JCTC Journal of Chemical Theory and Computation

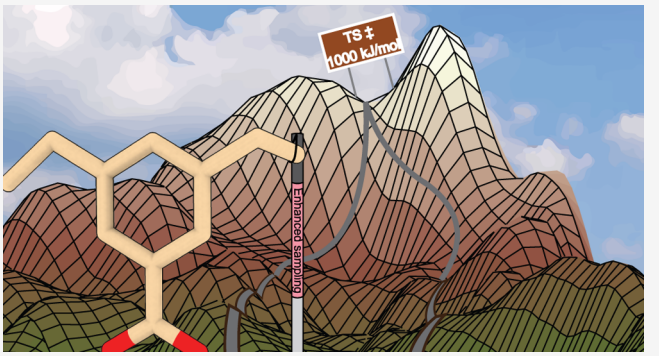
pubs.acs.org/JCTC
Review

Perspectives on Ligand/Protein Binding Kinetics Simulations: Force Fields, Machine Learning, Sampling, and User-Friendliness

Paolo Conflitti, Stefano Raniolo, and Vittorio Limongelli*

Cite This: J. Chem. Theory Comput. 2023, 19, 6047−6061
Read Online

ABSTRACT: Computational techniques applied to drug discovery have gained considerable popularity for their ability to filter potentially active drugs from inactive ones, reducing the time scale and costs of preclinical investigations. The main focus of these studies has historically been the search for compounds endowed with high affinity for a specific molecular target to ensure the formation of stable and long-lasting complexes. Recent evidence has also correlated the in vivo drug efficacy with its binding kinetics, thus opening new fascinating scenarios for ligand/protein binding kinetic simulations in drug discovery. The present article examines the state of the art in the field, providing a brief summary of the most popular and advanced ligand/protein binding kinetics techniques and evaluating their current limitations and the potential solutions to reach more accurate kinetic models. Particular emphasis is put on the need for a paradigm change in the present methodologies toward ligand and protein parametrization, the force field problem, characterization of the transition states, the sampling issue, and algorithms' performance, user-friendliness, and data openness.

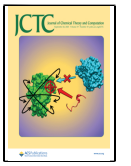


1. INTRODUCTION

The pharmacological properties of a drug are typically defined as the pharmacokinetic and pharmacodynamic properties. While pharmacokinetics regards the body effect on the drug defining its absorption, distribution, metabolism, excretion, and toxicity (i.e., ADMET), pharmacodynamics deals with the drug's effect on our body, which can be reasonably rationalized by its mechanism of action and the elucidation at atomistic scale of the drug binding interaction with its molecular target. The latter is of paramount relevance to guide drug discovery and is the topic of the present article. Despite the recent methodological advances, drug discovery remains a daunting task with poorly performing predictive models of in vivo drug efficacy,[1,2] dramatic time scale, and exorbitant costs (10−15 years and 2.6 billion US dollars on average to develop a new medication).[3,4] Considering also the low success rate of a drug to pass clinical trials (below 10%),[5] it is apparent that there is a tremendous need for techniques capable of increasing the probability that a ligand obtained from basic research could become a drug and at the same time reducing the costs of the research.[6−8] This need is even more urgent since December 2022 when the Food and Drug Administration (FDA) took the historic decision to replace the word "animal" with "nonclinical tests" in the law governing the agency's drug assessments, paving the way to nonanimal alternative methods, such as organoids, organs-on-a-chip and in silico modeling (Food and Drug Administration Modernization Act 2.0).[9] In this context, structure-based drug discovery (SBDD) will play an ever more prominent role. Traditionally, SBDD relies on drug/target binding studies (hereafter ligand/protein binding, LPB) focusing on ligand affinity to the target, expressed as ligand binding free energy ΔG, binding/dissociation constant K_b/K_d, or half maximal effective/inhibitory concentration EC50/IC50. However, this is an oversimplified representation of LPB, which is by far a more complex molecular process where the ligand can reach the final binding mode by passing through metastable states (alternative binding modes) and crossing even high energy barriers (Figure 1).

In other words, LPB can be seen as a combination of thermodynamic and kinetic problems, the first defining the ligand binding affinity to its target by estimating the ratio between the ligand concentration in the bound and unbound states and the second characterizing the ligand residence time in the target by computing the rate of ligand (un)binding. The thermodynamics and kinetics of LPB are properties of the

Received: June 14, 2023
Published: September 1, 2023



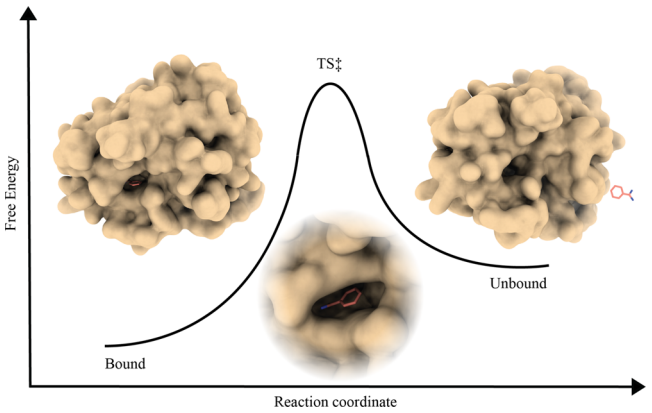


Figure 1. Artistic representation of the free-energy profile of LPB with bound (Bound), transition state (TS) and unbound state (Unbound). Target protein and ligand are shown as the surface and licorice, respectively.

system and are quantified as constants related by the following eq 1:

$$\Delta G_b = -\frac{1}{\beta}\ln(C^0 K_b) = -\frac{1}{\beta}\ln\left(C^0 \frac{k_{on}}{k_{off}}\right) \quad (1)$$

where ΔG is the free-energy difference between the ligand-bound and unbound state, K_b is the binding constant (K_d is the inverse of K_b), β is the inverse of the product of the Boltzmann constant and temperature, C^0 is the standard concentration of 1 M, also expressed as 1/1660 Å^{-3}, and k_{on} and k_{off} are the two kinetics constants for the binding and unbinding processes, respectively.

Noteworthily, recent evidence has correlated ligand binding rates (kinetics) more than ligand binding affinity (thermodynamics) to in vivo drug efficacy.[10−16] In addition, ligand binding kinetics have also been related to the type of ligand activity, agonist or antagonist, letting glimpse the potential impact of including ligand binding kinetics in drug discovery.[17] However, drug discovery studies are still rooted in binding affinity predictive models, and no drug has been developed based on ligand binding rate prediction so far. The reasons for this are both historical and methodological. While plenty of binding free-energy methods have been developed in the last four decades, ligand-binding kinetic models have only appeared in the last two.[18,19] Furthermore, the prediction of ligand binding rates requires the characterization of LPB transition states (TSs), which are high-energy, transient (short-lived) states that are elusive to standard structural biology methodologies. Encouraging results come from time-resolved X-ray crystallography and cryogenic electron microscopy, which provide an augmented spatial and temporal resolution of nonequilibrium macromolecule states.[20,21] Similarly, nuclear magnetic resonance (NMR),[22] surface plasmon resonance,[23,24] and other techniques (see Bernetti et al. 2019 for review)[14] have made significant progress in providing a dynamic description of LPB for kinetic predictions. However, all these techniques are time-demanding and costly, and they hardly provide mechanistic structural details useful to guide drug design.[25] To this end, computational approaches are valuable, since they can sample TSs and obtain transition rates from one energy minimum to another. Having the atomistic structure of TS at the saddle point and the ligand binding mode provides a unique, comprehensive description of LPB that might impact the quality and success of structure−activity relationship (SAR) studies and drug discovery in general. However, so far molecular simulations are used to reproduce experimental kinetic data, with only few examples where calculations are presented together with or followed by experimental validation.[26−29] This is mainly due to two limiting factors that are the time required to complete a kinetic study, in terms of both simulation and real time for analysis, and the accuracy of the calculations that largely rely on system-dependent simulation settings, which hamper a routine and automated use of such techniques in drug discovery. We discuss such aspects in the following paragraphs, describing the state-of-the-art techniques employed in LPB kinetics studies and their limits as predictive tools. We further delineate the future directions of in silico approaches in order to have a significant impact on drug discovery.

2. STATE OF THE ART

Many approaches aimed at studying LPB kinetics have been developed so far. They can be grouped into unbiased and biased MD-based methods.

The first category comprises techniques that focus on sampling the transition between metastable states by massively parallelizing the simulation, allowing direct calculation of the kinetic properties. Milestoning with MD or Brownian Dynamics (BD) simulations,[30−33] its notable variation called Markovian Milestoning with Voronoi Tessellations (MMVT),[34] Weighted Ensemble Methods (WEM and Markovian-WEM),[35−39] and Adaptive Multilevel Splitting (AMS)[40,41] are four examples of this kind of approach. Their key concept is discretizing the configuration space using various descriptors (e.g., grids, distances, native contacts). Multiple simulations are then spawned or killed semi-independently to ensure a statistically meaningful exploration of the transition pathways in a reasonable time. Although these techniques are all based on unbiased approaches, i.e., the single simulations are not affected by external bias, the sampling algorithm may follow a nonequilibrium strategy. In fact, in sampling at equilibrium, the system explores the forward and backward transitions of energy barriers without altering the probability distribution. On the other hand, in nonequilibrium samplig the reconstruction of the equilibrium ensemble is possible by optimizing the number of simulations performed, their distribution, or relative weights, so as to properly sample the rare event and promote the transition in a specific direction. According to this definition, Milestoning and MMVT are equilibrium approaches, whereas WE and AMS are not.

At variance with unbiased simulations, in biased techniques, user-defined degrees of freedom of the system are accelerated by adding an external potential or force, which might be used to compute the correct (unbiased) estimate of the energetic barriers crossed during the simulation. Umbrella Sampling (US),[42] Steered MD (stMD),[43] Targeted MD (TMD) and Dissipation-Corrected Targeted MD (dcTMD),[44−46] Smoothed MD (SMD),[47,48] Adiabatic-bias MD (AbMD),[49,50] Metadynamics MD (MetaD) (i.e., infrequent Metadynamics, frequency-adaptive metadynamics),[51−54] On-the-fly Probability-Enhanced Sampling (OPES, and its flooding variant OPES$_f$),[55−57] Gaussian Accelerated MD (GaMD) and its variants (Pep-GaMD, LiGaMD),[58−60] and τ-random acceleration MD (τRAMD)[61] are notable examples of this family of out-of-equilibrium approaches. Among these, we note that in

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

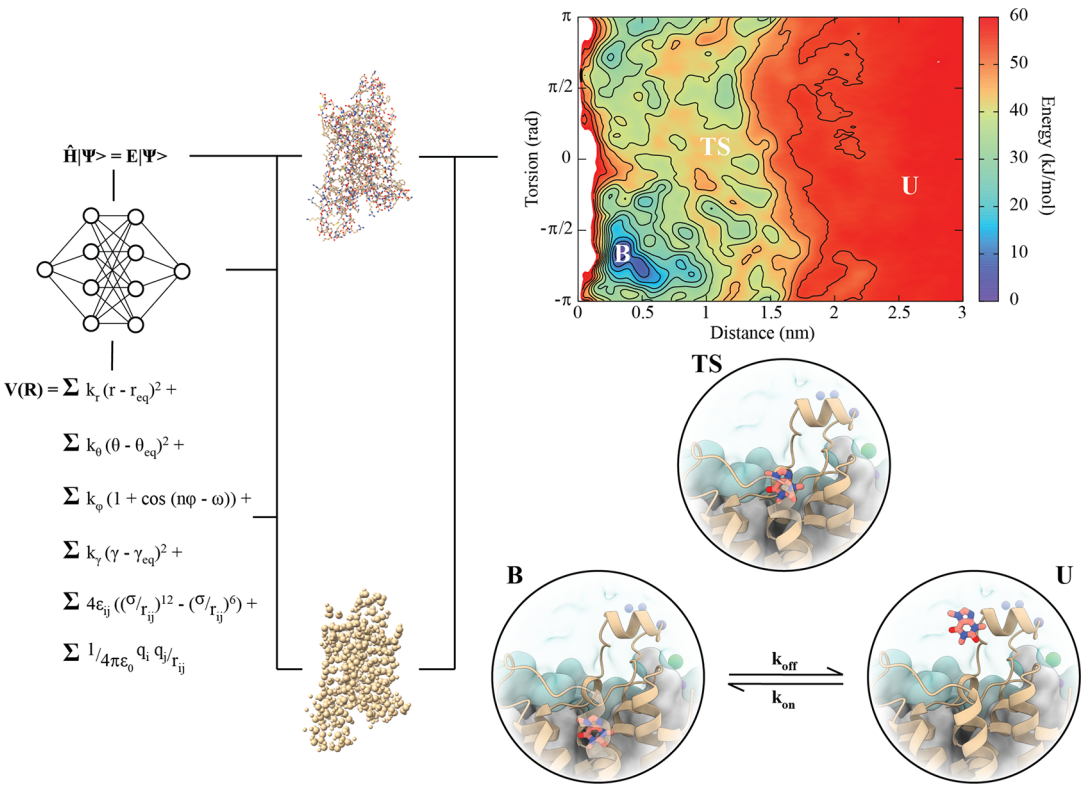


Figure 2. Pipeline of LPB kinetics calculations for a sample system (i.e., adenosine GPCR A2A in complex with caffeine). The LPB can be parametrized using classic or ML-based FFs and then simulated employing one of the computational techniques described in the text. The output is the identification of the possible energy-minima LPB complexes (B and U), the transition state (TS), and rates between them (k_{on} and k_{off}).

infrequent MetaD and OPES$_f$, while the sampling of the basins is accelerated, no bias should be applied to the system during the transition between them. These techniques can be further classified according to how the external potential, or force, is applied to the system and how the kinetic properties are estimated. US, MetaD, AbMD, stMD, TMD, and dcTMD require the definition of specific Collective Variables (CVs), which should describe the slowest degrees of freedom of the process investigated.[42,43,45,46,50,51,62] On the other hand, in GaMD, harmonic boost potentials are directly applied to the potential energy of the system.[60] Similarly, in SMD, the potential energy function is scaled back by a user-defined factor.[47] In τRAMD, randomly oriented forces are applied to the ligand to accelerate its unbinding.[61] Some of these methods compute the kinetic data by estimating the energetic barriers for the binding or unbinding events using Kramers' rate theory or the Eyring equation. This is the case of stMD, dcTMD, and GaMD.[43,44,60] This approach could be also applied to MetaD and OPES. Using the latter techniques, it is also possible to directly estimate the accelerated ligand residence time, which is then rescaled to the correct (unbiased) one using the bias deposited during the simulation.[53,56] The association (binding) rate might be finally derived from the dissociation constant and the ligand residence time.[53] Finally, SMD and τRAMD provide a computational residence time that can be correlated with the available experimental data, whereas abMD allows for estimating an energetic score that can be correlated to the experimental residence time.

In between biased and unbiased computational methods, we find Markov State Models (MSMs).[63−65] MSMs do not have a defined simulation protocol but rather represent an a posteriori analysis strategy that can be applied to any set of biased or unbiased calculations.[53,66−68] The only requirements are that these simulations should be capable of discretizing the phase space into a number of different, energetically relevant states, thus allowing the computation of the transition matrix with the exchange probability among these states.

We refer the reader to ref 19 for a review and deeper discussion of these methods. In the following sections, we focus on the limitations and advantages of these and other computational strategies, providing insights into much-needed improvements required to allow a breakthrough in the field. They can be summarized in three points:

- the force fields issue;
- the sampling issue;
- performance, user-friendliness, and data openness.

3. FORCE FIELDS

Force Fields (FFs) are a set of parameters through which interatomic forces are computed, allowing the system to evolve during an atomistic simulation, be it MD or Monte Carlo. As such, FFs should describe the physics underlying the simulated phenomenon, and their accuracy is crucial for the predictive power of simulations (Figure 2).

Over the years, continuous efforts have been paid in the development of reliable FFs for biological macromolecules (e.g., Amber, CHARMM, OPLS)[69−74] and small organic molecules (e.g., GAFF, CGenFF, OPLS, OpenFF, SMIRNOFF99Frosst).[69,73−77] Typically, atomistic FFs employ classical physics equations to describe interatomic interactions with fixed, point-charge models, defined without considering the surrounding chemical environment. The parameters of

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

these equations are derived from experimental or quantum mechanical (QM) calculations data, often using models with a finite number of atoms. These are then adapted into more generalized terms to serve a wider variety of chemical entities.[69−74] Despite the intrinsic limitations of these models justified by historical and practical reasons (FFs were developed using significantly lower computational capabilities compared to today), they have served very well the purpose of advancing our understanding of biophysical processes. But are classical FFs accurate enough for modern research in the age of GPU accelerators and machine learning (ML) techniques? The answer is highly dependent on the problem of interest. We can start by saying that in the case of LPB kinetics traditional FFs are not accurate enough.

An accurate estimate of kinetic rates requires a proper identification of TSs.[10] However, subpar parameters might overly stabilize specific ligand (and protein) conformations, thus altering the identification and energetic evaluation of metastable and transient states.[78,79] It has been shown that inadequacies in the van der Waals and electrostatic terms of FFs may influence partition functions,[80,81] and also impact the osmotic coefficients of some chemical entities.[82] Using the paradigmatic system benzamidine/trypsin, we have recently shown that proper parametrization of the ligand torsional potentials is essential to reconstruct the correct ligand binding conformation and free-energy surface.[78] In particular, the default benzamidine parameters generated by different FFs underestimated the energy of the barrier characterizing the rotation of the amidine group with respect to the aromatic ring. While the numerical value of the ligand/protein binding free energy was not affected by this issue, differences might be found in the energetic profile and binding mechanism obtained using the tailored or default torsional potential. A second important player is the mathematical formulation of molecular charges. Most FFs consider the electrostatic distribution fixed, even though it depends on the phase state, conformation, and the environment surrounding the molecule.[83] The electron densities of a ligand in the unbound and bound state may also be significantly different.[84] Although these approximations have a limited impact—in the absolute estimate of electrostatics—on the results quality of MD calculations when taken individually, they introduce errors that can reach tens of kJ/mol, as shown by Kaminsky and Jensen.[85] Even if such inaccuracies do not affect the detection of the metastable states, the passage through these states, which defines kinetic rates, could be altered. In this regard, Vitalini et al. reported that different FFs might provide contrasting kinetic rates and dynamics for the same system, even when the identified metastable states are in general agreement.[86] The simplest solution to these issues could be a thorough refinement of the existing FFs to improve the bonded and nonbonded terms. Still, it would not solve the fundamental limit of the classical physics description, especially regarding the molecular electrostatic properties. Instead of lingering on such approaches, we feel that it is time for FFs to step up to a higher tier of modelization, leveraging the computational power of current hardware and more accurate levels of theory. In this sense, the strategy followed by polarizable FFs is particularly promising.

Polarisable FFs adopt a different philosophy when treating molecular electronic densities, allowing the electron cloud to move around the atoms to induce dipole effects and mimic anisotropic electrostatic distributions, such as σ holes. This approach results in a more accurate depiction of molecular charges at the cost of increased computational power.[87] Several polarizable FFs are available nowadays (e.g., Amber ff02pol, CHARMM Drude, SIBFA, AMOEBA)[88−91] and they have consistently demonstrated higher capabilities than standard FFs in reproducing thermodynamic data.[86,92,93] Particularly noteworthy are the results of the SAMPL8 challenge achieved by AMOEBA.[94] Sadly, these FFs are rarely used to investigate LPB processes, especially due to their high computational cost. They usually find application in studying ion-absorbing, host/guest, or phase transition processes, DNA systems, and protein stability, phenomena where electrostatic interactions play a significant role.[87,95−97] This is also the case for LPB, so we expect that polarizable FFs will soon become standard for these studies. While polarizable FFs can achieve better results than classic ones, they are still not accurate enough to consistently reproduce kinetic data, as shown by Kaminsky and Jensen.[85] A further optimization step in FFs development might be the inclusion of kinetic data sets in the FF parameters generation. This proposal is nothing new, seeing as it was already suggested in 2015 by Vitalini et al.[86] Doing so would, however, require setting up novel parametrization schemes to consider kinetic and thermodynamic measurements. A pragmatic solution to this hurdle could be resorting to ML techniques.

ML has already been employed to optimize existing FFs, like the Lennard-Jones (LJ) parameters for Drude oscillators,[98,99] or to develop new ones.[100,101] In particular, LJ parameters are typically computed by taking the experimental hydration-free energy of a compound as a reference. However, their optimization is a lengthy and daunting task requiring extensive validation with multiple experimental data sets. Such an approach is necessary to ensure the reliability and accuracy of a FF. In Chatterjee et al. and Rupakheti et al., a ML approach has been employed to optimize the LJ parameters for several classes of compounds, improving the estimation of different molecular properties, including molecular volume, vaporization and sublimation enthalpies, and dielectric constant.[98,99] Specifically, Rupakheti et al. reported an improvement in the estimate of molecular volumes and heat of vaporization of the druglike small molecules of 2% and 9%, respectively. Similar improvements were also found in compounds not included in the training set. The average error with respect to the experimental data was 0.46 kcal/mol, which is significantly lower than the average error of 2.0 kcal/mol reported for GAFF using the same compounds.[98] As previously mentioned, ML is also employed to parametrize new FF from scratch.[100,101] ML FFs do not employ physics-based equations to describe atom−atom interaction and can be trained using even complex data sets, such as ab initio quantum-mechanical data. The main advantages are that the ML FFs might not require an a priori definition of how the atoms are bonded since the chemical bonds are directly inferred from the pairwise atomic distances, and quantum-mechanics effects can be embedded in the model at significantly reduced computational costs. Several strategies have been implemented to develop ML FFs, with diverse pros and cons and different accuracies. For a more detailed discussion on ML FFs, we refer the reader to the review of Unke et al.[101] The most notable example of ML FF, as well as one of the most promising, is NequIP from Batzner et al.[102] Fu et al.[103] have recently benchmarked it together with other state-of-the-art ML FFs, such as DeepPot-SE,[104] SchNet,[105] ForceNet,[106] and GemNet-T,[107] using different molecules and observables. In detail, Fu et

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

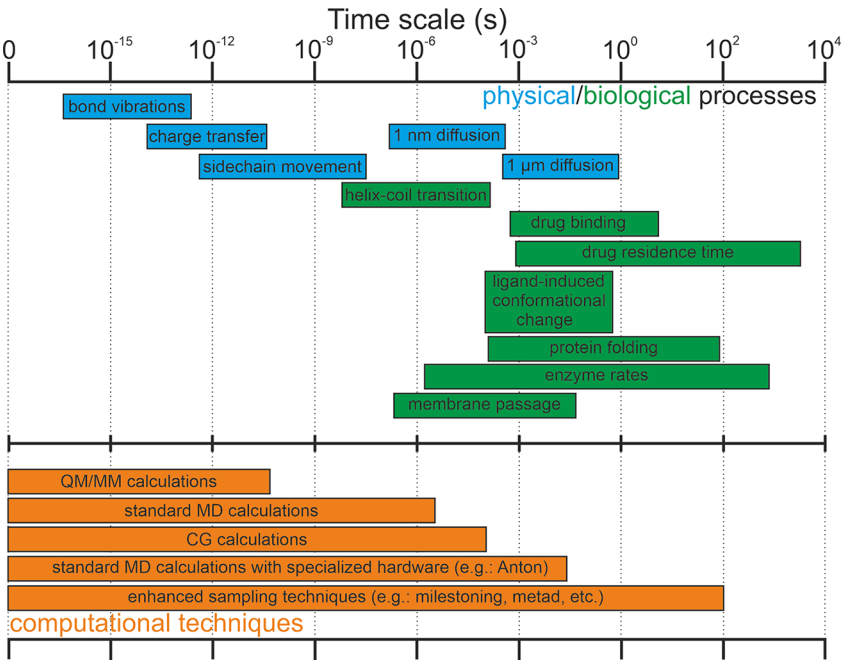


Figure 3. Time scales of the physical and biological processes investigated by simulations. The bars related to the diffusion refer to diffusion coefficients measured in water. At the bottom time scales achievable by computational techniques.

al. assessed the ML FFs capability to predict the correct forces, distribution of interatomic distances, and sampling of the free-energy surface of organic compounds (ethanol, naphthalene, and salicylic acid), a small molecule (aspirin), and a peptide (alanine dipeptide). They also evaluated whether these ML FFs could reproduce the correct radial distribution functions and diffusion coefficients of condensed phase systems such as liquid water and crystalline superionic lithium. Finally, the numerical stability of each system during the MD calculations was assessed. Overall, NequIP performed consistently better than the other ML FFs, with the exception of the simulation speed (i.e., frames per second). This is due to its construction, since NequIP uses local descriptors that require a higher amount of data to be processed. Notwithstanding, NequIP, as all the other ML FFs, has difficulties reproducing the sampling behavior of standard simulations in systems as simple as alanine dipeptide. In fact, when the simulation of the alanine dipeptide is started from low-density metastable states, NequIP is unable to reproduce the correct probability distribution, resulting in an incorrect free-energy surface. In some cases, simulations starting from these states could also be affected by numerical instabilities and crashes. The authors hypothesize that such issues are due to the quality of data set used to train the model, which has poor statistics of low-density states and lacks high-energy conformations. They remark that the latter could be particularly useful since it could improve the models' reliability.[103] Interestingly, ML techniques have also been applied to generalize atomistic FFs at the Coarse-Grained (CG) level with the final aim of reducing the dimensionality of the molecular representation and capturing the most relevant (slow) degrees of freedom of the simulated system.[100,108] CG models achieve this scope by merging the atomic particles into single entities called "beads" whose energetic interaction is computed through equations inspired by atomistic FFs or classical physics. The final result is reproducing the macroscopic properties of systems that would have required an unfeasible amount of time and resources using all-atom representation, however maintaining a physically sound description of the molecular entities involved.[109−111] CG FFs have been successfully employed to study mesoscale processes, supramolecular assembly, protein folding, and protein/lipid interactions.[110,112] In some cases, they were also used to reconstruct kinetic information related to protein−protein association mechanisms,[113−115] but for a long time, they were not employed to investigate small molecules and, in particular, LPB.[112,116] The recent works of Dandekar et al.,[117] Negami et al.,[118] and Souza et al.[119] have dramatically changed the status quo of CG FFs. Using different versions of the popular Martini FF, these authors demonstrated its potential in investigating the binding of small molecules to various classes of targets. In particular, Souza et al. employed the recently released third version of Martini FF,[111,119] which addressed several limitations of the previous ones.[120−122] Dandekar et al. and Negami et al. were instead able to reconstruct the LPB kinetics with good approximation.[117,118] Doing so, however, requires estimating the acceleration factor introduced by coarse-graining, which is known to reduce molecular friction and lower energy barriers due to entropy loss.[116] This is rarely done in practice. Usually, a generic, empirical acceleration factor from three to eight is provided as a correction to the kinetic estimates. These values are speed-up factors roughly estimated by comparing the diffusion coefficients calculated in pure water systems using Martini and atomistic simulations, and not considering the system under investigation.[117,118] Despite being useful for mesoscale applications, applying CG FFs to LPB problems suffers from limitations mainly correlated to their coarse nature that should be carefully taken into account. In particular, the lower accuracy of CG modeling implies that many details in the ligand/protein interactions may be approximated. Furthermore, the lack of a bijective function between AA and CG models leads to the obvious consequence that different AA models could be mapped to the same CG structure. In all cases, we recommend dutifully validating the kinetic rates obtained via CG MD calculations

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

against experimental data to adequately estimate the acceleration factor introduced by the CG FFs. These aspects and other limitations of classic CG FFs, such as their thermodynamics-based design and configurational-based CG mapping, could be overcome by ML techniques.[100,108,123−125] While we expect that these FFs, and the Martini in particular, will continue to play a leading role in investigating many phenomena, they could be integrated or replaced by ML-guided CG FFs in the future. As discussed for the atomistic FFs, the potentiality of ML methods in parametrizing, optimizing, and including additional measurables, such as kinetic data sets, could significantly improve the accuracy of FFs and, therefore, cannot be overlooked. CG-ML FFs would also benefit from the more consistent, data-driven approach in deciding how to map atoms into single beads, as opposed to current strategies that only consider the configurational aspect.[124] In doing so, they could include protein conformational freedom, which plays a determining role in ligand binding processes ruled by induced fit or conformational selection and is currently neglected by CG models.

In conclusion, the current findings indicate that existing atomistic and CG FFs are not accurate enough to reconstruct LPB kinetics properly. For this aim, a higher level of theory and a significant change in the parametrization strategies are required. ML FFs hold promise, but regrettably, they are not yet fully developed for general applications. They are facing essential challenges related to instabilities, representation issues, and limitations in transferability, particularly when applied to larger systems.[103,124,126] However, their rapid development cycle is encouraging, setting them apart from traditional FFs, with numerous new ML FFs being introduced annually.[100,101,103,108,123,125] For instance, Fu et al. have provided valuable insights into the limitations of current training methods for ML FFs, which are likely to inspire further advancements in the field.[103] The atomistic ML FFs hold great promise in achieving quantum mechanics-level accuracy in MD simulations, with the potential to introduce greater flexibility in handling configurational and conformational variations in small molecules, peptides, and proteins. This could include simulating tautomeric shifts or changes in protonation states, which are typically not feasible in MD calculations without dedicated algorithms. However, it is realistic to expect another five-ten years of progress to be necessary for a significant breakthrough. A similar time frame was observed in neural networks for visual pattern recognition, which were first developed in 1988 but only surpassed human accuracy thresholds in 2015.[127]

4. SAMPLING

The residence time of a drug in its molecular target can range from milliseconds to hours (Figure 3); therefore, LPB kinetics simulations should reach a comparable time scale.

In addition, to quantitatively estimate the kinetic rates, a relatively large number of transitions between energy minima should be observed to collect meaningful statistics. This makes this kind of study time and computationally demanding.[128,129] From a theoretical point of view, the study might focus on simulating either the binding or the unbinding, with the estimate of k_{on} or k_{off}, respectively. Once one of the two constants has been computed, the other could be derived according to eq 1 and supposed that K_b is known.[14] It has to be said that binding rates are faster than unbinding ones since the unbound state is higher in energy than the bound state, and consequently, the barriers to cross from the unbound state are lower in energy if compared to the same from the bound state. Therefore, the simulation of ligand binding is less time demanding with respect to unbinding. On the other hand, the system in the unbound state has a larger entropy, with multiple isoenergetic conformations possible and only one or a few of them functional for the ligand to reach the final binding mode. This ends in a probabilistic problem that requires a significant number of simulations for sampling the correct ligand binding. At variance with ligand binding, ligand unbinding is characterized by higher energy barriers, but only a few paths can reach the unbound state. This makes the ligand unbinding process and estimate of k_{off} computationally less challenging. However, in such simulations, the ligand binding mode should be a priori known. In both cases, kinetic calculations are demanding, and despite the increasing computing power, such extensive simulations are not accessible to many groups. Therefore, the practical solution is to decompose the problem into shorter and simpler tasks or to enhance the sampling.[130−135] As previously introduced, to estimate kinetic rates and simulate the entire (un)binding process, the transition paths between all of the energetic minimum states have to be sampled. This rules out computational techniques such as end-point and alchemical transformation methods, which focus on calculating the binding thermodynamics based on ligand binding modes, while neglecting the ligand binding mechanism dynamics.[136,137]

The simplest solution to deal with LPB kinetics is to run brute force MD calculations, organized in a number of parallel, independent replicas, followed by an algorithm that computes rates to go from one state to another one.[13] However, this approach simply shortens the time required to acquire the results without reducing the computational cost. As a result, slow kinetic rates, beyond milliseconds, remain unaccessible to this kind of approach. An elegant solution comes from the concept of a collective variable (CV): a reaction coordinate that distinguishes the diverse states assumed by the system during the binding process. CVs can be either easy and intuitive to define (e.g., geometrical features such as distances between molecules or groups of atoms) or complex in nature (e.g., functions of molecular properties), and they should represent the slowest degrees of freedom of the system under investigation. At the end of the simulation, the potential of mean force (PMF) computed as a CV function describes the evolution of the free energy along the LPB process, allowing the identification of transition states (Figure 1). In real-case scenarios, LPB might be characterized by a relatively large number of degrees of freedom with different dimensions (interatomic distances, angles, coordination shells, etc.), which could be difficult to identify. In addition, ligand binding might be characterized by different phases ruled by diverse slow degrees of freedom, especially in cryptic binding pockets and long kinetics during which the system might significantly evolve through large protein conformational changes, solvation, and other long time scale effects. Considering the complexity of the problem, it is clear that CV definition is a dimensionality reduction problem and user-defined CVs might not be accurate enough.[138−140] Calculations based on a bad CV choice might lead to simulation issues, including hysteresis and a lack of convergence. In the era of machine learning and semiautomated algorithms, many ways to reduce the dimensionality of the problem have been proposed, including principal component analysis, time-lagged independent

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

component analysis, and various ML approaches.[138,141−148] These techniques work with a given simulation data set and combine the slowest degrees of freedom of the system using linear or nonlinear functions. In this theoretical framework, the quality of the data set and the time scale of the preparatory simulations heavily affect the results. So far, ML-driven CV definition has been reported on a number of systems and more applications in general cases—where slow degrees of freedom are unknown—are necessary to assess better their predictive power.[13,15,149−152]

Furthermore, it is important to note that the system's slow degrees of freedom might not necessarily correspond to those functional for LPB procecess. This makes CV selection even more difficult. A possible solution could be to identify the degrees of freedom—and the CVs—functional for LPB (e.g., specific ligand/protein interactions, conformational changes, etc.) by training a ML model with short simulations at relevant metastable states of the system.[153] A distinct discussion merits the path-based methods, used to analyze the transition between two states by determining a transition-path ensemble constructed from an initial guess.[154−157] The initial guess pass is enriched by new branches, created, and added to the ensemble by running MD simulations or MC calculations. Despite being elegant in its formulation, this approach can be quite costly for systems with high barriers and rough free-energy profiles, for which possible solutions have been proposed.[158] To reduce the computational effort, the reaction pathway can be fragmented into discrete states, and the transitions between them can be simulated independently in parallel. This strategy is applied in state-based methods encompassing string methods, Markov state-based models, weighted ensemble, and milestoning, in which the CV can distinguish specific conformations or anchors in Voronoi tesselation.[32,37,38,159−161] Simulation protocols based on such approaches have been purposely designed for biological systems to consider the presence of multiple metastable states and the typically complex LPB free-energy surface. An interesting, more recent approach is reconstructing the correct sampling probability by operating on enhanced-sampling trajectories with nonoptimized CVs (e.g., variational conformational dynamics, OPES, and similar techniques).[55,162−166] For example, a customized version of the OPES technique has been applied in the calculation of kinetic rates for the benzamidine/trypsin system,[56] disclosing the role of water after the definition of a tailored solvation CV by the authors.[57] Such methods have demonstrated the ability to correct a skewed sampling if the proper degrees of freedom are included; however, as for other similar applications, their predictive power has to be assessed in real cases where little information is known about the system's slow degrees of freedom.

Whatever method one decides to use to study LPB kinetics and identify TSs, a postprocessing validation step is recommended and good practice. The most intuitive and old procedure implies sampling along nonintersecting hypersurfaces connecting two energy minima A and B identified in previous simulations and computing the committor probability. This is the probability that a given system conformation, belonging to the transition path connecting two minima, ends into A or B.[167−169] The committor ranges from 0 to 1 and is 0.5 at the TS, which means that running a relatively large number of independent simulations starting from the putative TS, half should fall into A and half into B. It could also be used to identify TSs in simple systems, e.g., with only two energy minima, which is seldom the case in LPB. In addition, statistical tests like the Kolmogorov−Smirnov test might further be used to assess the Markovian transition of the saddle points and the Poissonian distribution of the computed kinetic rates.[13,53,170,171]

Furthermore, committors might be employed to identify optimal reaction coordinates and meaningful CVs, even using ML approaches.[172−174] This is an active field of research that could be further expanded in the near future.[175] In conclusion, defining CVs that could at least alleviate the sampling issue is a most reasonable strategy and a daunting task that continues to attract the community.[138,176] Despite the encouraging results coming from automated and semiautomated ML algorithms, the extensive preparatory simulations required for this kind of calculations limit their applicability.[177] An interesting direction could be to define the best CVs automatically on-the-fly during simulation using an unsupervised algorithm as suggested by Bhakat.[177] Independently from the method chosen, the operator is asked to possess a solid understanding of the technique and the investigated system, which dramatically impacts the method's user-friendliness, as discussed in the next section.

5. FRIENDLINESS AND PERFORMANCE

In a world more focused on energy efficiency[178,179] and data openness,[180,181] aspects such as computing performances, user-friendliness, and code accessibility or reuse are becoming increasingly important. This section briefly discusses the most popular methods for computing the LPB kinetics from this perspective. First, we compare the simulation times required by diverse techniques to compute the association or dissociation rates and the accuracy of their estimates. Then, we discuss code availability, ease of usage, and data openness. We conclude this section by analyzing how the open data approach can be applied to FFs.

Providing a fair comparison of the computational performance of the different techniques is difficult. Nonetheless, some baselines can still be drawn using a common test case like the paradigmatic system benzamidine/trypsin where ligand unbinding occurs with a millisecond time scale.[32,34,39,46,53,57,60,66−68,161,163,182−186]

Table 1 summarizes the results achieved over the years using different methodologies, organized chronologically. Please note that in Table 1, for the sake of discussion, we use the names of the general techniques as defined in the section 2. From the reported data, it can be seen that optimization of the computing protocols, i.e., reduction of computer time needed for LPB simulation, has been a primary focus. From 2016 onward, starting with the work based on the AMS method,[184] new techniques typically required less than ten microseconds of calculations to provide an estimation of the association and dissociation rates, except for CG MD.[117] Notable results were achieved by Markovian-WEM (M-WEM)[39] and OPES$_f$[57] in delivering accurate values of K_{off} in less than four microseconds. Similarly, MMVT[34,161] gave an excellent estimate of the K_{on} in only five microseconds. The kinetic estimations significantly differed from the experimental ones in all of the other cases. However, one should reckon that such inaccuracy might be due to FFs issues, as discussed in the previous section, and not necessarily to the methods themselves. Considering the continuous improvement in the hardware field[133,187] and the potentialities of the upcoming quantum computing technology,[188] it is likely that in the future, the real

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

Table 1. Comparison of Methods Employed for Computing the LPB Kinetics in the Benzamidine/Trypsin System

Method	K_{on} (10^6 M^{-1} s^{-1})	K_{off} (s^{-1})	Simulation Time[a] (μs)
Experimental[182]	29	600 ± 300	–
MD + MSM[66]	150 ± 20	95000 ± 33000	50
MD + MSM[67]	440	28000	1
MetaD[53]	11.8 ± 10	9.1 ± 2.5	2
MD + MSM[183]	64	13100 ± 10900	149.1
US + MD + MSM[68]	–[b]	1170 ± 276.5	58.28
AMS[184]	–[b]	260 ± 240	2.3
MD + BD[32]	21 ± 3	83 ± 14	19
WEM[185]	–[b]	5556	4.1
MetaD[163]	–[b]	4176 ± 324	1.2
WEM[186]	–[b]	266	8.75
dcTMD[46]	8.7 ± 0.5	270 ± 40	10000[c]
CG MD[117]	368	690000	398
MMVT[34]	120 ± 5	174 ± 9	4.4
LiGaMD[60]	11.5 ± 7.9	3.53 ± 1.41	5
M-WEM[39]	7.6 ± 3.8	769 ± 261	0.73
MMVT[161]	24 ± 2	990 ± 130	5
$OPES_f$[57]	–[b]	687[d]	3.3

[a]Only production runs are considered. [b]Data not reported/computed. [c]dcTMD employs coarse-graining of the degrees of freedom of the system and increased integration time step (up to 10 fs), significantly reducing the computational time and power needed for simulations[46] [d]This value is the slowest unbinding rate identified in 60% of the performed simulations. No estimate of the standard deviation is reported in the paper.[57]

time needed to achieve the calculation convergence will be significantly shorter. With this in mind, achieving a more accurate, comprehensive description of the interatomic forces (as discussed in section 3) at the cost of more demanding calculations seems to be acceptable or even desirable.

Access to higher computational capabilities or more accurate modeling of molecular properties will be a game changer, but we should not focus on only these aspects. We should also continue to enforce the recent push for good code availability and reusability practices to make these techniques easier to use and comprehend, fostering their diffusion in the scientific community. In principle, computational methods should have clear documentation, self-explaining interfaces, maintained repositories with a tracked list of changes and, when possible, open databases of protocols for reproducibility practices.[180] For example, MetaD and MMVT algorithms are supported by their specific libraries PLUMED[189] and SEEKER,[34,161] with plenty of tutorials and documentation available. Funnel-metadynamics, a variant of metadynamics designed to study LPB, has been recently furnished with a user-friendly protocol, named Funnel-Metadynamics Advanced Protocol (FMAP), and a graphical user interface.[190] Furthermore, the PLUMED consortium has published PLUMED-NEST, a public repository containing the computational protocols employed in PLUMED-assisted MD calculations, for the sake of data reproducibility.[191] LiGaMD, on the other hand, has been directly integrated into the Amber software.[192] We believe that the next step in making this software more accessible and easy to use is integrating it into a web server to generate and validate simulation inputs and create public databases for storing trajectories related to kinetic experiments. Examples of such an approach can already be seen in the CHARMM-GUI Web server,[193] the 3-dimensional structure Representation Sharing (3dRS),[194] and the GPCRmd repository,[195] among others.[196−198]

Lastly, we believe existing and new FFs should be developed following the principles of data openness. Most FFs use diverse definitions for residue names, atom names, or types, which may confuse a novice user. In addition, they have different parametrization routines, which often involve diverse pools of model compounds.[69,73−76,98] Such discrepancies affect the bonded and nonbonded parameters obtained from the parametrization tasks, leading to divergent behaviors in simulation, especially for small molecules, when diverse FFs are applied to the same problem. Multiple authors have observed such outcomes in different circumstances, in both thermodynamic and kinetic calculations. The SAMPL6 challenge highlighted that GAFF and OPLS systematically overestimated the octanol/water partition coefficients, whereas CGenFF gave more accurate predictions.[81] Kashefolgheta et al. reported slightly different behaviors of the most popular ligand FFs when evaluating their capability to reproduce experimental cross-solvation free energies.[80] Zhu showed similar results in its benchmark calculations of ligands FFs when compared to experimental osmotic coefficient data.[82] Amore et al. demonstrated that different FFs provide varying conformer energies and geometries during the optimization of multiple molecules and fragments.[79] We documented that the parametrization of benzamidine using default GAFF parameters led to unsatisfactory outcomes, altering the ligand binding conformation and the free-energy landscape of the benzamidine/trypsin binding process.[78] Kaminsky and Jensen reported that the conformational transitions of various amino acid derivatives varied depending on the FF employed. On average, the errors in the interconversion barriers were in the order of 10 kJ/mol.[85] While such discrepancies are usually minor and may not be an issue in most cases, they could significantly influence the outcome of LPB kinetic studies, especially when working with poorly parametrized functional groups and atom types,[98,199] as also shown by the comparison between different versions of OPLS.[74,200,201] Developing standardized FFs with improved physicochemical description, uniform parametrization protocols, unified validation tests, and reproducible results over a wide array of functional groups would be advantageous for accurately predicting kinetic data and MD calculations.[75] Creating universal definitions of atom types and names, shared among all FFs, would also help with such standardization efforts. In this context, abandoning the historical classification of atomic entities into atom types, which unnecessarily complicates present FFs due to redundancy issues, in favor of alternative approaches such as the one presented by Mobley et al. could support the development of a gold standard.[202] A first step in this direction could be the development of public repositories of model compounds with theoretical and experimental data for the parametrization of FFs to avoid discrepancies in the reference data pools. To date, the Open Force Field initiative is the only consortium to host the complete data set employed for the parametrization of its FF in an openly accessible form.[75,76]

6. CONCLUSIONS

LPB kinetic calculations hold great promise in drug discovery to achieve more accurate prediction models of in vivo drug activity.[10−16] In the past decade, computational methods have demonstrated the ability to characterize LPB ki-

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

netics.[32,34,39,46,53,57,60,66−68,161,163,182−186] However, the applicability and accuracy of such approaches remain limited (Table 1), mainly due to issues related to FFs and sampling capability. The hardware and software improvements—also considering the emerging era of quantum computing—together with the continuously evolving ML techniques will undoubtedly play a leading role in the coming years, hopefully making this kind of calculations less demanding and useful for ligand database screening protocols. In this regard, one could expect that even quantum mechanical calculations—now considered unfeasible for ligand binding studies[203]—might at least integrate the atomistic level description of LPB. In the present article, we have discussed three macro areas that could be further improved, which are molecular properties parametrization, sampling, and data openness. In addition, to make kinetics calculations the "gold standard" in the near future, the following requisites should also be fulfilled: (i) prediction of kinetic rates; (ii) assessment of the results; (iii) release of the atomistic structure of rate-determining states. To date, kinetics calculations have been employed to reproduce experimental rates, with only a few examples where calculations are presented together with or followed by experimental validation.[26−29] We expect that a posteriori experimental validation of the simulation data can become a standard practice. This would represent an important acknowledgment of the predictive power of this kind of calculations. Furthermore, the assessment of the results is another crucial point. As introduced in section 4, committor analysis and statistical postprocessing tests are available and should represent a standard practice to assess kinetics constants estimates and identification of TSs. The latter leads to the last point, which is the release of the atomistic structure of the rate-determining states. These are high-energy, short-lived, hence transient, states that are elusive to experimental structural biology techniques like crystallography, Cryo-EM, and NMR. At variance with these techniques, atomistic simulations have the capability of detecting and disclosing the atomistic structure of TSs and these should be reported in any kinetic study. These structures would indeed increase the impact of the work and represent unprecedented structural information helpful for medicinal chemists in the design of ligands with tailored binding kinetic properties. In such a way, the LPB kinetics models could mark a breakthrough in drug discovery and be in the pool of the in silico models as nonanimal alternative methods for drug assessments approved by the FDA (Food and Drug Administration Modernization Act 2.0).[9]

AUTHOR INFORMATION

Corresponding Author

Vittorio Limongelli − *Faculty of Biomedical Sciences, Euler Institute, Universitá della Svizzera italiana (USI), 6900 Lugano, Switzerland; Department of Pharmacy, University of Naples "Federico II", 80131 Naples, Italy;* orcid.org/0000-0002-4861-1199; Phone: +41 58 666 4293; Email: vittoriolimongelli@gmail.com

Authors

Paolo Conflitti − *Faculty of Biomedical Sciences, Euler Institute, Universitá della Svizzera italiana (USI), 6900 Lugano, Switzerland;* orcid.org/0000-0002-5440-6167

Stefano Raniolo − *Faculty of Biomedical Sciences, Euler Institute, Universitá della Svizzera italiana (USI), 6900 Lugano, Switzerland;* orcid.org/0000-0003-4835-8696

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.jctc.3c00641

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by a grant from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme ("CoMMBi" ERC grant agreement No. 101001784) and the Swiss National Supercomputing Centre (CSCS) under project ID s1116.

REFERENCES

(1) Hay, M.; Thomas, D. W.; Craighead, J. L.; Economides, C.; Rosenthal, J. Clinical development success rates for investigational drugs. *Nat. Biotechnol.* **2014**, *32*, 40−51.

(2) Sun, D.; Gao, W.; Hu, H.; Zhou, S. Why 90% of clinical drug development fails and how to improve it? *Acta Pharm. Sin. B* **2022**, *12*, 3049−3062.

(3) Goldberg, N. H.; Schneeweiss, S.; Kowal, M. K.; Gagne, J. J. Availability of Comparative Efficacy Data at the Time of Drug Approval in the United States. *JAMA* **2011**, *305*, 1786−1789.

(4) Mullard, A. New drugs cost US$2.6 billion to develop. *Nat. Rev. Drug Discovery* **2014**, *13*, 877.

(5) Wong, C. H.; Siah, K. W.; Lo, A. W. Estimation of clinical trial success rates and related parameters. *Biostatistics* **2019**, *20*, 273−286.

(6) Zhu, T.; Cao, S.; Su, P.; Patel, R.; Shah, D.; Chokshi, H. B.; Szukala, R.; Johnson, M. E.; Hevener, K. E. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. *J. Med. Chem.* **2013**, *56*, 6560−6572.

(7) DiMasi, J. A.; Reichert, J. M.; Feldman, L.; Malins, A. Clinical Approval Success Rates for Investigational Cancer Drugs. *Clin. Pharmacol. Ther.* **2013**, *94*, 329−335.

(8) Slater, O.; Kontoyianni, M. The compromise of virtual screening and its impact on drug discovery. *Expert Opin. Drug Discovery* **2019**, *14*, 619−637.

(9) Moutinho, S. Researchers and regulators plan for a future without lab animals. *Nat. Med.* **2023** DOI: 10.1038/s41591-023-02362-z.

(10) Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discovery* **2006**, *5*, 730−739.

(11) Lu, H.; Tonge, P. J. Drug-Target Residence Time: Critical Information for Lead Optimization. *Curr. Opin. Chem. Biol.* **2010**, *14*, 467−474.

(12) Copeland, R. A. The drug-target residence time model: a 10-year retrospective. *Nat. Rev. Drug Discovery* **2016**, *15*, 87−95.

(13) Dickson, A.; Tiwary, P.; Vashisth, H. Kinetics of Ligand Binding Through Advanced Computational Approaches: A Review. *Curr. Top. Med. Chem.* **2017**, *17*, 2626−2641.

(14) Bernetti, M.; Masetti, M.; Rocchia, W.; Cavalli, A. Kinetics of Drug Binding and Residence Time. *Annu. Rev. Phys. Chem.* **2019**, *70*, 143−171.

(15) Nunes-Alves, A.; Kokh, D. B.; Wade, R. C. Recent progress in molecular simulation methods for drug binding kinetics. *Curr. Opin. Struct. Biol.* **2020**, *64*, 126−133.

(16) van der Velden, W. J. C.; Heitman, L. H.; Rosenkilde, M. M. Perspective: Implications of Ligand-Receptor Binding Kinetics for Therapeutic Targeting of G Protein-Coupled Receptors. *ACS Pharmacol. Transl. Sci.* **2020**, *3*, 179−189.

(17) De Benedetti, P. G.; Fanelli, F. Computational modeling approaches to quantitative structure-binding kinetics relationship in drug discovery. *Drug Discovery Today* **2018**, *23*, 1396−1406.

(18) Bruce, N. J.; Ganotra, G. K.; Kokh, D. B.; Sadiq, S. K.; Wade, R. C. New approaches for computing ligand-receptor binding kinetics. *Curr. Opin. Struct. Biol.* **2018**, *49*, 1−10.
(19) Limongelli, V. Ligand binding free energy and kinetics calculation in 2020. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2020**, *10*, No. e1455.
(20) Papasergi-Scott, M. M.; Pérez-Hernández, G.; Batebi, H.; Gao, Y.; Eskici, G.; Seven, A. B.; Panova, O.; Hilger, D.; Casiraghi, M.; He, F.; Maul, L.; Gmeiner, P.; Kobilka, B. K.; Hildebrand, P. W.; Skiniotis, G. Time-resolved cryo-EM of G protein activation by a GPCR. 2023; DOI: 10.1101/2023.03.20.533387. bioRxiv ePrint archive. https://www.biorxiv.org/content/early/2023/03/21/2023.03.20.533387.1 (accessed March 23, 2023).
(21) Schmidt, M. Time-Resolved Macromolecular Crystallography at Modern X-Ray Sources. *Methods Mol. Biol.* **2017**, *1607*, 273−294.
(22) Lindorff-Larsen, K.; Best, R. B.; Depristo, M. A.; Dobson, C. M.; Vendruscolo, M. Simultaneous determination of protein structure and dynamics. *Nature* **2005**, *433*, 128−132.
(23) Zhang, P.; Ma, G.; Dong, W.; Wan, Z.; Wang, S.; Tao, N. Plasmonic scattering imaging of single proteins and binding kinetics. *Nat. Methods.* **2020**, *17*, 1010−1017.
(24) O'Connell, N. Protein Ligand Interactions Using Surface Plasmon Resonance. *Methods Mol. Biol.* **2021**, *2365*, 3−20.
(25) De Vivo, M.; Masetti, M.; Bottegoni, G.; Cavalli, A. Role of Molecular Dynamics and Related Methods in Drug Discovery. *J. Med. Chem.* **2016**, *59*, 4035−4061.
(26) Bai, F.; Xu, Y.; Chen, J.; Liu, Q.; Gu, J.; Wang, X.; Ma, J.; Li, H.; Onuchic, J. N.; Jiang, H. Free energy landscape for the binding process of Huperzine A to acetylcholinesterase. *Proc. Natl. Acad. Sci. U.S.A.* **2013**, *110*, 4273−4278.
(27) Bernetti, M.; Rosini, E.; Mollica, L.; Masetti, M.; Pollegioni, L.; Recanatini, M.; Cavalli, A. Binding Residence Time through Scaled Molecular Dynamics: A Prospective Application to hDAAO Inhibitors. *J. Chem. Inf. Model.* **2018**, *58*, 2255−2265.
(28) Gobbo, D.; Piretti, V.; Di Martino, R. M. C.; Tripathi, S. K.; Giabbai, B.; Storici, P.; Demitri, N.; Girotto, S.; Decherchi, S.; Cavalli, A. Investigating Drug−Target Residence Time in Kinases through Enhanced Sampling Simulations. *J. Chem. Theory Comput.* **2019**, *15*, 4646−4659.
(29) Bai, F.; Jiang, H. Computationally Elucidating the Binding Kinetics for Different AChE Inhibitors to Access the Rationale for Improving the Drug Efficacy. *J. Phys. Chem. B* **2022**, *126*, 7797−7805.
(30) Cho, S. S.; Levy, Y.; Wolynes, P. G. P versus Q: Structural reaction coordinates capture protein folding on smooth landscapes. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 586−591.
(31) Elber, R. A milestoning study of the kinetics of an allosteric transition: Atomically detailed simulations of Deoxy SCAPHARCA hemoglobin. *Biophys. J.* **2007**, *92*, L85−L87.
(32) Votapka, L. W.; Jagger, B. R.; Heyneman, A. L.; Amaro, R. E. SEEKR: Simulation Enabled Estimation of Kinetic Rates, A Computational Tool to Estimate Molecular Kinetics and Its Application to Trypsin−Benzamidine Binding. *J. Phys. Chem. B* **2017**, *121*, 3597−3606.
(33) Elber, R. Milestoning: An efficient approach for atomically detailed simulations of Kinetics in Biophysics. *Annu. Rev. Biophys.* **2020**, *49*, 69−85.
(34) Jagger, B. R.; Ojha, A. A.; Amaro, R. E. Predicting ligand binding kinetics using a Markovian milestoning with Voronoi tessellations multiscale approach. *J. Chem. Theory Comput.* **2020**, *16*, 5348−5357.
(35) Huber, G.; Kim, S. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* **1996**, *70*, 97−110.
(36) Zhang, B. W.; Jasnow, D.; Zuckerman, D. M. The "Weighted Ensemble" path sampling method is statistically exact for a broad class of stochastic processes and Binning procedures. *J. Chem. Phys.* **2010**, *132*, 054107.
(37) Dickson, A.; Brooks, C. L. WExplore: Hierarchical Exploration of High-Dimensional Spaces Using the Weighted Ensemble Algorithm. *J. Phys. Chem. B* **2014**, *118*, 3532−3542.
(38) Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of methodology, applications, and software. *Annu. Rev. Biophys.* **2017**, *46*, 43−57.
(39) Ray, D.; Stone, S. E.; Andricioaei, I. Markovian Weighted Ensemble Milestoning (M-WEM): Long-Time Kinetics from Short Trajectories. *J. Chem. Theory Comput.* **2022**, *18*, 79−95.
(40) Cérou, F.; Guyader, A. Adaptive multilevel splitting for Rare event analysis. *Stoch. Anal. Appl.* **2007**, *25*, 417−443.
(41) Cérou, F.; Guyader, A.; Leliévre, T.; Pommier, D. A multiple replica approach to simulate reactive trajectories. *J. Chem. Phys.* **2011**, *134*, 054108.
(42) You, W.; Tang, Z.; Chang, C.-e. A. Potential Mean Force from Umbrella Sampling Simulations: What Can We Learn and What Is Missed? *J. Chem. Theory Comput.* **2019**, *15*, 2433−2443.
(43) Spiriti, J.; Wong, C. F. Qualitative Prediction of Ligand Dissociation Kinetics from Focal Adhesion Kinase Using Steered Molecular Dynamics. *Life* **2021**, *11*, 74.
(44) Wolf, S.; Amaral, M.; Lowinski, M.; Vallée, F.; Musil, D.; Güldenhaupt, J.; Dreyer, M. K.; Bomke, J.; Frech, M.; Schlitter, J.; Gerwert, K. Estimation of protein−ligand unbinding kinetics using non-equilibrium targeted molecular dynamics simulations. *J. Chem. Inf. Model.* **2019**, *59*, 5135−5147.
(45) Wolf, S.; Stock, G. Targeted molecular dynamics calculations of free energy profiles using a nonequilibrium friction correction. *J. Chem. Theory Comput.* **2018**, *14*, 6175−6182.
(46) Wolf, S.; Lickert, B.; Bray, S.; Stock, G. Multisecond ligand dissociation dynamics from atomistic simulations. *Nat. Commun.* **2020**, *11*, 2918.
(47) Mollica, L.; Decherchi, S.; Zia, S. R.; Gaspari, R.; Cavalli, A.; Rocchia, W. Kinetics of protein-ligand unbinding via smoothed potential molecular dynamics simulations. *Sci. Rep.* **2015**, *5*, DOI: 10.1038/srep11539.
(48) Mollica, L.; Theret, I.; Antoine, M.; Perron-Sierra, F.; Charton, Y.; Fourquez, J.-M.; Wierzbicki, M.; Boutin, J. A.; Ferry, G.; Decherchi, S.; Bottegoni, G.; Ducrot, P.; Cavalli, A. Molecular dynamics simulations and kinetic measurements to estimate and predict protein−ligand residence times. *J. Med. Chem.* **2016**, *59*, 7167−7176.
(49) Marchi, M.; Ballone, P. Adiabatic bias molecular dynamics: A method to navigate the conformational space of complex molecular systems. *J. Chem. Phys.* **1999**, *110*, 3697−3702.
(50) Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. Decoding the role of water dynamics in ligand−protein unbinding: CRF1R as a test case. *J. Chem. Inf. Model.* **2015**, *55*, 1857−1866.
(51) Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 12562−12566.
(52) Tiwary, P.; Parrinello, M. From metadynamics to dynamics. *Phys. Rev. Lett.* **2013**, *111*, DOI: 10.1103/PhysRevLett.111.230602.
(53) Tiwary, P.; Limongelli, V.; Salvalaglio, M.; Parrinello, M. Kinetics of protein-ligand unbinding: Predicting pathways, rates, and rate-limiting steps. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 386−391.
(54) Wang, Y.; Valsson, O.; Tiwary, P.; Parrinello, M.; Lindorff-Larsen, K. Frequency adaptive metadynamics for the calculation of rare-event kinetics. *J. Chem. Phys.* **2018**, *149*, 072309.
(55) Invernizzi, M.; Parrinello, M. Rethinking Metadynamics: From Bias Potential to Probability Distributions. *J. Phys. Chem. Lett.* **2020**, *11*, 2731−2736.
(56) Ray, D.; Ansari, N.; Rizzi, V.; Invernizzi, M.; Parrinello, M. Rare Events Kinetics from Adaptive Bias Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 6500−6509.
(57) Ansari, N.; Rizzi, V.; Parrinello, M. Water regulates the residence time of benzamidine in Trypsin. *Nat. Commun.* **2022**, *13*, 5438.
(58) Miao, Y. Acceleration of biomolecular kinetics in gaussian accelerated molecular dynamics. *J. Chem. Phys.* **2018**, *149*, 072308.

(59) 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*, 154109−154109.
(60) Miao, Y.; Bhattarai, A.; Wang, J. Ligand Gaussian accelerated molecular dynamics (LiGaMD): Characterization of ligand binding thermodynamics and kinetics. *J. Chem. Theory Comput.* **2020**, *16*, 5526−5547.
(61) Kokh, D. B.; Amaral, M.; Bomke, J.; Grädler, U.; Musil, D.; Buchstaller, H.-P.; Dreyer, M. K.; Frech, M.; Lowinski, M.; Vallee, F.; Bianciotto, M.; Rak, A.; Wade, R. C. Estimation of drug-target residence times by τ-random acceleration molecular dynamics simulations. *J. Chem. Theory Comput.* **2018**, *14*, 3859−3869.
(62) Barducci, A.; Bussi, G.; Parrinello, M. Well-tempered metadynamics: A smoothly converging and tunable free-energy method. *Phys. Rev. Lett.* **2008**, *100*, 020603.
(63) Singhal, N.; Snow, C. D.; Pande, V. S. Using path sampling to build better Markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *J. Chem. Phys.* **2004**, *121*, 415.
(64) Paul, F.; Wehmeyer, C.; Abualrous, E. T.; Wu, H.; Crabtree, M. D.; Schöneberg, J.; Clarke, J.; Freund, C.; Weikl, T. R.; Noé, F. Protein-peptide association kinetics beyond the seconds timescale from atomistic simulations. *Nat. Commun.* **2017**, *8*, 1095.
(65) Wan, H.; Voelz, V. A. Adaptive Markov State model estimation using short reseeding trajectories. *J. Chem. Phys.* **2020**, *152*, 024103.
(66) Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2011**, *108*, 10184−10189.
(67) Doerr, S.; De Fabritiis, G. On-the-fly learning and sampling of ligand binding by high-throughput molecular simulations. *J. Chem. Theory Comput.* **2014**, *10*, 2064−2069.
(68) Wu, H.; Paul, F.; Wehmeyer, C.; Noé, F. Multiensemble Markov models of molecular thermodynamics and kinetics. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, E3221−E3230.
(69) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.
(70) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696−3713.
(71) Huang, J.; MacKerell, A. D., Jr CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J. Comput. Chem.* **2013**, *34*, 2135−2145.
(72) Yu, Y.; Krämer, A.; Venable, R. M.; Brooks, B. R.; Klauda, J. B.; Pastor, R. W. CHARMM36 Lipid Force Field with Explicit Treatment of Long-Range Dispersion: Parametrization and Validation for Phosphatidylethanolamine, Phosphatidylglycerol, and Ether Lipids. *J. Chem. Theory Comput.* **2021**, *17*, 1581−1595.
(73) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; Mackerell, A. D., Jr CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *J. Comput. Chem.* **2010**, *31*, 671−690.
(74) Lu, C.; Wu, C.; Ghoreishi, D.; Chen, W.; Wang, L.; Damm, W.; Ross, G. A.; Dahlgren, M. K.; Russell, E.; Von Bargen, C. D.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS4: Improving Force Field Accuracy on Challenging Regimes of Chemical Space. *J. Chem. Theory Comput.* **2021**, *17*, 4291−4300.
(75) Slochower, D. R.; Henriksen, N. M.; Wang, L.-P.; Chodera, J. D.; Mobley, D. L.; Gilson, M. K. Binding Thermodynamics of Host−Guest Systems with SMIRNOFF99Frosst 1.0.5 from the Open Force Field Initiative. *J. Chem. Theory Comput.* **2019**, *15*, 6225−6242.
(76) Qiu, Y.; Smith, D. G. A.; Boothroyd, S.; Jang, H.; Hahn, D. F.; Wagner, J.; Bannan, C. C.; Gokey, T.; Lim, V. T.; Stern, C. D.; Rizzi, A.; Tjanaka, B.; Tresadern, G.; Lucas, X.; Shirts, M. R.; Gilson, M. K.; Chodera, J. D.; Bayly, C. I.; Mobley, D. L.; Wang, L.-P. Development and Benchmarking of Open Force Field v1.0.0—the Parsley Small-Molecule Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 6262−6280.
(77) Nerenberg, P. S.; Head-Gordon, T. New developments in force fields for biomolecular simulations. *Curr. Opin. Struct. Biol.* **2018**, *49*, 129−138.
(78) Raniolo, S.; Limongelli, V. Improving Small-Molecule Force Field Parameters in Ligand Binding Studies. *Front. Mol. Biosci.* **2021**, *8*, 760283.
(79) D'Amore, L.; Hahn, D. F.; Dotson, D. L.; Horton, J. T.; Anwar, J.; Craig, I.; Fox, T.; Gobbi, A.; Lakkaraju, S. K.; Lucas, X.; Meier, K.; Mobley, D. L.; Narayanan, A.; Schindler, C. E. M.; Swope, W. C.; in 't Veld, P. J.; Wagner, J.; Xue, B.; Tresadern, G. Collaborative Assessment of Molecular Geometries and Energies from the Open Force Field. *J. Chem. Inf. Model.* **2022**, *62*, 6094−6104.
(80) Kashefolgheta, S.; Oliveira, M. P.; Rieder, S. R.; Horta, B. A.; Acree, W. E.; Hünenberger, P. H. Evaluating classical force fields against experimental cross-solvation free energies. *J. Chem. Theory Comput.* **2020**, *16*, 7556−7580.
(81) Fan, S.; Iorga, B. I.; Beckstein, O. Prediction of octanol-water partition coefficients for the SAMPL6-log *P* molecules using molecular dynamics simulations with OPLS-AA, AMBER and CHARMM force fields. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 543−560.
(82) Zhu, S. Validation of the Generalized Force Fields GAFF, CGenFF, OPLS-AA, and PRODRGFF by Testing Against Experimental Osmotic Coefficient Data for Small Drug-Like Molecules. *J. Chem. Inf. Model.* **2019**, *59*, 4239−4247.
(83) Baker, C. M. Polarizable force fields for molecular dynamics simulations of biomolecules. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2015**, *5*, 241−254.
(84) Hensen, C.; Hermann, J. C.; Nam, K.; Ma, S.; Gao, J.; Höltje, H.-D. A Combined QM/MM Approach to Protein-Ligand Interactions: Polarization Effects of the HIV-1 Protease on Selected High Affinity Inhibitors. *J. Med. Chem.* **2004**, *47*, 6673−6680.
(85) Kaminský, J.; Jensen, F. Conformational Interconversions of Amino Acid Derivatives. *J. Chem. Theory Comput.* **2016**, *12*, 694−705.
(86) Vitalini, F.; Mey, A. S.; Noé, F.; Keller, B. G. Dynamic properties of force fields. *J. Chem. Phys.* **2015**, *142*, 084101.
(87) Jing, Z.; Liu, C.; Cheng, S. Y.; Qi, R.; Walker, B. D.; Piquemal, J.-P.; Ren, P. Polarizable force fields for biomolecular simulations: Recent Advances and Applications. *Annu. Rev. Biophys.* **2019**, *48*, 371−394.
(88) Wang, Z.-X.; Zhang, W.; Wu, C.; Lei, H.; Cieplak, P.; Duan, Y. Strike a balance: Optimization of backbone torsion parameters of amber polarizable force field for simulations of proteins and peptides. *J. Comput. Chem.* **2006**, *27*, 781−790.
(89) Baker, C. M.; Lopes, P. E.; Zhu, X.; Roux, B.; MacKerell, A. D. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the charmm drude polarizable force field. *J. Chem. Theory Comput.* **2010**, *6*, 1181−1198.
(90) Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J.-P. Anisotropic, Polarizable Molecular Mechanics Studies of Inter- and Intramolecular Interactions and Ligand-Macromolecule Complexes. A Bottom-Up Strategy. *J. Chem. Theory Comput.* **2007**, *3*, 1960−1986.
(91) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A. J.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the Amoeba Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(92) Shi, Y.; Zhu, C. Z.; Martin, S. F.; Ren, P. Probing the Effect of Conformational Constraint on Phosphorylated Ligand Binding to an SH2 Domain Using Polarizable Force Field Simulations. *J. Phys. Chem. B* **2012**, *116*, 1716−1727.
(93) Qi, R.; Jing, Z.; Liu, C.; Piquemal, J.-P.; Dalby, K. N.; Ren, P. Elucidating the Phosphate Binding Mode of Phosphate-Binding Protein: The Critical Effect of Buffer Solution. *J. Phys. Chem. B* **2018**, *122*, 6371−6376.

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061

(94) Chung, M. K. J.; Miller, R. J.; Novak, B.; Wang, Z.; Ponder, J. W. Accurate Host−Guest Binding Free Energies Using the AMOEBA Polarizable Force Field. *J. Chem. Inf. Model.* **2023**, *63*, 2769.
(95) Ponder, J. W.; Wu, C.; Ren, P.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A. J.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current Status of the AMOEBA Polarizable Force Field. *J. Phys. Chem. B* **2010**, *114*, 2549−2564.
(96) Shi, Y.; Xia, Z.; Zhang, J.; Best, R.; Wu, C.; Ponder, J. W.; Ren, P. Polarizable Atomic Multipole-Based AMOEBA Force Field for Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 4046−4063.
(97) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. J. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116*, 4983−5013.
(98) Rupakheti, C. R.; MacKerell, A. D. J.; Roux, B. Global Optimization of the Lennard-Jones Parameters for the Drude Polarizable Force Field. *J. Chem. Theory Comput.* **2021**, *17*, 7085−7095.
(99) Chatterjee, P.; Sengul, M. Y.; Kumar, A.; MacKerell, A. D. J. Harnessing Deep Learning for Optimization of Lennard-Jones Parameters for the Polarizable Classical Drude Oscillator Force Field. *J. Chem. Theory Comput.* **2022**, *18*, 2388−2407.
(100) Gkeka, P.; Stoltz, G.; Barati Farimani, A.; Belkacemi, Z.; Ceriotti, M.; Chodera, J. D.; Dinner, A. R.; Ferguson, A. L.; Maillet, J.-B.; Minoux, H.; Peter, C.; Pietrucci, F.; Silveira, A.; Tkatchenko, A.; Trstanova, Z.; Wiewiora, R.; Lelièvre, T. Machine learning force fields and coarse-grained variables in molecular dynamics: Application to materials and Biological Systems. *J. Chem. Theory Comput.* **2020**, *16*, 4757−4775.
(101) Unke, O. T.; Chmiela, S.; Sauceda, H. E.; Gastegger, M.; Poltavsky, I.; Schütt, K. T.; Tkatchenko, A.; Müller, K.-R. Machine Learning Force Fields. *Chem. Rev.* **2021**, *121*, 10142−10186.
(102) Batzner, S.; Musaelian, A.; Sun, L.; Geiger, M.; Mailoa, J. P.; Kornbluth, M.; Molinari, N.; Smidt, T. E.; Kozinsky, B. E(3)-equivariant graph neural networks for data-efficient and accurate interatomic potentials. *Nat. Commun.* **2022**, *13*, 2453.
(103) Fu, X.; Wu, Z.; Wang, W.; Xie, T.; Keten, S.; Gomez-Bombarelli, R.; Jaakkola, T. Forces are not Enough: Benchmark and Critical Evaluation for Machine Learning Force Fields with Molecular Simulations. *arXiv:2210.07237* **2022**, DOI: 10.48550/arXiv.2210.07237.
(104) Hu, W.; Shuaibi, M.; Das, A.; Goyal, S.; Sriram, A.; Leskovec, J.; Parikh, D.; Zitnick, C. L. ForceNet: A Graph Neural Network for Large-Scale Quantum Calculations. *arXiv:2103.01436v1* **2021**, DOI: 10.48550/arXiv.2103.01436.
(105) Schütt, K. T.; Kindermans, P.-J.; Sauceda, H. E.; Chmiela, S.; Tkatchenko, A.; Müller, K.-R. *SchNet: A Continuous-Filter Convolutional Neural Network for Modeling Quantum Interactions. Proceedings of the 31st International Conference on Neural Information Processing Systems*; von Luxburg, U., Guyon, I., Bengio, S., Wallach, H., Fergus, R., Eds; Long Beach California, USA, December 4−9, 2017; Curran Associates, Inc.: Red Hook, NY, United States, 2017; pp 992−1002.
(106) Zhang, L.; Han, J.; Wang, H.; Saidi, W.; Car, R.; E, W. End-to-end Symmetry Preserving Inter-atomic Potential Energy Model for Finite and Extended Systems. *Advances in Neural Information Processing Systems*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Montréal, Canada, December 2−8, 2018; Curran Associates, Inc., 2018.
(107) Gasteiger, J.; Becker, F.; Günnemann, S. GemNet: Universal Directional Graph Neural Networks for Molecules. *Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P. S., Wortman Vaughan, J., Eds.; December 6−14, 2021; Curran Associates, Inc.: Red Hook, NY, United States, 2021; pp 6790−6802.
(108) Wang, J.; Olsson, S.; Wehmeyer, C.; Pérez, A.; Charron, N. E.; de Fabritiis, G.; Noé, F.; Clementi, C. Machine Learning of Coarse-Grained Molecular Dynamics Force Fields. *ACS Cent. Sci.* **2019**, *5*, 755−767.
(109) Lelimousin, M.; Limongelli, V.; Sansom, M. S. P. Conformational Changes in the Epidermal Growth Factor Receptor: Role of the Transmembrane Domain Investigated by Coarse-Grained MetaDynamics Free Energy Calculations. *J. Am. Chem. Soc.* **2016**, *138*, 10611−10622.
(110) Singh, N.; Li, W. Recent Advances in Coarse-Grained Models for Biomolecules and Their Applications. *Int. J. Mol. Sci.* **2019**, *20*, 3774.
(111) Souza, P. C. T.; Alessandri, R.; Barnoud, J.; Thallmair, S.; Faustino, I.; Grünewald, F.; Patmanidis, I.; Abdizadeh, H.; Bruininks, B. M. H.; Wassenaar, T. A.; Kroon, P. C.; Melcr, J.; Nieto, V.; Corradi, V.; Khan, H. M.; Domański, J.; Javanainen, M.; Martinez-Seara, H.; Reuter, N.; Best, R. B.; Vattulainen, I.; Monticelli, L.; Periole, X.; Tieleman, D. P.; de Vries, A. H.; Marrink, S. J. Martini 3: a general purpose force field for coarse-grained molecular dynamics. *Nat. Methods.* **2021**, *18*, 382−388.
(112) Marrink, S. J.; Monticelli, L.; Melo, M. N.; Alessandri, R.; Tieleman, D. P.; Souza, P. C. Two decades of martini: Better beads, broader scope. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2023**, *13*, 13.
(113) Galanti, M.; Fanelli, D.; Piazza, F. Conformation-controlled binding kinetics of antibodies. *Sci. Rep.* **2016**, *6*, DOI: 10.1038/srep18976.
(114) Xie, Z.-R.; Chen, J.; Wu, Y. Predicting protein−protein association rates using coarse-grained simulation and machine learning. *Sci. Rep.* **2017**, *7*, DOI: 10.1038/srep46622.
(115) He, Z.; Paul, F.; Roux, B. A critical perspective on Markov state model treatments of protein−protein association using coarse-grained simulations. *J. Chem. Phys.* **2021**, *154*, 084101.
(116) Souza, P. C.; Limongelli, V.; Wu, S.; Marrink, S. J.; Monticelli, L. Perspectives on high-throughput ligand/protein docking with Martini MD simulations. *Front. Mol. Biosci.* **2021**, *8*, 8.
(117) Dandekar, B. R.; Mondal, J. Capturing Protein−Ligand Recognition Pathways in Coarse-Grained Simulation. *J. Phys. Chem. Lett.* **2020**, *11*, 5302−5311.
(118) Negami, T.; Shimizu, K.; Terada, T. Coarse-grained molecular dynamics simulation of protein conformational change coupled to ligand binding. *Chem. Phys. Lett.* **2020**, *742*, 137144.
(119) Souza, P. C. T.; Thallmair, S.; Conflitti, P.; Ramírez-Palacios, C.; Alessandri, R.; Raniolo, S.; Limongelli, V.; Marrink, S. J. Protein−ligand binding with the coarse-grained Martini model. *Nat. Commun.* **2020**, *11*, 3714.
(120) Alessandri, R.; Souza, P. C. T.; Thallmair, S.; Melo, M. N.; de Vries, A. H.; Marrink, S. J. Pitfalls of the Martini Model. *J. Chem. Theory Comput.* **2019**, *15*, 5448−5460.
(121) Filipe, H. A. L.; Esteves, M. I. M.; Henriques, C. A.; Antunes, F. E. Effect of Protein Flexibility from Coarse-Grained Elastic Network Parameterizations on the Calculation of Free Energy Profiles of Ligand Binding. *J. Chem. Theory Comput.* **2020**, *16*, 4734−4743.
(122) Jarin, Z.; Newhouse, J.; Voth, G. A. Coarse-Grained Force Fields from the Perspective of Statistical Mechanics: Better Understanding of the Origins of a MARTINI Hangover. *J. Chem. Theory Comput.* **2021**, *17*, 1170−1180.
(123) Sadeghi, M.; Noé, F. Large-scale simulation of biomembranes incorporating realistic kinetics into coarse-grained models. *Nat. Commun.* **2020**, *11*, 2951.
(124) Jin, J.; Pak, A. J.; Durumeric, A. E. P.; Loose, T. D.; Voth, G. A. Bottom-up Coarse-Graining: Principles and Perspectives. *J. Chem. Theory Comput.* **2022**, *18*, 5759−5791.
(125) Yang, W.; Templeton, C.; Rosenberger, D.; Bittracher, A.; Nüske, F.; Noé, F.; Clementi, C. Slicing and Dicing: Optimal Coarse-Grained Representation to Preserve Molecular Kinetics. *ACS Cent. Sci.* **2023**, *9*, 186−196.
(126) Durumeric, A. E.; Charron, N. E.; Templeton, C.; Musil, F.; Bonneau, K.; Pasos-Trejo, A. S.; Chen, Y.; Kelkar, A.; Noé, F.; Clementi, C. Machine learned coarse-grained protein force-fields: Are we there yet? *Curr. Opin. Struct. Biol.* **2023**, *79*, 102533.
(127) Alom, M. Z.; Taha, T. M.; Yakopcic, C.; Westberg, S.; Sidike, P.; Nasrin, M. S.; Esesn, B. C. V.; Awwal, A. A. S.; Asari, V. K. History Began from AlexNet: A Comprehensive Survey on Deep Learning

Approaches. 2018; arXiv:1803.01164. arXiv.org ePrint archive. https://arxiv.org/abs/1803.01164 (accessed Feb 13, 2023).
(128) Piana, S.; Lindorff-Larsen, K.; Shaw, D. E. Protein folding kinetics and thermodynamics from atomistic simulation. *Proc. Natl. Acad. Sci. U.S.A.* **2012**, *109*, 17845−17850.
(129) Pan, A. C.; Borhani, D. W.; Dror, R. O.; Shaw, D. E. Molecular determinants of drug−receptor binding kinetics. *Drug Discovery Today* **2013**, *18*, 667−673.
(130) Shaw, D. E.; Grossman, J.; Bank, J. A.; Batson, B.; Butts, J. A.; Chao, J. C.; Deneroff, M. M.; Dror, R. O.; Even, A.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Greskamp, B.; Ho, C. R.; Ierardi, D. J.; Iserovich, L.; Kuskin, J. S.; Larson, R. H.; Layman, T.; Lee, L.-S.; Lerer, A. K.; Li, C.; Killebrew, D.; Mackenzie, K. M.; Mok, S. Y.-H.; Moraes, M. A.; Mueller, R.; Nociolo, L. J.; Peticolas, J. L.; Quan, T.; Ramot, D.; Salmon, J. K.; Scarpazza, D. P.; Schafer, U. B.; Siddique, N.; Snyder, C. W.; Spengler, J.; Tang, P. T. P.; Theobald, M.; Toma, H.; Towles, B.; Vitale, B.; Wang, S. C.; Young, C. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *SC '14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*. New Orleans, LA, November 16−21, 2014; IEEE: New York, NY, United States, 2014.
(131) Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2014**, *24*, 98−105.
(132) Pan, A. C.; Weinreich, T. M.; Piana, S.; Shaw, D. E. Demonstrating an Order-of-Magnitude Sampling Enhancement in Molecular Dynamics Simulations of Complex Protein Systems. *J. Chem. Theory Comput.* **2016**, *12*, 1360−1367.
(133) Shaw, D. E.; Adams, P. J.; Azaria, A.; Bank, J. A.; Batson, B.; Bell, A.; Bergdorf, M.; Bhatt, J.; Butts, J. A.; Correia, T.; Dirks, R. M.; Dror, R. O.; Eastwood, M. P.; Edwards, B.; Even, A.; Feldmann, P.; Fenn, M.; Fenton, C. H.; Forte, A.; Gagliardo, J.; Gill, G.; Gorlatova, M.; Greskamp, B.; Grossman, J.; Gullingsrud, J.; Harper, A.; Hasenplaugh, W.; Heily, M.; Heshmat, B. C.; Hunt, J.; Ierardi, D. J.; Iserovich, L.; Jackson, B. L.; Johnson, N. P.; Kirk, M. M.; Klepeis, J. L.; Kuskin, J. S.; Mackenzie, K. M.; Mader, R. J.; McGowen, R.; McLaughlin, A.; Moraes, M. A.; Nasr, M. H.; Nociolo, L. J.; OĎonnell, L.; Parker, A.; Peticolas, J. L.; Pocina, G.; Predescu, C.; Quan, T.; Salmon, J. K.; Schwink, C.; Shim, K. S.; Siddique, N.; Spengler, J.; Szalay, T.; Tabladillo, R.; Tartler, R.; Taube, A. G.; Theobald, M.; Towles, B.; Vick, W.; Wang, S. C.; Wazlowski, M.; Weingarten, M. J.; Williams, J. M.; Yuh, K. A. Anton 3: Twenty Microseconds of Molecular Dynamics Simulation before Lunch. *SC '21: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*; St. Louis, Missouri, November 14−19, 2021; Association for Computing Machinery: New York, NY, United States, 2021.
(134) Bernardi, R. C.; Melo, M. C. R.; Schulten, K. Enhanced sampling techniques in molecular dynamics simulations of biological systems. *Biochim. Biophys. Acta - Gen. Subj.* **2015**, *1850*, 872−877.
(135) Lazim, R.; Suh, D.; Choi, S. Advances in Molecular Dynamics Simulations and Enhanced Sampling Methods for the Study of Protein Systems. *Int. J. Mol. Sci.* **2020**, *21*, 6339.
(136) Williams-Noonan, B. J.; Yuriev, E.; Chalmers, D. K. Free energy methods in drug design: Prospects of "Alchemical perturbation" in medicinal chemistry. *J. Med. Chem.* **2018**, *61*, 638−649.
(137) Wang, E.; Sun, H.; Wang, J.; Wang, Z.; Liu, H.; Zhang, J. Z. H.; Hou, T. End-Point Binding Free Energy Calculation with MM/PBSA and MM/GBSA: Strategies and Applications in Drug Design. *Chem. Rev.* **2019**, *119*, 9478−9508.
(138) Noé, F.; Clementi, C. Collective variables for the study of long-time kinetics from molecular trajectories: theory and methods. *Curr. Opin. Struct. Biol.* **2017**, *43*, 141−147.
(139) Sultan, M. M.; Pande, V. S. Automated design of collective variables using supervised machine learning. *J. Chem. Phys.* **2018**, *149*, 094106.
(140) Zhang, Y.-Y.; Niu, H.; Piccini, G.; Mendels, D.; Parrinello, M. Improving collective variables: The case of crystallization. *J. Chem. Phys.* **2019**, *150*, 094509.
(141) Schwantes, C. R.; Pande, V. S. Modeling Molecular Kinetics with tICA and the Kernel Trick. *J. Chem. Theory Comput.* **2015**, *11*, 600−6008.
(142) Sultan, M. M.; Pande, V. S. tICA-Metadynamics: Accelerating Metadynamics by Using Kinetically Selected Collective Variables. *J. Chem. Theory Comput.* **2017**, *13*, 2440−2447.
(143) Bernetti, M.; Bertazzo, M.; Masetti, M. Data-Driven Molecular Dynamics: A Multifaceted Challenge. *Pharmaceuticals* **2020**, *13*, 253−283.
(144) Wang, Y.; ao Marcelo Lamim Ribeiro, J.; Tiwary, P. Machine learning approaches for analyzing and enhancing molecular dynamics simulations. *Curr. Opin. Struct. Biol.* **2020**, *61*, 139−145.
(145) Schultze, S.; Grubmüller, H. Time-Lagged Independent Component Analysis of Random Walks and Protein Dynamics. *J. Chem. Theory Comput.* **2021**, *17*, 5766−5776.
(146) Bonati, L.; Rizzi, V.; Parrinello, M. Data-Driven Collective Variables for Enhanced Sampling. *J. Phys. Chem. Lett.* **2020**, *11*, 2998−3004.
(147) Bonati, L.; Piccini, G.; Parrinello, M. Deep learning the slow modes for rare events sampling. *Proc. Natl. Acad. Sci. U.S.A.* **2021**, *118*, No. e2113533118.
(148) Badaoui, M.; Buigues, P. J.; Berta, D.; Mandana, G. M.; Gu, H.; Földes, T.; Dickson, C. J.; Hornak, V.; Kato, M.; Molteni, C.; Parsons, S.; Rosta, E. Combined Free-Energy Calculation and Machine Learning Methods for Understanding Ligand Unbinding Kinetics. *J. Chem. Theory Comput.* **2022**, *18*, 2543−2555.
(149) Paul, F.; Wu, H.; Vossel, M.; de Groot, B. L.; Noé, F. Identification of kinetic order parameters for non-equilibrium dynamics. *J. Chem. Phys.* **2019**, *150*, 164120.
(150) Ludwig, J.; Smith, J.; Pfaendtner, J. Analyzing the Long Time-Scale Dynamics of Uremic Toxins Bound to Sudlow Site II in Human Serum Albumin. *J. Phys. Chem. B* **2021**, *125*, 2910−2920.
(151) ao Marcelo Lamin Ribeiro, J.; Provasi, D.; Filizola, M. A combination of machine learning and infrequent metadynamics to efficiently predict kinetic rates, transition states, and molecular determinants of drug dissociation from G protein-coupled receptors. *J. Chem. Phys.* **2020**, *153*, 124105.
(152) Kokh, D. B.; Kaufmann, T.; Kister, B.; Wade, R. C. Machine Learning Analysis of τRAMD Trajectories to Decipher Molecular Determinants of Drug-Target Residence Times. *Front. Mol. Biosci.* **2019**, *6*, DOI: 10.3389/fmolb.2019.00036.
(153) Mendels, D.; de Pablo, J. J. Collective Variables for Free Energy Surface Tailoring: Understanding and Modifying Functionality in Systems Dominated by Rare Events. *J. Phys. Chem. Lett.* **2022**, *13*, 2830−2837.
(154) Ensing, B.; Baerends, E. Reaction path sampling of the reaction between iron(II) and hydrogen peroxide in aqueous solution. *J. Phys. Chem. A* **2002**, *106*, 7902−7910.
(155) Morelli, M. J.; Allen, R. J.; Tanase-Nicola, S.; ten Wolde, P. R. Eliminating fast reactions in stochastic simulations of biochemical networks: a bistable genetic switch. *J. Chem. Phys.* **2008**, *128*, 045105.
(156) Velez-Vega, C.; Borrero, E. E.; Escobedo, F. A. Kinetics and mechanism of the unfolding native-to-loop transition of Trp-cage in explicit solvent via optimized forward flux sampling simulations. *J. Chem. Phys.* **2010**, *133*, 105103.
(157) Becker, N. B.; Allen, R. J.; ten Wolde, P. R. Non-stationary forward flux sampling. *J. Chem. Phys.* **2012**, *136*, 174118.
(158) van Erp, T. S.; Bolhuis, P. G. Elaborating transition interface sampling methods. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *205*, 157−181.
(159) Weinan, E.; Ren, W.; Vanden-Eijnden, E. Simplified and improved string method for computing the minimum energy paths in barrier-crossing events. *J. Chem. Phys.* **2007**, *126*, 164103.
(160) Maragliano, L.; Vanden-Eijnden, E. On-the-fly string method for minimum free energy paths calculation. *Chem. Phys. Lett.* **2007**, *446*, 182−190.

(161) Votapka, L. W.; Stokely, A. M.; Ojha, A. A.; Amaro, R. E. SEEKR2: Versatile Multiscale Milestoning Utilizing the OpenMM Molecular Dynamics Engine. *J. Chem. Inf. Model.* **2022**, *62*, 3253−3262.
(162) McCarty, J.; Parrinello, M. A variational conformational dynamics approach to the selection of collective variables in metadynamics. *J. Chem. Phys.* **2017**, *147*, 204109.
(163) Brotzakis, Z. F.; Limongelli, V.; Parrinello, M. Accelerating the calculation of protein−ligand binding free energy and residence times using dynamically optimized collective variables. *J. Chem. Theory Comput.* **2019**, *15*, 743−750.
(164) Ribeiro, J. M. L.; Bravo, P.; Wang, Y.; Tiwary, P. Reweighted autoencoded variational Bayes for enhanced sampling (RAVE). *J. Chem. Phys.* **2018**, *149*, 072301.
(165) Tsai, S.-T.; Smith, Z.; Tiwary, P. SGOOP-d: Estimating Kinetic Distances and Reaction Coordinate Dimensionality for Rare Event Systems from Biased/Unbiased Simulations. *J. Chem. Theory Comput.* **2021**, *17*, 6757−6765.
(166) Invernizzi, M.; Parrinello, M. Exploration vs Convergence Speed in Adaptive-Bias Enhanced Sampling. *J. Chem. Theory Comput.* **2022**, *18*, 3988−3996.
(167) Du, R.; Pande, V. S.; Grosberg, A.; Tanaka, T.; Shakhnovich, E. On the transition coordinate for protein folding. *J. Chem. Phys.* **1998**, *108*, 334−350.
(168) Eyring, H. Transmission coefficient in reaction rate. *Rev. Mod. Phys.* **1962**, *34*, 616−619.
(169) Maragliano, L.; Fischer, A.; Vanden-Eijnden, E.; Ciccotti, G. String method in collective variables: Minimum free energy paths and isocommittor surfaces. *J. Chem. Phys.* **2006**, *125*, 024106.
(170) Tiwary, P.; Mondal, J.; Morrone, J. A.; Berne, B. J. Role of water and steric constraints in the kinetics of cavity−ligand unbinding. *Proc. Natl. Acad. Sci. U.S.A.* **2015**, *112*, 12015−12019.
(171) Casasnovas, R.; Limongelli, V.; Tiwary, P.; Carloni, P.; Parrinello, M. Unbinding Kinetics of a p38 MAP Kinase Type II Inhibitor from Metadynamics Simulations. *J. Am. Chem. Soc.* **2017**, *139*, 4780−4788.
(172) He, Z.; Chipot, C.; Roux, B. Committor-Consistent Variational String Method. *J. Phys. Chem. Lett.* **2022**, *13*, 9263−9271.
(173) Chen, H.; Roux, B.; Chipot, C. Discovering Reaction Pathways, Slow Variables, and Committor Probabilities with Machine Learning. *J. Chem. Theory Comput.* **2023**, *19*, 4414−4426.
(174) Rotskoff, G. M.; Mitchell, A. R.; Vanden-Eijnden, E. Active Importance Sampling for Variational Objectives Dominated by Rare Events: Consequences for Optimization and Generalization. *Proceedings of the 2nd Mathematical and Scientific Machine Learning Conference*; Bruna, J., Hesthaven, J., Zdeborova, L., Eds.; August 16−19, 2021; PMLR, 2022; pp 757−780.
(175) Roux, B. Transition rate theory, spectral analysis, and reactive paths. *J. Chem. Phys.* **2022**, *156*, 134111.
(176) Chen, M. Collective variable-based enhanced sampling and machine learning. *Eur. Phys. J. B* **2021**, *94*, 1−17.
(177) Bhakat, S. Collective variable discovery in the age of machine learning: reality, hype and everything in between. *RSC Adv.* **2022**, *12*, 25010−25024.
(178) Huang, S.; Xiao, S.; Feng, W. On the energy efficiency of graphics processing units for scientific computing. *2009 IEEE International Symposium on Parallel & Distributed Processing*; Rome, Italy, May 23−29, 2009; IEEE: New York, NY, United States, 2009.
(179) Mastelic, T.; Oleksiak, A.; Claussen, H.; Brandic, I.; Pierson, J.-M.; Vasilakos, A. V. Cloud computing: Survey on Energy Efficiency. *ACM Comput. Surv.* **2015**, *47*, 1−36.
(180) Wilson, G.; Bryan, J.; Cranston, K.; Kitzes, J.; Nederbragt, L.; Teal, T. K. Good enough practices in scientific computing. *PLoS Comput. Biol.* **2017**, *13*, e1005510.
(181) Ivie, P.; Thain, D. Reproducibility in scientific computing. *ACM Comput. Surv.* **2019**, *51*, 1−36.
(182) Guillain, F.; Thusius, D. Use of proflavine as an indicator in temperature-jump studies of the binding of a competitive inhibitor to trypsin. *J. Am. Chem. Soc.* **1970**, *92*, 5534−5536.
(183) Plattner, N.; Noé, F. Protein conformational plasticity and complex ligand-binding kinetics explored by atomistic simulations and Markov models. *Nat. Commun.* **2015**, *6*, 7653.
(184) Teo, I.; Mayne, C. G.; Schulten, K.; Leliévre, T. Adaptive multilevel splitting method for molecular dynamics calculation of benzamidine-trypsin dissociation time. *J. Chem. Theory Comput.* **2016**, *12*, 2983−2989.
(185) Dickson, A.; Lotz, S. D. Multiple Ligand Unbinding Pathways and Ligand-Induced Destabilization Revealed by WExplore. *Biophys. J.* **2017**, *112*, 620−629.
(186) Donyapour, N.; Roussey, N. M.; Dickson, A. REVO: Resampling of ensembles by variation optimization. *J. Chem. Phys.* **2019**, *150*, 244112.
(187) Leiserson, C. E.; Thompson, N. C.; Emer, J. S.; Kuszmaul, B. C.; Lampson, B. W.; Sanchez, D.; Schardl, T. B. There's plenty of room at the top: What will drive computer performance after moore's law? *Science* **2020**, *368*, 6495.
(188) Outeiral, C.; Strahm, M.; Shi, J.; Morris, G. M.; Benjamin, S. C.; Deane, C. M. The prospects of Quantum Computing in computational molecular biology. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2021**, *11*, e1481.
(189) Tribello, G. A.; Bonomi, M.; Branduardi, D.; Camilloni, C.; Bussi, G. PLUMED 2: New feathers for an old bird. *Comput. Phys. Commun.* **2014**, *185*, 604−613.
(190) Raniolo, S.; Limongelli, V. Ligand binding free-energy calculations with funnel metadynamics. *Nat. Protoc.* **2020**, *15*, 2837−2866.
(191) The PLUMED consortium. Promoting transparency and reproducibility in enhanced molecular simulations. *Nat. Methods* **2019**, *16*, 670−673.
(192) Salomon-Ferrer, R.; Case, D. A.; Walker, R. C. An overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2013**, *3*, 198−210.
(193) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859−1865.
(194) Bayarri, G.; Hospital, A.; Orozco, M. 3dRS, a Web-Based Tool to Share Interactive Representations of 3D Biomolecular Structures and Molecular Dynamics Trajectories. *Front. Mol. Biosci.* **2021**, *8*, 726232 DOI: 10.3389/fmolb.2021.726232.
(195) Rodríguez-Espigares, I.; Torrens-Fontanals, M.; Johanna; Aranda-García, D.; Ramírez-Anguita, J. M.; Stepniewski, T. M.; Worp, N.; Varela-Rial, A.; Morales-Pastor, A.; Medel-Lacruz, B.; Pándy-Szekeres, G.; Mayol, E.; Giorgino, T.; Carlsson, J.; Deupi, X.; Filipek, S.; Filizola, M.; Gómez-Tamayo, J. C.; González, A. F.; Gutiérrez-de Terán, H. GPCRMD uncovers the dynamics of the 3D-GPCRome. *Nat. Methods.* **2020**, *17*, 777−787.
(196) Marchetto, A.; Si Chaib, Z.; Rossi, C. A.; Ribeiro, R.; Pantano, S.; Rossetti, G.; Giorgetti, A. CGMD platform: Integrated web servers for the preparation, running, and analysis of coarse-grained molecular dynamics simulations. *Molecules* **2020**, *25*, 5934.
(197) Meyer, T.; D'Abramo, M.; Hospital, A.; Rueda, M.; Ferrer-Costa, C.; Pérez, A.; Carrillo, O.; Camps, J.; Fenollosa, C.; Repchevsky, D.; Gelpí, J. L.; Orozco, M. MoDEL (Molecular Dynamics Extended Library): ADdatabase of Atomistic Molecular Dynamics Trajectories. *Structure* **2010**, *18*, 1399−1409.
(198) Hospital, A.; Andrio, P.; Cugnasco, C.; Codo, L.; Becerra, Y.; Dans, P. D.; Battistini, F.; Torres, J.; Goñi, R.; Orozco, M.; Gelpí, J. L. BIGNASim: A NoSQL database structure and analysis portal for Nucleic Acids Simulation Data. *Nucleic Acids Res.* **2016**, *44*, D272−D278.
(199) Ehrman, J. N.; Lim, V. T.; Bannan, C. C.; Thi, N.; Kyu, D. Y.; Mobley, D. L. Improving small molecule force fields by identifying and characterizing small molecules with inconsistent parameters. *J. Comput.-Aided Mol. Des.* **2021**, *35*, 271−284.
(200) Harder, E.; Damm, W.; Maple, J.; Wu, C.; Reboul, M.; Xiang, J. Y.; Wang, L.; Lupyan, D.; Dahlgren, M. K.; Knight, J. L.; Kaus, J. W.; Cerutti, D. S.; Krilov, G.; Jorgensen, W. L.; Abel, R.; Friesner, R. A. OPLS3: A Force Field Providing Broad Coverage of Drug-like

Small Molecules and Proteins. *J. Chem. Theory Comput.* **2016**, *12*, 281−296.
(201) Roos, K.; Wu, C.; Damm, W.; Reboul, M.; Stevenson, J. M.; Lu, C.; Dahlgren, M. K.; Mondal, S.; Chen, W.; Wang, L.; Abel, R.; Friesner, R. A.; Harder, E. D. OPLS3e: Extending Force Field Coverage for Drug-Like Small Molecules. *J. Chem. Theory Comput.* **2019**, *15*, 1863−1874.
(202) Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Slochower, D. R.; Shirts, M. R.; Gilson, M. K.; Eastman, P. K. Escaping Atom Types in Force Fields Using Direct Chemical Perception. *J. Chem. Theory Comput.* **2018**, *14*, 6076−6092.
(203) Raghavan, B.; Paulikat, M.; Ahmad, K.; Callea, L.; Rizzi, A.; Ippoliti, E.; Mandelli, D.; Bonati, L.; De Vivo, M.; Carloni, P. Drug Design in the Exascale Era: A Perspective from Massively Parallel QM/MM Simulations. *J. Chem. Inf. Model.* **2023**, *63*, 3647−3658.

https://doi.org/10.1021/acs.jctc.3c00641
J. Chem. Theory Comput. 2023, 19, 6047−6061