD) employs biasing potentials designed to reduce the free energy barrier determining the frequency of dissociations. Because the bias affects the dynamics, correction terms are required to recover the unbiased \( k_{\text{off}} \) from the biased rates. A second group is represented by path sampling approaches such as weighted ensemble and milestoning. These rigorously generate an ensemble of trajectories by iteratively restarting the (unbiased) simulations from selected configurations (typically closer to the transition state than expected from the equilibrium distribution) with the aim of increasing the likelihood of observing dissociations. Finally, Markov state models (MSMs) can provide a complete picture of the metastable states of the system and transition rates between them by analyzing molecular simulation data.

In this review, we summarize principles and applications of the three approaches outlined above (Sections 2–4). Next, we discuss the impact of force fields on the accuracy of the calculations (Section 5). Finally, we provide a perspective on how machine learning, along with exascale computing, could constitute one way to address these challenges (Section 6).

### 1.1 Scope

Many methods have been developed for the calculation of rate constants in biomolecular simulations. Here, we review methodology that have been applied to the calculation of binding dissociation rates \( (k_{\text{off}}) \) of protein-ligand complexes with a focus on the effect of the potential energy function. In particular, for the sake of conciseness, we do not cover methods that have been applied only to other types of systems/problems (e.g., supramolecular host-guest dissociations, peptide folding rates) and methods that enable relative comparisons of \( k_{\text{off}} \) between different ligands. For these methods, we refer the reader to the other excellent resources on the topic (Chong et al., 2017; Bruce et al., 2018; Nunes-Alves et al., 2020).

### 2 BIASED MD METHODS

In this class of methods, the system is biased (by adding a potential term to the Hamiltonian, or adding external forces) to favor the observation of unbinding events. The bias is designed to enhance the exploration along low-dimensional collective variables (CV), which are represented as differentiable functions \( s(x) \) of the atomic coordinates \( x \). These describe the slow degrees of freedom governing the unbinding process. The CV must be able to distinguish the metastable states involved in the process i.e., configurations in different states should correspond to different values of the CV. The identification of optimal CVs (whenever they are not intrinsic in the technique) is a complicated task, and their identification is at the center of a heated debate that is still open (Sittel and Stock, 2018). Because biasing terms alter the dynamics, methods which recover the kinetic parameters of the unbiased system from its free energy surface have been devised. The majority of biased methods adopt specific corrections based on Kramers’ rate theory

\[
 k_{AB} = \omega_A \kappa_A \frac{Z^*}{Z_A} 
\]

where \( k_{AB} \) is the rate of transition from state A to B (in this case the bound and unbound states), \( \omega_A \) is typically associated with the curvature of the free energy surface, \( \kappa_A \) is the transmission coefficient, and \( Z^* \) and \( Z_A \) are the configurational partition functions of the transition state and state A, respectively. These methods require the identification of the transition state ensemble, defined as the set of conformations of highest free energy along the (un)binding pathway. This is in general a challenging task for drug binding processes, which can involve multiple dissociation pathways due to the conformational flexibility of the protein (Plattner and Noé, 2015). Approaches of this kind have been developed for Gaussian accelerated molecular dynamics (Miao et al., 2020) (see Section 2.1), dissipation-corrected Targeted Molecular Dynamics (Wolf and Stock, 2018) (see Section 2.2), and \( \tau \)-random acceleration molecular dynamics (Kokh et al., 2018) (see Section 2.3). If no bias is deposited on the region of the transition state(s), the kinetic correction can be assumed not to depend on \( \kappa_A \) and \( Z^* \) (Voter and Doll, 1985; Hänggi et al., 1996; Truhlar et al., 1996). This simplifies dramatically the rate estimation, and it is used for ligand unbinding in the kinetics-oriented flavors of metadynamics (Tiwary and Parrinello, 2013; Wang et al., 2018) (see Section 2.4).

### 2.1 Ligand Gaussian Accelerated MD

#### 2.1.1 Basic principles

In this approach (Miao, 2018), two harmonic potentials are added to the non-bonded component of the potential energy so as to lower the binding/unbinding free energy barrier (Figure 1). These potentials act on the following CVs: 1) the ligand-environment potential energy and (optionally) 2) the rest of the system potential energy. Both biasing potentials are capped at user-defined thresholds. Computing the correction to recover the unbiased transition rate requires the estimation of the potential of mean force (PMF) profile and free energy barrier as a function of a separate CV describing the binding process e.g., a distance between ligand and protein atoms (Miao, 2018). In the

---

1 These techniques include, among others, scaled MD (Sinko et al., 2013; Bernet et al., 2018), steered MD (Paci and Karplus, 2000; Potterton et al., 2019; Spiriti and Wong, 2021), targeted MD (Schlutter et al., 1994; Wolf et al., 2019), GAMES (Debnath and Parrinello, 2020) path-reweighting methods (Chodera et al., 2011; Donati et al., 2017; Kieninger and Keller, 2021) metadynamics of paths (Madelli et al., 2020) and many transition path sampling-derived methods (Pratt, 1986; Delgado et al., 1998; Van Erp et al., 2003). A brief review of some of these methods (namely scaled MD, targeted MD and GAMES) is given in the supplementary material.

2 A Kramers’ rate theory correction has been developed for scaled MD as well but it has not been applied to the calculation of full dissociation rates (see Supplementary Material S3.3).
closely related Pep-GaMD method, developed specifically for simulating peptides unbinding from their target protein, the harmonic “boost” potentials are applied to the total potential (both non-bonded and bonded components) along the CVs (Wang and Miao, 2020). The application of the additional boost potential to the bonded component of the peptide potential energy accelerates the sampling of its conformational flexibility.

2.1.2 Applications
So far, the approach has been successfully applied to the ligand benzamidine targeting the trypsin enzyme (Miao, Bhattarai and Wang, 2020), using the AMBER14SB (Maier et al., 2015) and GAFF (Wang et al., 2004) force-fields. The calculated $k_{\text{off}} = 3.53 \pm 1.41 \text{s}^{-1}$ was two orders of magnitude smaller than the experimental value of $600 \pm 300 \text{s}^{-1}$ (Guillian and Thusias, 1970). The simulations required a cumulative 5 μs of MD for the estimation of $k_{\text{off}}$. Pep-GaMD has been used to investigate the un/binding of three model peptides that target the SH3 domain—one of which (PDB:1CKB) has an experimentally determined $k_{\text{off}}$ available for comparison to the computed value. Employing the AMBER14SB (Maier et al., 2015) force field and an aggregate simulation time of 3 μs, a $k_{\text{off}}$ of $1.45 \pm 1.17 \cdot 10^3 \text{s}^{-1}$ was computed for 1CKB; a result that is in close agreement with the experimental value of $8.9 \cdot 10^3 \text{s}^{-1}$ (Xue et al., 2014).

2.2 Dissipation-Corrected Targeted Molecular Dynamics (dcTMD)

2.2.1 Basic principles
This method (Wolf and Stock, 2018) assumes that unbinding processes (along with binding processes) can be described by the 1-dimensional Langevin dynamics of a suitable CV. The approach requires the determination of the free energy profile and the Langevin friction coefficient as a function of such a CV. These can be calculated from a nonequilibrium targeted molecular dynamics simulation (Schlitter et al., 1994) (see Supplementary Material S4), where a pulling force drives the system at a constant speed along the CV. Dissociation rates could then be obtained by performing the unbiased 1-dimensional Langevin dynamics simulations (Wolf and Stock, 2018). Despite the simplification, the timescales of ligand unbinding processes at room temperature still lead to prohibitively expensive simulations. To tackle this problem, the authors later introduced an approach that uses Kramers’ theory to correct the rates obtained from Langevin simulations performed at higher temperatures. (Wolf et al., 2020).

2.2.2 Applications
The method has been successfully applied to the calculation of $k_{\text{off}}$ of the trypsin-benzamidine complex, and the complex between a resorcinol scaffold-based inhibitor and the HSP90 protein. The calculated values $270 \pm 40 \text{s}^{-1}$ and $1.6 \pm 0.2 \cdot 10^{-3} \text{s}^{-1}$ respectively, agree well with the experimental values of $600 \pm 300 \text{s}^{-1}$ (Guillian and Thusias, 1970) and $3.4 \pm 0.2 \cdot 10^{-2} \text{s}^{-1}$ (Amaral et al., 2017). These predictions required an aggregate of $\sim 1.5 \times 10^4 \text{ms}$ of Langevin simulations and used the AMBER99SB* force-field (Best and Hummer, 2009).

2.3 τRAMD

2.3.1 Basic principles
The τ-random acceleration molecular dynamics (τRAMD) (Kokh et al., 2018) protocol is a quasi-biased method that evolved from random acceleration molecular dynamics (RAMD) (Lüdemann et al., 2000). τRAMD simulations of ligand-protein systems proceed similarly to standard MD simulations, without the need for any prior parameter fitting, characterization of CVs or binding pathways. The user specifies the magnitude of a randomly oriented force that is applied to the ligand to accelerate its dissociation from the binding pocket at each...
applied force on the receptor-ligand coupling (which previously
theory-based rescaling factor to correct for the in
application was demonstrated by Maximova and co-workers
computed
focused on qualitatively ranking ligands according to their
2.3.2 Applications
from different starting con
checkpoint, allowing for the observation of dissociation pathways
within several nanoseconds of simulation time. The magnitude of
the force dictates the duration of simulation time that is required
and is reported to have a minimal effect on the accuracy of
computed residence times. The direction of the force is reassigned
after each checkpoint until the ligand COM moves past a certain
distance threshold from its previous position. If the deviation of
the ligand COM meets or exceeds this threshold after the force is
applied, the direction of the force is retained until the following
checkpoint. An ensemble of unbinding simulations is spawned
from different starting configurations and velocities, and the
ensemble-averaged residence time is calculated from the
bootstrapped distribution of the time taken for dissociation
2.3.2 Applications
The earliest applications of rRAMD for unbinding kinetics
focused on qualitatively ranking ligands according to their
computed $k_{\text{off}}$ values (see Supplementary Material S5) (Kokh
et al., 2018, 2019, 2020). Recently, the first quantitative rRAMD
application was demonstrated by Maximova and co-workers
(Maximova et al., 2021), who formulated a Kramers’ rate
theory-based rescaling factor to correct for the influence of the
applied force on the receptor-ligand coupling (which previously
limited the method to qualitative ranking) to obtain a quantitative
$k_{\text{off}}$ estimate for the drug Isoniazid unbinding from the catalase
enzyme. Using seven trajectories (with applied forces of different
magnitudes), and the CHARMM36 forcefield (Best et al., 2012), a
$k_{\text{off}}$ of $2.8 \pm 3.7 \cdot 10^{-2} \text{s}^{-1}$ was computed—a result which agreed
very well with the experimental equivalent of $2.0 \pm 0.3 \cdot 10^{-2}
(Singh et al., 2008).

2.4 Metadynamics-Derived Methods
2.4.1 Basic principles
Well-tempered Metadynamics (MetaD) (Laio and Parrinello,
2002) is an exact free-energy method (Barducci et al., 2008;
Dama et al., 2014). It draws inspiration from earlier CV-based
enhanced sampling techniques such as local elevation (Huber
et al., 1994), Wang-Landau (Wang and Landau, 2001),
conformational flooding (Grubmüller, 1995), and adaptive
umbrella sampling techniques (Hooft et al., 1992; Bartels and
Karplus, 1997). In MetaD, a history-dependent bias potential
$B_t(s)$ is built iteratively by adding Gaussian functions (as
approximations of CV histograms) to the potential at regular
intervals throughout the simulations. Several different bias-
deposition schemes have been devised (Bussi and Laio, 2020).
Ultimately, convergence is achieved when the sum of the free
energy surface and the bias potential produces a flat landscape
that results in diffusive dynamics in CV space (see Figure 2). It is
then possible to compute the free energy surface along the CV via
rewighting methods, such as Weighted Histogram Analysis
Method (WHAM) (Kumar et al., 1992), Multistate Bennet
Acceptance Ratio (MBAR) (Shirts and Chodera, 2008), or
other estimators (Tiwary and Parrinello, 2015; Schäfer and
Settanni, 2020).

MetaD has been extended to allow recovery of the kinetics of
the unbiased ensemble. The method speeds up the calculation of
kinetic rates by filling up the starting free energy basin so as to
reduce the activation free energy barrier to $\sim k_B T$. This way, the
biased residence time of the system in the initial state is small
enough to allow multiple observations of the transition.
Transition times obtained in the biased ensemble are then
scaled to recover the unbiased kinetics. Following the
approaches of Grubmüller (Conformational flooding
(Grubmüller, 1995)) and Voter (Hyperdynamics (Voter,
1997)), the unbiased transition time is connected to the biased
time by:

$$ t_{\text{unbiased}} = \alpha t_{\text{biased}} $$

where $\beta = (k_B T)^{-1}$, $B_t(s)$ is the history-dependent bias potential,
and $\Delta t$ is the time step. For this last equation to be valid, no bias
should be present on the transition state. In the so-called
infrequent MetaD variant (Tiwary and Parrinello, 2013), the
Gaussians are deposited less frequently in barrier regions than
they are in standard MetaD, thus lowering the probability of
adding bias to the transition state. In frequency-adaptive (FA)
MetaD (Wang et al., 2018), the time interval between bias
depositions is gradually increased as the system approaches
the transition state. After an initial fast filling of the free
energy minimum, the same deposition rate as infrequent
MetaD is achieved. This way, results are obtained at a lower

![Figure 2](image-url)
computational cost compared to standard infrequent MetaD. Recently, an alternative method to infrequent and frequency-adapted MetaD has been presented. (Ansari et al., 2022) This method builds on a variant of MetaD called on-the-fly probability enhanced sampling (OPES). (Invernizzi and Parrinello, 2020) In the new approach, called OPES-flooding, the bias is constructed in a fast but controlled manner to fill the starting metastable basin up to a user-defined threshold value to automatically avoid depositing bias on the transition state. Usually, the standard protocol adopted in infrequent, FA-MetaD and OPES-flooding consists of running multiple independent simulations that yield an empirical distribution of residence times. A statistical analysis based on the Kolmogorov-Smirnov (KS) test (Salvalaglio et al., 2014), details in Supplementary Material S1) is then used to verify a posteriori that the transition state was indeed untainted.

### 2.4.2 Applications

Infrequent MetaD simulations based on the OPLS force-field (Kaminski et al., 2001) were used to study the unbinding of the ligand dasatinib from its target c-Src kinase (Tiwary et al., 2017). The CVs were the distance between the ligand and the binding pocket and the solvation state of the binding pocket. The calculated \( k_{\text{off}} \) of \( 4.8 \pm 2.4 \cdot 10^{-2} \text{s}^{-1} \) of dasatinib to c-Src obtained from 12 unbinding trajectories agreed well with an experimental value \( (5.6 \cdot 10^{-2} \text{s}^{-1}, \text{measured indirectly from } k_{\text{on}}) \) published by (Shan et al., 2009), but differed from a second value obtained for a fluorophore-tagged analogue \( (1.8 \cdot 10^{-3} \text{ to } 7.9 \cdot 10^{-4} \text{s}^{-1}) \) (Kwarcinski et al., 2016). A similar protocol was used to calculate \( k_{\text{off}} \) for 1-(3-(tert-butyl)-1-(p-tolyl)-1H-pyrazol-5-yl) urea, an inhibitor of p38 MAP II kinase belonging to the BIRB-796 family, this time using AMBER99SB-ILDN (Hornak et al., 2006; Lindorff-Larsen et al., 2010) and GAFF force-fields (Wang et al., 2004; Wang J. et al., 2006). After 17 independent unbinding events, the calculated \( k_{\text{off}} \) \((0.020 \pm 0.011 \text{s}^{-1}) \) (Casasnovas et al., 2017) was almost one order of magnitude lower than the experimental value of \( 0.14 \text{s}^{-1} \) (Regan et al., 2003).

Two other CVs yielded very similar results simulating 10 unbinding events each, suggesting that the discrepancy between the calculated and experimental values most likely arises from uncertainty in the force field rather than the choice of CVs.

FA-MetaD and Infrequent MetaD were used by Wang and co-authors (Wang et al., 2018) to obtain \( k_{\text{off}} \) values for benzene and indole ligands from the binding pocket of the 1.99A mutant of T4 lysozyme using CHARMM22 (MacKerell et al., 1998; MacKerell et al., 2004) and CGenFF (Vanommeslaeghe et al., 2011). The calculated \( k_{\text{off}} \) for benzene lay within the range of \( 4 \text{–} 10 \text{s}^{-1} \), around 100-fold lower than the experimental value of \( 950 \pm 20 \text{s}^{-1} \) (Feher et al., 1996). Both MetaD protocols used the same force-field, sample size (20 replicas), and path-CVs (Branduardi et al., 2007; Wang et al., 2017). CHARMM36-based (Best et al., 2012) infrequent MetaD simulations (Mondal et al., 2018) yielded a \( k_{\text{off}} \) for benzene \( (270 \pm 100 \text{s}^{-1}) \) that was considerably closer to the experimental value. Although only the displacement between binding pocket and ligand centers-of-mass was used as the CV, and the sample size was smaller than that of the previous study by Wang et al, it is tempting to conclude that even a different version of the same force-field (CHARMM in this case) may significantly impact the result.

More recently, AMBER14SB-based (Maier et al., 2015) FA-MetaD simulations were applied to study the unbinding kinetics of a radioligand, iperoxo, from the M3 human muscarinic acetylcholine receptor (Capelli et al., 2020). The calculated \( k_{\text{off}} \) \((3.7 \pm 0.7 \cdot 10^{-4} \text{s}^{-1}) \) was two orders of magnitude smaller than the experimental value \( (1.0 \pm 0.2 \cdot 10^{-2} \text{s}^{-1}) \). Density Functional Theory (DFT)-based QM/MM calculations suggested that this estimation discrepancy may be ascribed, at least in part, to the lack of polarization and charge transfer effects lacking in standard biomolecular force fields (Capelli et al., 2020).
OPES-flooding simulations based on AMBER14SB (Maier et al., 2015) and GAFF (Wang et al., 2004) were recently applied to study the unbinding kinetics of the trypsin-benzamidine complex, unveiling the role of water in regulating the residence time. Notably, the authors identified two different unbinding pathways and were able to calculate the corresponding rates separately. The slowest rate of 687 s\(^{-1}\) that is supposed to dominate the residence time is in good agreement with the experimental value of 600 ± 300 s\(^{-1}\) (Guillain and Thusias, 1970).

3 MARKOV STATE MODELS

3.1 Basic Principles

Markov state models (MSMs) (Singhal et al., 2004) are discrete models describing the dynamics of a system in terms of transition probabilities between a finite set of metastable states. The fundamental ingredients of the method are 1) a discretization of the conformational space into (kinetically fast) microstates and 2) a transition matrix that describes the probability of observing the system in another microstate after a fixed lag time \(t\). An interpretable, coarse-grained model is then built by defining kinetically metastable macrostates as collections of microstates, and this model can provide \(k_{\text{off}}\) values. Figure 3 shows a simplified schematic depiction of the MSM construction pipeline. The lag time \(t\) must be long enough to ensure that transitions between states are approximately Markovian\(^4\) and short enough for the model to represent all relevant fast processes. It should be chosen to be faster than association events to avoid systematic overestimation of the residence time (Paul et al., 2017).

When this is not possible, \(k_{\text{off}}\) can still be estimated from rate matrices rather than transition matrices (Kalbfeisch and Lawless, 1985; Crommelin and Vanden-Eijnden, 2009). However, rate matrix estimation is not unique and different approaches can result in residence times that differ even by an order of magnitude (Paul et al., 2017).

The input data to build MSMs can come from an ensemble of unbiased MD trajectories that sample dissociation events. However, generating this data is usually prohibitively expensive. Hence, several powerful schemes have been designed to enable the estimation of second-long residence times from relatively short MD simulations. These include adaptive restarting strategies (Bowman et al., 2010; Wan and Voelz, 2020) and/or biased simulations (Wu et al., 2016; Paul et al., 2017; Stelzl et al., 2017). In particular, recently developed estimators such as transition-based reweighting analysis TRAM (Wu et al., 2016) and its MBAR variant TRAMMBAR (Paul et al., 2017) require only irreversible visits to metastable states in the unbiased MD (as long as these states are sampled reversibly in the biased ones) and can greatly alleviate the sampling problem.

\(^4\)i.e., the probability of observing the system in a state \(y\) after the lag time given that it was in state \(x\) does not depend on the states of the system before \(x\).

3.2 Applications

MSM calculations on the trypsin-benzamidine complex (Plattner and Noé, 2015) (methodological details in Supplementary Material S6) yielded a \(k_{\text{off}}\) of 131 ± 109 × 10\(^{-3}\) s\(^{-1}\), which compares fairly well with experiments (\(k_{\text{off}} = 600 ± 300\) s\(^{-1}\)) (Guillain and Thusias, 1970). However, the high level of uncertainty suggests that sampling of unbinding events might be insufficient despite the large amount of aggregate simulation time (149.1 μs in this case). The dissociation of benzene from the L99A mutant T4 Lysozyme was investigated in a hybrid MSM/infrequent MetaD study (Mondal et al., 2018) using the CHARMM36 force-field (Best et al., 2012). The MSM was constructed from unbiased MD trajectories, and gave a \(k_{\text{off}}\) of 310 ± 130 s\(^{-1}\), which was marginally closer to the experimental \(k_{\text{off}}\) (950 ± 20 s\(^{-1}\)) (Feher et al., 1996) than the value reported by the accompanying infrequent MetaD simulations (\(k_{\text{off}} = 270 ± 100\) s\(^{-1}\)) (Mondal et al., 2018) and considerably closer than the previous FA-MetaD-based predictions (see Table 1) (Wang et al., 2018). However, the statistical uncertainty in the MSM-derived \(k_{\text{off}}\) was quite large, and the calculation required more simulation time (60 μs) compared biased MD approaches to obtain similar uncertainties: FA-MetaD/Infrequent MetaD studies typically require 6–12 μs (Casasnovas et al., 2017; Wang et al., 2018; Capelli et al., 2020) and LiGaMD (Miao et al., 2020) required ~5 μs.

The use of biased simulations can greatly reduce the sampling requirements. Wu et al. (2016) showed that by integrating unbiased MD with umbrella sampling simulation data, only 5%-10% of the unbiased data was necessary to estimate the dissociation rate of the trypsin-benzamidine complex up to statistical significance (\(k_{\text{off}} = 1170\) s\(^{-1}\) [617 s\(^{-1}\), 2120 s\(^{-1}\)]). A combination of 500 μs of unbiased MD and 1 μs of Hamiltonian replica exchange simulation was used to create an MSM model describing the binding of the oncoprotein fragment Mdm2 and a peptide inhibitor PMI. Estimates based on two different post-processing schemes yielded values of \(k_{\text{off}} = 0.125 \pm 0.025\) s\(^{-1}\), 0.66 s\(^{-1}\) and \(k_{\text{off}} = 1.13 \pm 0.48\) s\(^{-1}\), 1.33 s\(^{-1}\), corresponding to a 10–30-fold overestimation relative to experiments (\(k_{\text{off}} = 0.037\) s\(^{-1}\) [0.029 s\(^{-1}\), 0.04 s\(^{-1}\)]) (Paul et al., 2017).

4 PATH SAMPLING METHODS

Path sampling methods focus on generating an ensemble of transition pathways between bound and unbound states. Typically, this class of methods accelerates the unbinding event by exploiting restarting strategies to favor the sampling of short trajectories in the vicinity of the transition state, which are then used to reconstruct the full unbinding process. Weighted Ensemble (WE) (Huber and Kim, 1996), milestoning (Cho et al., 2006; Elber, 2007), transition state-partial path interface sampling (TS-PPTIS) (Jurasek et al., 2013), and adaptive multilevel splitting (AMS) (Cérou and Guyader, 2007; Cérou et al., 2011) are path sampling methods that were employed in calculations of \(k_{\text{off}}\) for ligand/protein complexes.
| Target                  | Technique    | T [K] | Force field       | $k_{\text{off}}$ (sim) [s$^{-1}$] | $k_{\text{off}}$ (Exp) [s$^{-1}$] | Simulation time [µs] | Ref                                      | Year  |
|------------------------|--------------|-------|-------------------|----------------------------------|----------------------------------|----------------------|-----------------------------------------|-------|
| Trypsin/Benzamidine    | SEEKR        | 298   | Amber14SB + GAFF   | 83 ± 14                          | 600 ± 300                        | 19                   | 10.1021/jacsocb.8b003988               | 2017  |
| Trypsin/Benzamidine    | SEEKR        | 298   | Amber14SB + GAFF   | 174 ± 9                          | 600 ± 300                        | 4.4                  | 10.1021/jacsocb.8c00495               | 2020  |
| Trypsin/Benzamidine    | SEEKR2       | 298   | Amber14SB + GAFF   | 990 ± 70                         | 600 ± 300                        | 5                    | 10.26434/chemrxv-2021-pfps             | 2021  |
| Trypsin/Benzamidine    | M-WEM        | 298   | Amber14SB + GAFF   | 791 ± 197                        | 600 ± 300                        | 0.48                 | 10.1021/jacsocb.1c00803              | 2022  |
| Trypsin/Benzamidine    | Inf-MetaD    | 300   | Amber99SB-ILDN     | 9.1 ± 2.5                        | 600 ± 300                        | 5                    | 10.1073/pnas.1424461112              | 2015  |
| Trypsin/Benzamidine    | Inf-MetaD    | 300   | Amber14SB + GAFF   | 4176 ± 324                       | 600 ± 300                        | —                    | 10.1021/jacsocb.8b00934              | 2019  |
| Trypsin/Benzamidine    | MSM          | 298   | Amber99SB + GAFF   | (9.5 ± 3.3)·10$^4$               | 600 ± 300                        | 50                   | 10.1073/pnas.1103547108              | 2011  |
| Trypsin/Benzamidine    | MSM          | —     | —                 | 2.8 ·10$^4$                      | 600 ± 300                        | 7.7                  | 10.1021/citc.004919                   | 2014  |
| Trypsin/Benzamidine    | MSM          | —     | Amber99SB + GAFF   | 131 ± 109                        | 600 ± 300                        | 149.1                | 10.1038/ncomms8663                   | 2015  |
| Trypsin/Benzamidine    | WExplore     | 300   | Charmm36 + CGenFF  | 5.56 ·10$^4$                     | 600 ± 300                        | 4.1                  | 10.1016/j.bpj.a2017.01.006           | 2017  |
| Trypsin/Benzamidine    | REVO         | 300   | Charmm36 + CGenFF  | 2660                             | 600 ± 300                        | 8.75                 | 10.1063/jacsocb.1003521              | 2019  |
| Trypsin/Benzamidine    | LiGaMD       | 300   | Amber14SB + GAFF   | 3.53 ± 1.41                      | 600 ± 300                        | 5                    | 10.1021/jacsocb.6c00395              | 2020  |
| Trypsin/Benzamidine    | dcTMD        | 290   | Amber99SB*         | 270 ± 40                         | 600 ± 300                        | 10000$^c$            | 10.1038/s41467-020-16655-1            | 2020  |
| Trypsin/Benzamidine    | AMS          | 298   | Charmm36 + CGenFF  | 260 ± 240                        | 600 ± 300                        | 2.3                  | 10.1021/jacsocb.6b00277              | 2016  |
| Trypsin/Benzamidine    | OPSES        | 300   | Amber14SB + GAFF   | 887                              | 600 ± 300                        | 3.2                  | arXiv:2204.05572                     | 2022  |
| T4L L99A-Benzene       | In-MetaD     | 300   | Charmm22*          | 6.0 ± 2.2                        | 950 ± 200                        | 6.7                  | 10.1039/c7sc01627a                    | 2017  |
| T4L L99A-Benzene       | FA-MetaD     | 300   | Charmm22*          | 5.7 ± 2.3                        | 950 ± 200                        | 5.5                  | 10.1063/jacsocb.15024679             | 2018  |
| T4L L99A-Benzene       | In-MetaD     | 303   | Charmm36           | 270 ± 100                        | 950 ± 200                        | —                    | 10.1371/journal.pcbi.1006180         | 2018  |
| T4L L99A-Benzene       | MSM          | 303   | Charmm36           | 310 ± 130                        | 950 ± 200                        | 60                   | 10.1371/journal.pcbi.1006180         | 2018  |
| T4L L99A-Indole        | In-MetaD     | 300   | Charmm22* + CGenFF | 9.8 ± 10.2                       | 325 ± 75b                       | 4.5                  | 10.1063/jacsocb.15024679             | 2018  |
| T4L L99A-Indole        | FA-MetaD     | 300   | Charmm22* + CGenFF | 6.0 ± 3.7                        | 325 ± 75b                       | 2.0                  | 10.1063/jacsocb.15024679             | 2018  |
| µOpioid receptor-      | In-MetaD     | 300   | Charmm36 + CGenFF  | (5.7 ± 0.5)·10$^{-2}$            | (2.3 ± 0.2)·10$^{-2}$            | 6                    | 10.1063/5.0019100                     | 2020  |
| µOpioid receptor-       | In-MetaD     | 300   | Charmm36 + CGenFF  | (2.1 ± 0.3)·10$^{-2}$            | (1.8 ± 0.3)·10$^{-3}$            | 19                   | 10.1063/5.0019100                     | 2020  |
| µOpioid receptor-Fentanyl | In-MetaD   | 310   | Charmm36m + CGenFF | (2.6 ± 0.8)·10$^{-2}$ (HD) (3.8 ± 1.4)·10$^{-3}$ (HIP) (D1)6.4 · 10$^{-3}$ (D2)6.6,7,10$^1$ (D3)6.4 · 10$^{-3}$ (D4)4.1 · 10$^{-3}$ (A)6.0 · 10$^{-4}$ (D1-D4 different docked poses) | 4.2 · 10$^{-3}$ | 6 | 10.1021/jacsocb.1c00341 | 2021 |
| TSPO-PK11195            | REVO         | 300   | Charmm36 + CGenFF  | 4.9 · 10$^{-4}$                  | 40                               |                      | 10.1016/j.bpj.2020.11.015            | 2021  |
| c-Src kinase-dasatinib | In-MetaD     | 300   | OPLS               | (4.8 ± 2.4)·10$^{-2}$            | 5.6 · 10$^{-2}$                  | 7–8                  | 10.1126/sciadv.1700014                | 2017  |
| Src kinase-imatinib    | TS-PPTIS     | 305   | Amber99SB*-ILDN + GAFF (QM/MM) | 0.026 | 0.11 ± 0.08 | — | 10.1021/jacsocb.6b00687 | 2018 |

(Continued on following page)
### 4.1 Weighted Ensemble Methods

#### 4.1.1 Basic principles

A set of unbiased molecular dynamics trajectories with equivalent statistical weights are spawned in parallel from a ligand/protein complex in the ground-state configuration (Huber and Kim, 1996). The configuration space is then subdivided into bins, which the trajectories/walkers navigate through. The weighted ensemble (WE) method aims to maintain a fixed number (N) of walkers per bin. Thus, the occupancy of the bins is calculated at specific resampling intervals $\tau_{\text{res}}$. If the number of walkers in a given bin is lower than $N$, one or more of the walkers are cloned, with each daughter trajectory receiving a fraction of the weight of the original. Conversely, in regions populated by a number of walkers exceeding $N$, two or more trajectories are merged, with the resulting trajectory inheriting the weights of its constituents (Zuckerman and Chong, 2017). This process results in a resampled trajectory space spanning the bound, intermediate and unbound states from which $k_{\text{off}}$ values can be obtained (Zhang et al., 2010).

Notably, the method does not require a detailed a priori definition of differentiable collective variables, and it is embarrassingly parallel. Given that the availability of Tier-0 and Tier-1 machines has grown significantly since the method was first formulated, several scalable open-source implementations have emerged. These include WExplore (Dickson and Brooks, 2014), Wepy (Lötz and Dickson, 2020), and REVO (Resampling of Ensembles by Variation Optimization) (Donyapour et al., 2019). The latter is a method featuring a novel resampling algorithm replacing bins in configurational space with a system-specific all-to-all pairwise distance matrix between walkers, thereby decreasing the correlation between trajectories. The novel concurrent adaptive sampling (CAS) algorithm (Ahn et al., 2017) builds on the traditional WE method by adaptively constructing macromolecules (represented by n-dimensional Voronoi cells) while approximating the committor function of each macromstate, and clustering the macromstates according to their committor functions as the simulation progresses. This improves the efficiency of WE simulations in high-dimensional systems, by directing computational power to sampling portions of configuration space that are closer to the “product” configuration.

#### 4.1.2 Applications

The $k_{\text{off}}$ of the trypsin-benzamidine complex as calculated by WExplore (5560 s$^{-1}$) (Dickson and Lotz, 2017) overestimated by one order of magnitude the experimental value (600 s$^{-1}$) (Güllian and Thusias, 1970). This value was calculated from five independent WExplore runs, corresponding to an aggregate simulation time of 4.1 μs. Using clustering-based confirmation space network analysis techniques (Dickson and Brooks, 2013), three distinct ligand exit pathways were unearthed from the trajectories. The trypsin-benzamidine system was later investigated again with REVO (Donyapour et al., 2019). Based on five independent REVO runs, giving a total of 8.75 μs, a $k_{\text{off}}$ of 2660 s$^{-1}$ was predicted—an important improvement on WExplore, but an overestimation nonetheless. WExplore was also employed to estimate the dissociation rate of the TPPU inhibitor from soluble epoxide hydrolase. The calculated $k_{\text{off}}$ (2.4 $\cdot$ 10$^{-2}$ s$^{-1}$ [3.6 $\cdot$ 10$^{-2}$ s$^{-1}$, 4.4 $\cdot$ 10$^{-2}$ s$^{-1}$]) was one order of magnitude greater than the experimental value of 3.6 $\cdot$ 10$^{-3}$ s$^{-1}$ (Lötz and Dickson, 2018), and required 6 μs of simulation time to compute. However, the reason for the systematic overestimations of $k_{\text{off}}$ is not explicitly addressed. REVO was recently employed (Dixon et al., 2021) to quantify $k_{\text{off}}$ values for five distinct bound poses of the PK11195 radioligand in complex with TSPO (see Table 1), using a cumulative 5.18 μs of simulation time per pose. The calculated

### Table 1 (Continued) Quantitative in silico calculations (we highlighted in boldface the simulations that are below one order of magnitude for the predicted results with respect to the experimental ones)

| Target | Technique | $T$ [K] | Force field | $k_{\text{off}}$ ($\text{sim}$) [s$^{-1}$] | $k_{\text{off}}$ (Exp) [s$^{-1}$] | Simulation time [μs] | Ref | Year |
|--------|-----------|---------|-------------|---------------------------------|------------------|---------------------|-----|------|
| Epoxide Hydrolase-TPPU | WExplore | 300 | Charmm36 + Cx2FF + Amber99SB-ILDN + GAFF | 2.4 $\cdot$ 10$^{-2}$ [3.6 $\cdot$ 10$^{-2}$ s$^{-1}$, 4.4 $\cdot$ 10$^{-2}$ s$^{-1}$] | 1.5 $\cdot$ 10$^{-3}$ | 6 | 10.1021/jacs.7b08572 | 2018 |
| p38 kinase-1 (3-tert-buty)-1- (p-tolyl)-1H-pyrazol-5-ylurea | In-MetaD | 300 | Amber99SB-ILDN + GAFF | 0.020 ± 0.011 | 0.14 | 6.8 | 10.1021/jacs.6b12950 | 2017 |
| M2 muscarinic receptor/iperoxo | FA-MetaD | 310 | Amber14SB + GAFF | (3.7 ± 0.7)$\cdot 10^{-4}$ | (1.0 ± 0.2)$\cdot 10^{-2}$ | 8 | 10.1021/acscplett.0c00999 | 2020 |
| HSP90-inhibitor | dcTMD | 300 | Amber99SB + GAFF | (1.6 ± 0.2)$\cdot 10^{-3}$ | (3.4 ± 0.2)$\cdot 10^{-2}$ | 5000c | 10.1038/s41467-020-16655-1 | 2020 |
| Mdm2/PMI | MSM | 300 | Amber99SB-ILDN | 0.125 [0.025, 0.68] | 1.13 [0.48, 1.33] | Different rate matrix estimators | 0.037 [0.029, 0.04] | 500 | 10.1038/s41467-017-01163-6 | 2017 |
| Mdm2/p53 | MSM | 300 | Amber99SB-ILDN | 1.9$\cdot 10^{5}$ | 2.1 | 831 | 10.1016/ | jbbp.2017.07.009 | 2017 |
| SH3 Domain—1CKB | Pep-GaMD | 300 | Amber99SB-NMR | (1.45 ± 1.17)$\cdot 10^{-3}$ | 8.9 $\cdot 10^{-3}$ | 3 | 10.1063/5.0021399 | 2020 |
| MIKαT3—Isoniazid | rRAMD + extrapolation | 300 | CHARMM36 + SwissParam | (2.8 ± 3.7)$\cdot 10^{-2}$ | (2.0 ± 0.3)$\cdot 10^{-2}$ | — | 10.1021/acscplett.1c02952 | 2021 |

*The Authors in the original work considered the experimental $k_{\text{off}}$ at 293 K (800 ± 200 s$^{-1}$), while they simulated the system at 300 K. Here we choose to put the value at the closest temperature available in experiments (303K). Both the experimental values come from (Feher et al., 1996). The experimental value has been measured at 293 K.*

*For dcTMD, computational time is referred to 1D Langevin simulator, and the authors says that “1 ms of simulation time at a 5 fs time step take ~6 h of wall-clock time on a single CPU.”
values for the poses spanned five orders of magnitude, and the pose with the most favorable docking score (pose D1, $k_{off} = 6.4 \times 10^{-5} \text{s}^{-1}$) exhibited the closest agreement with the experimental value ($4.9 \times 10^{-4} \text{s}^{-1}$) out of all the docked poses. All the of the studies described here made use of the CHARMM36 (Best et al., 2012) and CGenFF (Vanommeslaeghe et al., 2011) force fields. At present, the CAS method described in Section 4.1.1 has been successfully applied to host-guest systems only (Ahn et al., 2020).

4.2 Milestoning

4.2.1 Basic principles

Here, the configuration space is treated as a coarse mesh characterized by slowly relaxing variables, such as native contacts and/or distances between chemical groups that describe the ligand unbinding process (Cho et al., 2006; Elber, 2007). The mesh must be coarse enough for distinct long-lived metastable states to emerge, but fine enough to ensure that transitions between the interfaces between the mesh’s cells or “milestones” are accessible in MD simulations. Equilibrium configurations for each milestone are usually generated with “pulling” SMD simulations, or with a series of MD simulations in which the diffusing group is harmonically restrained to the milestone surface. Afterward, a set of trajectories is spawned from each milestone, and whenever a trajectory reaches a new neighboring milestone, it is terminated. In practice, the criteria for termination of trajectories vary depending on the implementation. The lifetime and flux (i.e., number of trajectories passing through the milestone per unit time) associated with each milestone may be used to compute the ligand residence time (Elber, 2020).

Practical implementations of milestone in ligand unbinding studies fall into two categories: 1) The Simulation Enabled Estimation of Kinetic Rates (SEEKR) (Volapka et al., 2017) approach, which exploits milestone theory in a multiscale framework based on MD and Brownian Dynamics (BD) simulations (Luty et al., 1993). The milestones are nested spherical shells surrounding the binding pocket. Transitions between milestones close to the binding pocket are simulated using all-atom MD. Meanwhile, transitions between the more diffuse milestones further away are described by cheaper BD simulations—where fast sampling of rigid body interactions is more important than detailed sampling of ligand conformations. An updated implementation of SEEKR, named MMVT SEEKR, has been subsequently proposed (Jagger et al., 2020): it circumvents the need to compute the equilibrium distribution for all the milestones, reducing the computational time needed to compute kinetics constants. 2) The recently formulated weighted ensemble milestoneing (WEM) methods combine milestoneing theory with WE methods. Here, the configurational space between the milestones is split into bins, and WE simulations are run in between milestones to achieve faster convergence (Ray and Andricioaei, 2020).

4.2.2 Applications

All applications to kinetics of biological systems so far are based on Amber14SB (Maier et al., 2015) and Gaff (Wang et al., 2004) force fields and applied to the trypsin-benzamidine complex. SEEKR yielded a $k_{off}$ of $83 \pm 14 \text{s}^{-1}$ for the trypsin-benzamidine system using 19 μs of aggregate MD and ten spherical milestones. These results underestimate the experimental value ($600 \pm 300 \text{s}^{-1}$) (Guillian and Thusias, 1970). MMVT SEEKR improved the $k_{off}$ estimate ($174 \pm 9 \text{s}^{-1}$), with only a quarter of the aggregate simulation time (4.4 μs) used in the prior SEEKR study. WEM (Ray and Andricioaei, 2020) gave a further improvement $k_{off} = 791 \pm 197 \text{s}^{-1}$, using a mere 0.5 μs of simulation time (Ray et al., 2022).

4.3 Transition State-Partial Path Transition Interface Sampling

4.3.1 Basic principles

In transition state-partial path transition interface sampling (TS-PPTIS) (Iuraszek et al., 2013) an initial metadynamics calculation is performed to determine the transition state and the free energy barrier along a given CV. Then, the transmission coefficient is estimated, similarly to the PPTIS method (Van Erp et al., 2003; Moroni et al., 2004) by foliating the barrier region along the CV with interfaces and sampling short trajectories spanning three consecutive interfaces. These trajectories are sampled using transition path sampling (Pratt, 1986; Dellago et al., 1998). Under the assumptions that the dynamics in the barrier region is diffusive and there are no memory effects for travelled distances beyond two interfaces, the kinetic rates are independent of the CV.

4.3.2 Applications

TS-PPTIS was used to compute the $k_{off}$ of the imatinib-Src kinase complex (Morando et al., 2016). The calculation used 5 CVs: 2 path collective variables (Branduardi et al., 2007), a CV counting the number of water molecules interacting with the ligand and the binding cavity, and two distances between key residues of Src characterizing the motion of the kinase A-loop. Using AMBER99SB*-ILDN and Gaff, the authors computed a value of $k_{off} = 0.0114 \text{s}^{-1} \pm 0.001 \text{s}^{-1} \pm 0.139 \text{s}^{-1}$, which is slow (but within statistical significance) compared to experiments ($k_{off} = 0.11 \pm 0.08 \text{s}^{-1}$). In a separate work (Haldar et al., 2018), the authors refined the prediction by computing a free energy correction from the MM to a hybrid quantum mechanics/molecular mechanics Hamiltonian using a replica exchange thermodynamic integration scheme (Woods et al., 2003) and Metropolis-Hastings Monte Carlo sampling (Woods et al., 2008). This correction does not account for dynamical effects but only for changes in the free energy. The computed correction to $k_{off}$ was small but consistent with faster dissociation dynamics obtaining a corrected value of $k_{off} = 0.026 \text{s}^{-1}$.

4.4 Adaptive Multilevel Splitting

4.4.1 Basic principles

Similarly to WE, adaptive multilevel splitting (AMS) (Cérou and Guyader, 2007; Cérou et al., 2011) relies on a set of trajectories that are systematically cloned or killed. However, AMS does not require bins. Instead, the algorithm is initialized by generating a set of “loop” trajectories starting and ending in the bound state. At each iteration, the replica that travelled the least distance $d$
from the bound state (measured through a CV) is killed, and a new loop is created by restarting a simulation from a point at the same distance $d$ previously visited by one of the remaining replicas. This is repeated until all loops travelled a distance above a threshold value (defining the unbound state) before returning to the bound state. The dissociation rate is then estimated from this collection of trajectories.

### 4.4.2 Applications

AMS was used to calculate the dissociation rate of trypsin-benzamidine using the CHARMM36 force field (Best et al., 2012) for trypsin and CGenFF (Vanommeslaeghe et al., 2011) for the ligand (Teo et al., 2016). As the CV, the authors used the distance between the center of mass of benzamidine and the alpha carbons of the amino acids close to the binding site. A suitable value for the threshold value of the CV was obtained through a steered MD simulation. Furthermore, 130 ns unbiased MD value for the threshold value of the CV was obtained through a steered MD simulation. As the CV, the authors used the distance between the center of mass of benzamidine and the alpha carbons of the amino acids close to the binding site. A suitable value for the threshold value of the CV was obtained through a steered MD simulation. Furthermore, 130 ns unbiased MD simulation was run to estimate the average time of a looping trajectory under the assumption that the short loops thus sampled represented the large majority of loops and thus dominated the average loop duration. In total, 2.3 μs of simulations were used to obtain a $k_{\text{off}} = 260 \text{ s}^{-1} \pm 240 \text{ s}^{-1}$, in good agreement with experimental measurements.

### 5 LIMITATIONS ASSOCIATED WITH FORCE FIELDS

Table 1 summarizes the $k_{\text{off}}$ predictions of various ligand/protein systems obtained using the methods discussed in previous sections. For completeness, we also report the temperature, total simulation time, and force field used. In about one-third of the cases, spanning all different classes of methodologies and force fields, the theoretical predictions are in the same order of magnitude as the experimental values, and in a few cases (shown in boldface in Table 1) they reproduce them within statistical error. In most cases, however, calculated values show discrepancies from 1 to 2 orders of magnitude, regardless of the method and force field. Similarly, the only predictive study reported so far (Paul et al., 2017) reports values with an error of 1–2 orders of magnitude (albeit with large statistical errors) relative to experimental data performed afterwards. All these results, taken together, lead us to suggest that regular force fields may be, at times, not accurate enough to predict $k_{\text{off}}$ values.

Dissociation rate is then estimated from this collection of trajectories. The dissociation rate is then estimated from this collection of trajectories. The dissociation rate is then estimated from this collection of trajectories. The dissociation rate is then estimated from this collection of trajectories.

#### 5.1 Force Field Dependence of the Results

The published data indicate that careful parametrization of the force fields is essential to obtain $k_{\text{off}}$ predictions. Comparison between the results obtained from brute force MD calculations on a set of ligands binding to a β-cyclodextrin (βCD) host showed that $k_{\text{off}}$ predictions parametrizing βCD with the Q4MD force field (Cézard et al., 2011) were consistently more accurate (within one order of magnitude of experimental values) than the GAFF-based (Wang et al., 2004) estimates (Tang and Chang, 2018). On the other hand, the $k_{\text{on}}$ estimates were consistently better for the GAFF model, which points to the difficulty of obtaining transferable potentials. In the case of benzene unbinding from L99A T4 lysozyme, infrequent MetaD simulations using CHARMM22 (MacKerrell et al., 1998; MacKerrell Jr. et al., 2004) yielded a significantly underestimated $k_{\text{off}}$ in the range of 4–10 s$^{-1}$ (Wang et al., 2018), while the same method combined with CHARMM36 (Best et al., 2012) produced a $k_{\text{off}}$ (270 ± 100 s$^{-1}$) (Mondal et al., 2018) considerably closer to the experimental value of 950 ± 20 s$^{-1}$ (Feher et al., 1996). Although different CVs were used in these two works (see Section 2.4.2), the effect of the force field cannot be ruled out. Indeed, the two force fields differ only in a few dihedral potential terms (Best et al., 2012) that control the rigidity of secondary structures, and in particular two helices of T4 which control benzene’s access to the binding pocket. Finally, we mention here the work of (Vitalini et al., 2015), where it was shown that slow relaxation timescales of two small peptides using five protein force fields (AMBER99SB-ILDN (Lindorff-Larsen et al., 2010), AMBERff03 (Duan et al., 2003), OPLS-AA/L (Kaminski et al., 2001), CHARMM27 (MacKerell et al., 2000), and GROMOS43a1 (Daura et al., 1998)) differ up to two orders of magnitude. Given the importance of slow protein conformational changes in unbinding kinetics (Plattner and Noé, 2015), this result further highlights the role of force fields for accurate rate calculations.

#### 5.2 Polarization and Charge Transfer Effects

Traditional force fields describe electrostatics using fixed point charges. This representation is extremely efficient and works remarkably well, even in the case of systems with high electric fields (Mironenko et al., 2021). However, such a scheme cannot adapt to changes in the electrostatic environment observed during ligand unbinding. Recently, some of us (Capelli et al., 2020) found that electrostatic effects contribute significantly to the force field misrepresentation of protein-ligand interactions at the transition state of the M2-iperoxo complex. Furthermore, the work of Haldar and coworkers (Haldar et al., 2018) showed that accounting for changes in charge distribution resulted in free energy corrections ranging from 1.9 to 4.7 kcal/mol as the ligand progressed from the hydrophobic binding pocket to the solvated state. Metalloenzymes (representing 40%–50% of all proteins in the PDB database (Chen et al., 2019)) and highly charged protein-ligand systems are also quite challenging to describe with traditional force fields (Li and Merz, 2017). Indeed, for the
latter systems, FF-based binding free energy calculations resulted in significant systematic errors (Rocklin et al., 2013). Overall, these results show that going beyond standard fixed-charged models is in many cases desirable to improve accuracy.

6 PERSPECTIVES: FROM POLARIZABLE FORCE FIELDS TO QM/MM CALCULATIONS TOWARDS THE EXASCALE

Force fields have been overwhelmingly successful in predicting equilibrium properties such as free energies of binding (Karplus and McCammon, 2002; Wang et al., 2015; Robustelli et al., 2022). Indeed, force fields are traditionally fitted to reproduce equilibrium experimental measurements (ensemble averages) and geometries obtained with quantum mechanical methods. As a result, their performance is expected to peak in the regions near the free energy minima (e.g., the bound state) rather than near the kinetically relevant transition states, where small errors are exponentially amplified in $k_{\text{off}}$ predictions. After observing discrepancies of two orders of magnitude in the kinetic predictions of several force fields, Vitalini and coworkers (Vitalini et al., 2015) suggested that kinetic information should be included in the fitting process. In general, designing new parametrization strategies for force fields is still a very active area of research (He et al., 2020; Giannos et al., 2021; Qiu et al., 2021). This is not surprising, given the issues discussed in Section 5. For example, methods to include polarization effects within a fixed-charge scheme (Kelly and Smith, 2020) and multisite models for transition metal ions have been developed (Li and Merz, 2017).

A different direction pursued by the modeling community is instead to use potential energy functions that go beyond the biomolecular force fields’ simple representation of electrostatics (e.g., polarizable force fields, hybrid quantum mechanics/molecular mechanics (QM/MM) calculations, machine learning potentials). Without any claim of being comprehensive, here we provide a brief perspective on the role of these methods in the upcoming era of exascale computing.

6.1 Polarizable Force Fields

Polarizable force fields for biomolecules (Jing et al., 2019) including AMBER ff02pol (Wang Z. X. et al., 2006), AMOEBA (Ponder et al., 2010), CHARMM Drude (Baker et al., 2010), CHARMM-FQ (Patel and Brooks, 2003; Patel and Brooks, 2004), SIBFA (Piquemal et al., 2007), and ABEEMoFF (Liu et al., 2017) aim at providing an empirical description electronic polarizability. Simulations based on these potentials could dramatically improve the modeling of transition states in cases where electronic polarization and charge transfer may be linked to non-trivial rearrangements of hydrogen bonds and hydrophobic interactions (Schmidtko et al., 2011; Schiebel et al., 2018). Although polarizable force fields have recently shown excellent accuracy in prospective predictions of binding affinities in model systems (Amezcua et al., 2022), to the best of our knowledge, they have not been used for protein-ligand $k_{\text{off}}$ predictions yet. Notably, in a very recent paper (Yue et al., 2022), it was shown how using a polarizable force field improved the accuracy of the predictions of anion permeation rates in fluoride channels compared to predictions based on standard fixed charge schemes, highlighting the necessity of using polarizable models for treating such processes. Although this is not a ligand/protein system, this work further showcases the limitations of conventional force fields in treating electrostatic interactions as well as the potential of polarizable models.

6.2 QM/MM Simulations

DFT-based QM/MM simulations treat a small region of interest (in our case this could be a ligand and the protein residues interacting with it) at the DFT level, while the overall computational cost is balanced by MM treatment of other regions (Kulik, 2018). The form of the potential energy is a hybrid model between classical mechanics and quantum chemistry:

$$U = U_{\text{QM}} + U_{\text{MM}} + U_{\text{QM/MM}}$$

where $U_{\text{QM/MM}}$ denotes the interaction between atom groups assigned to the QM region and MM region. DFT-based QM/MM simulations include both electronic polarizability and charge transfer effects (Blumberger, 2008; Capelli et al., 2020), and they address the problem of transferability, as they do not rely on optimizations against predefined training data sets. These approaches can tackle important biomedicine problems such as the study of transition-metal-based drugs binding to proteins (Caladriini et al., 2015) or the description of enzymatic reactions. (Carlani et al., 2002; Liao and Thiel, 2013; Roston et al., 2016; Caldararu et al., 2018; Kulik, 2018; Piniello et al., 2021). However, these simulations are orders of magnitude more expensive than any of the potentials described so far, and hence achieving high statistical accuracy with such an approach is obviously extremely challenging.

6.2.1 Parallel Computing in DFT-Based QM/MM

Modern supercomputers are currently breaching the exascale limit in the United States (Schneider, 2022), Japan, and China.7 Exascale calculations however remain one of the major challenges in molecular simulations (Hospital et al., 2015; Päll et al., 2015). Recent advances in massively scalable QM/MM codes, such as that developed in Juelich in collaboration with European universities (Olsen et al., 2019; Bolnykh et al., 2020a) (see

$\text{The } k_{\text{off}} \text{ values depend exponentially on the height of the dissociation free energy barrier, so even small inaccuracies in the potential energy may impact dramatically kinetics calculations.}$

7According to the Top500 list (https://www.top500.org/), which ranks computers based on their performance on the HPLinpack benchmark (Dongarra et al., 2003), currently there are no machines that have exceeded the exascale limit. The Fugaku supercomputer in Japan showed performances above one EFlop/s, but on a different benchmark (Kudo et al., 2020). In China, one or two exascale supercomputers might be already operating (Ma et al., 2022; Schneider, 2022).
Supplementary Material S7) and their successful applications to predict free energy landscapes associated with biological processes (Bolnykh et al., 2020a; Chiariello et al., 2020) brings us to suggest that in a not-too-far future QM/MM calculations may exploit the unprecedented power of exascale computing for direct MD simulations of ligand (un)binding (Bolnykh et al., 2020b, Bolnykh et al., 2021).

6.2.2 Machine Learning in QM/MM
Neural network models of the potential energy function have emerged as a promising route to obtaining near-DFT accuracies (Unke et al., 2021; Kocer et al., 2022) at a computational cost only 1–2 orders of magnitude slower than force fields. Applications to the kinetics of chemical reactions have been published (Stocker et al., 2020; Yang et al., 2022) and in principle, they could be used to model DFT-based QM/MM predictions of ligand poses during the unbinding process. However, ML potentials are currently still limited to small molecule applications and robust solutions to model long-range interactions have yet to emerge (Yue et al., 2021). The advent of exascale computing could dramatically expand the domain of applicability of such approaches (Lu et al., 2021). Moreover, several approaches to solve these issues have been proposed based on hybrid machine learning/molecular mechanics models (Shen and Yang, 2018; Rufa et al., 2020; Böselt et al., 2021; Gastegger et al., 2021) (see Supplementary Material S8 for details).

7 CONCLUSION
We have reviewed an array of rather diverse methods able to predict unbinding kinetics constants using atomistic representations of the biomolecules involved. These techniques have shown tremendous progress in the last years: considering trypsin-benzamidine as a benchmark system (as seen in Table 1), we start from 2–3 orders of magnitude in $k_{off}$ error in the pioneering MSMs of De Fabritiis and co-workers (Buch et al., 2011) to an error of less than 1 order of magnitude in some of the most recent calculations (Plattner and Noé, 2015; Votapka et al., 2017; Brotzakis et al., 2019; Wolf et al., 2020). Despite these impressive methodological advances, the domain of applicability and accuracy appears to be still limited by current force fields. Better parametrization and polarizable force fields (Lin and MacKerell, 2019) promise to improve the quality of the potential energy model at a reasonable cost at a reasonable computational cost (Lemkul et al., 2016). Another possibility is the use of massively parallel DFT-QM/MM complemented by ML techniques, which include electronic polarizability as well as charge transfer. This approach could address the issue of transferability of current biomolecular force fields. However, the accuracy of these approaches is yet to be established.

Traditionally, computational drug discovery has used a combination of methods such as docking (Ferreira et al., 2015), quantitative structure-activity relationship (QSAR) modeling (Dossetter et al., 2013), free-energy methods (Cournia et al., 2017), and (recently) ML-based approaches (Zhao et al., 2020) to improve the binding affinity of a compound during lead optimization. Computer-aided ligand design campaigns could enormously profit from the design of so-called transition state analogues which, in the case of enzyme inhibitors, have been correlated with release rates that are orders of magnitude slower than product release (Schramm, 2013; Schramm 2015; Svensson et al., 2015). We hope that approaches beyond the use of standard force fields, such as those discussed here, will lead in a not-too-distant future to the accurate description of the energetics and structural determinants of the unbinding transition states, giving an unprecedented boost to the discovery of promising new small molecules and the optimization of known drugs.

AUTHOR CONTRIBUTIONS
KA is the first author. AR and RC made equal contributions as second authors. DM is third author, WL is the fourth. Prof. PC is the PI and the last author.

FUNDING
AR, PC, and DM acknowledge support from the Helmholtz European Partnering program ("Innovative high-performance computing approaches for molecular neuromedicine"). KA and PC acknowledge funding from the Human Brain Project (EU Horizon 2020). W.L. appreciates the National Natural Science Foundation of China No. 21505134.

ACKNOWLEDGMENTS
The authors are indebted to Prof. Michele Parrinello for many discussions on the topic of kinetics of ligand binding to proteins.

SUPPLEMENTARY MATERIAL
The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmolb.2022.899805/full#supplementary-material

REFERENCES
Ahn, S.-H., Grate, J. W., and Darve, E. F. (2017). Efficiently Sampling Conformations and Pathways Using the Concurrent Adaptive Sampling (CAS) Algorithm. J. Chem. Phys. 147, 074115. doi:10.1063/1.4999097
Ahn, S.-H., Jagger, B. R., and Amaro, R. E. (2020). Ranking of Ligand Binding Kinetics Using a Weighted Ensemble Approach and Comparison with a Multiscale Milestoning Approach. J. Chem. Inf. Model. 60, 5340–5352. doi:10.1021/acs.jcim.9b00968
Amaral, M., Kohl, D. B., Bornke, J., Wegener, A., Buchstaller, H. P., Eggenweiler, H. M., et al. (2017). Protein Conformational Flexibility Modulates Kinetics and Thermodynamics of Drug Binding. Nat. Commun. 8, 2276. doi:10.1038/ncomms12258
Ameczua, M., Setiadi, J., Ge, Y., and Mobley, D. L. (2022). An Overview of the SAMPL8 Host-Guest Binding Challenge. ChemRxiv, 1–42. doi:10.26434/chemrxiv-2022-lwdf0h
