Chapter 1

Free Energy Methods in Drug Discovery—Introduction

Zoe Cournia,[*,1] Christophe Chipot,[*,2,3,4] Benoît Roux,[*,5] Darrin M. York,[*,6,7,8] and Woody Sherman[*,9]

[1]Biomedical Research Foundation, Academy of Athens, 4 Soranou Ephesiou, 11527 Athens, Greece

[2]Laboratoire International Associé CNRS, UMR n°7019, Université de Lorraine, BP 70239, F-54506 Vandœuvre-lès-Nancy, France

[3]Department of Physics, University of Illinois at Urbana-Champaign, 1110 West Green Street, Urbana, Illinois 61801, United States

[4]Theoretical and Computational Biophysics Group, Beckman Institute for Advanced Science and Technology, University of Illinois at Urban-Champaign, 405 North Mathews, Urbana, Illinois 61801, United States

[5]Department of Chemistry, Gordon Center for Integrative Science, University of Chicago, Chicago, Illinois 60637, United States

[6]Laboratory for Biomolecular Simulation Research, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854-8076, United States

[7]Institute for Quantitative Biomedicine, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854-8076, United States

[8]Department of Chemistry and Chemical Biology, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, United States

[9]Roivant Sciences and Silicon Therapeutics, Boston, Massachusetts 02110, United States

[*]Email: zcournia@bioacademy.gr

[*]Email: chipot@illinois.edu

[*]Email: roux@uchicago.edu

[*]Email: darrin.york@rutgers.edu

[*]Email: woody.sherman@roivant.com

Complete understanding of most, if not all chemical processes requires at its very core the knowledge of the underlying free-energy change. In computer-aided drug design, for instance, such processes as binding of a drug to a protein or its spontaneous partitioning across the cell membrane cannot be predicted reliably without considering how the associated free energy varies. Owing to relentless theoretical developments, which have benefited from ever-growing computational resources, free-energy calculations leaning on statistical-mechanics simulations are now part of the arsenal of robust and well-characterized modeling tools.

Armacost and Thompson; Free Energy Methods in Drug Discovery: Current State and Future Directions
ACS Symposium Series; American Chemical Society: Washington, DC, 0.

However, as will be explained below and touched upon throughout the chapters of this book, it is still challenging to obtain accurate and reliable free-energy predictions for biomolecules due to the many nuances in the system setup and the unknown unknowns such as whether a given simulation is globally converged or only locally converged, perhaps in an incorrect free-energy basin. In all but the simplest cases, free-energy simulations still require experts in the field to prepare the system, run the calculations, and analyze the results in order to obtain robust predictions that can be confidently used to make decisions in drug discovery campaigns.

Introduction

Foundations of Free-Energy Simulations and Historical Backdrop

The calculation of free energies is among the most important applications of biomolecular simulations in drug discovery (*1*–*5*). Initial applications to biomolecular systems such as, for example, the calculations of hydration free energies (*6*, *7*) and of binding free energies go back to the 1980s (*8*–*13*). Early calculations were generally burdened by the limited sampling from short simulations, though the promise of free-energy methods was immediately recognized. Considerable progress has been made since then, due to advances in theoretical formulations, progress with simulation algorithms, and also increased availability of powerful computers. Importantly, the theoretical framework for carrying out various free energy computations has been greatly clarified (*4*).

Free energy simulations are arguably the most powerful and attractive approaches to estimate the binding free energy of ligands to macromolecules, which determine the thermodynamics of life. While relative and absolute free energy of binding calculations remain too demanding computationally for screening extremely large databases of compounds, recent successes in prospective lead optimization and the increase in computational power suggest that these methods are going to play an increasingly important role in drug discovery as computational resources continue to expand and the price of computational resources falls (*14*).

Democratization of free-energy calculations has been accompanied by the emergence of a host of algorithms that have contributed to improve not only the reliability, but also the efficiency of the methodology, with the unfortunate consequence of breeding confusion, and leaving the neophyte and the expert alike puzzled by how seemingly similar approaches could lead to significantly different results when compared, and which methodology is best to apply in their specific case of interest. In spite of carrying very distinct names, these methods are often conceptually related, and rest on a handful of basic ideas, which can be traced back to such trailblazers of the field as De Donder (*15*), Peierls (*16*), Landau (*17*, *18*), Kirkwood (*19*), Zwanzig (*20*), or Valleau (*21*, *22*). Still, addressing the question of the best method for a given problem remains eminently relevant, and ought to be rephrased in terms of the most cost-effective algorithm, or combination thereof, to obtain a reliable answer. From an applicative standpoint, free-energy calculations can be dichotomized in terms of geometrical and alchemical transformations. Whereas the former act directly on the spatial coordinates of the chemical objects at play to modify their positional, orientational, and conformational states, the latter exploit the malleability of the potential energy function to

interconvert between chemically distinct states. Beyond this rather coarse distinction, from a methodological perspective, geometrical and alchemical transformations can be carried out employing a variety of numerical schemes, which, in a broad sense, can be categorized in four classes, namely methods that rest upon (i) histograms and counting statistics (*21*–*23*), (ii) free-energy perturbation (FEP) (*17*, *20*), (iii) gradients and thermodynamic integration (TI) (*2*, *19*), and (iv) non-equilibrium ensemble (NEE) (*24*, *25*). The common denominator of these four classes of methods is the exploration of the configurational space of a reference state such that the relevant low-energy configurations of the target state are appropriately sampled, which constitutes the prerequisite for the accurate determination of a free-energy difference.

Since neither brute-force molecular dynamics, nor Monte Carlo, are particularly well suited to the above task of sampling all relevant energy minima, a number of strategies aimed at performing non-Boltzmann sampling have materialized at the dawn of the seventies. It is interesting to note that the groundwork for this methodology had been laid a long time before the advent of the computational era. Such is the case of the concept of an order, or generalized-extent parameter (thermodynamic coupling parameter lambda), introduced by Kirkwood in his derivation of integral equations for liquid-state theory (*19*), which reconciles statistical mechanics and the early notion of degree of evolution of a chemical reaction by De Donder (*15*). This concept was utilized by Valleau and Card to connect the reference and the target states (*23*), the low-energy regions which only marginally overlap, thereby establishing the foundation of a stratification strategy broadly utilized to break the total free-energy difference into a sum of finite free-energy differences between intermediate states for which the overlap is significantly greater. Pursuing the objective of sampling the reference state adequately to extract valuable information about the low-energy configurations of the target state, Torrie and Valleau put forth the idea of a non-Boltzmann weighting function, which could be introduced in the simulation, and ultimately removed to supply an unbiased probability distribution, provided that sufficient overlap in sampling between adjacent states that reside on steeply increasing (or steeply decreasing) parts of the (free) energy surface is achieved. This strategy forms the basis of the popular umbrella sampling algorithm (*21*, *22*), employed to determine potentials of mean force (PMFs) (*26*), and pertains to the first, histogram-based class of methods. Another approach based on histograms and counting statistics is λ-dynamics, whereby the probability of different alchemical states is determined from the dynamical evolution of a generalized coupling parameter (*27*–*29*). The development of tools such as the Weighted Histogram Analysis Method (WHAM) (*30*, *31*) has greatly contributed to the overall robustness of this class of methods.

Historically, the second class of methods, which rests on free-energy perturbation theory (*17*, *20*), has often been associated to alchemical transformations (*7*), even though gradient-based schemes, like thermodynamic integration (*19*), constitute a relevant alternative to obtain reliable free-energy differences between chemically distinct states. This implicit association of a method and an application can be understood in terms of the target state being the result of a chemical perturbation of the reference one, and rationalizes the prevalent use of "free-energy perturbation" to describe chemical alterations in molecular objects, as would be done experimentally, for instance, in site-directed mutagenesis. In practice, however, the absence of compelling evidence that one class of methods is computationally superior to the other suggests that either can potentially be employed to transform between chemical states (*5*, *13*, *32*–*36*).

A related class of methods, leaning on the knowledge of the free-energy gradient with respect to some coupling parameter, can be employed interchangeably for both geometrical and alchemical transformations. Whereas plain thermodynamic integration has been applied successfully in both

instances, with a general-extent parameter of a geometric or of a chemical nature, its many heirs have been used in the context of geometrical transformations (*37–40*),, (e.g. chemical reactions and conformational changes) and the mapping of free-energy landscapes, which offer valuable information to understand how chemical objects recognize and associate. In fact, one might argue that such free-energy landscapes, and their one-dimensional form, or PMF, provide a more satisfactory picture of recognition and association phenomena than alchemical transformations because they allow the molecules at play to be followed as they come towards each other and bind, even though the general-extent parameter utilized bears a certain arbitrariness, and, in general, does not necessarily mirror a physically meaningful pathway. Gradient-based methods, like the adaptive biasing force family of algorithms (*40*, *41*), have been employed fruitfully to determine within chemical accuracy protein-ligand (*42*, *43*) and protein-protein binding (*44*) affinities, following a purely geometric (configurational) route consisting of a series of PMF calculations.

All these methods, which assume an equilibrium sampling of the system, are built upon the reversible work theorem that relates free energy differences with the log of histogram ratios (*45*). In contrast, one last class of methods are based on the remarkable relationship put forth by Jarzynski (*24*) connecting an equilibrium free-energy difference and a suitable set of irreversible transformations performed between a reference and a target state of the system, i.e., time-dependent perturbations that are switched on rapidly via nonequilibrium simulations. Combined with steered molecular dynamics carried along an arbitrary direction of Cartesian space (*46*), application of the Jarzynski identity supplies in principle the one-dimensional free-energy, or PMF, change along this direction (*47*). In practice, however, an overly high pulling-speed regime is prone to hamper convergence of the Jarzynski identity due to an insufficient number of near-equilibrium realizations, thereby questioning the advantage of this route over equilibrium geometrical free-energy calculations.

Applications of free-energy methods in real chemical systems began to emerge in the 80s in conjunction with molecular dynamics (MD) or Monte Carlo (MC) sampling. The first study of absolute free energy calculations to study the free energy of cavity formation in water was performed in 1982 (*6*). The first study of relative free energy calculations was published by Jorgensen in 1985 (*7*), who computed the difference in the free energy of hydration between methanol and ethane, which compared favorably to the experimental result. Soon after, the first calculation of relative binding free energies (RBFE) in a biomolecular system followed by McCammon and co-workers, who applied FEP/MD to predict the difference in the free energy of binding of Cl^- and Br^- anions to the macrotricyclic receptor SC24 in water in excellent agreement with experimental data (*48*). In a series of follow-up studies, the first substrate-protein relative free energies of binding were predicted using FEP for hydrated trypsin, thermolysin, and subtilisin by the McCammon and Kollman groups, again in good agreement with experimental results (*49–51*). The predictive value of FEP became apparent when the RBFE of three thermolysin inhibitors was predicted prospectively by Merz and Kollman (*52*). The value of calculating free energy profiles across a coupling parameter was not only promising in host-guest systems but also for the investigation of chemical reactions. The first example of a chemical reaction studied with enhanced sampling calculations and stratification strategies was the investigation of the $Cl^- + CH3Cl$ *SN*2 reaction in both the gas phase and aqueous solution (*53*), which laid the foundations for quantum mechanical/molecular mechanical (QM/MM) calculations together with earlier seminal works (*54*).

These first applications set a precedent and showcased the feasibility and applicability of using free energy methods to predict fundamental chemical quantities such as binding or equilibrium

constants, solubilities, partition coefficients, relative pKa values, adsorption coefficients and others. For more examples and applications and a more detailed discussion the reader is referred to reviews in e.g. Refs. (*1*, *3*, *55–67*)

Methods

The reliability of the sampling algorithm used in a free energy calculation, such as for example MD simulations, is limited by the accuracy of the force field, the degree to which relevant phase space can be sampled, the degree to which the microscopic system set up *in silico* represents the macroscopic system *in vitro* or *in vivo*, and to a lesser extent the fundamental physical approximations inherent in classical non-relativistic treatment of the potential energy surface and equations of motion (*14*, *68–70*). In some cases, it is of interest to study the mechanisms associated with a chemical process such as a catalytic chemical reaction, in which case the transformation between states needs to occur along a physical pathway characterized by a free energy landscape parameterized by a set of appropriate coordinates. In other cases such as in computer-aided drug design applications, only the difference between thermodynamic states is desired, which may be much easier to compute using non-physical transformations.

In this book, free energy differences are discussed for processes involving similar but distinct molecules, where thermodynamic cycles can be constructed where pathways involve transformations from one molecule into another. These types of transformations are designated "alchemical", where changes occur between defined thermodynamic states of the system via non-physical pathways. Specifically, we focus on applications of free energy simulations within the context of the lead optimization stage of drug discovery (*14*, *71–75*). The direct simulation of the free energy of the physical process of ligand binding to a protein target is extremely difficult as it involves diffusion, desolvation, and conformational changes of both ligand and protein. However, the alchemical change between two similar ligands in solution and in a complex with the protein is often comparatively much simpler and more amenable to calculation. In particular, the calculation of RBFEs of sufficiently similar ligands can be routinely achieved to reasonable precision with current state-of-the-art methods, although significant challenges still remain.

A major goal in drug discovery is to make predictions about the binding thermodynamics, and in some cases kinetics, in order to guide the synthesis and further testing of compounds in lead refinement. Alchemical free energy simulations enable the prediction of absolute and/or relative binding free energies (*14*, *71–75*). Recent progress and improvements in computer hardware, simulation software, and free energy methods (*75–81*), including GPU-acceleration (*72*, *82–87*), have opened new doors by extending the accessible time scales of computer simulations and scope of applications.

As discussed earlier in the introduction, the change in free energy between two thermodynamic states can be rigorously formulated from equilibrium simulations using perturbation theory (*20*) or gradient-based methods such as thermodynamic integration (TI) (*2*, *19*), or through non-equilibrium ensemble (NEE) simulations using the Jarzynski equality and its equation variations (*24*, *25*, *88–90*). Both the TI and NEE approaches require formulation of a transformation pathway between states in order to connect them. Only the free energy perturbation (FEP) approach *formally* requires performing simulations only of one or both of the thermodynamic end states in order to obtain the free energy differences between states. This aspect has made this method appealing as a method for performing so-called "book-ending" end point corrections between molecular mechanical (MM) and quantum mechanical (QM) models so as to reduce or eliminate the need

for simulations that require the use of computationally intensive QM models (*88*, *91*, *92*). Further, the FEP approach has been used as the foundation for approximate end-point methods such as the Linear Interaction Energy (LIE) approach (LIE) (*93*).

All of the above classes of methods are critically sensitive to "phase space overlap": the degree to which the required averages from different phase space distributions can be reliably computed from numerical simulations (*88*, *92*, *94–96*). In practice, this implies that even for the class of FEP approaches, it is necessary to break up the transformation into multiple small steps for all but the most conservative transformations (e.g., such as book-ending approaches that use special "reference" potentials designed to optimize the phase space overlap with the QM end state). In this way, for drug discovery applications in practice, all the classes of methods above require defining a transformation pathway between states.

The free-energy methods outlined so far have been devised to improve sampling, while remaining consistent with a Boltzmann equilibrium distribution. Because they encourage sampling in regions that contribute substantially to the free energy, and that are only seldom visited by Boltzmann sampling, they are often referred to as importance-sampling methods. This terminology is often confused with enhanced-sampling strategies, which, by fostering ergodic sampling, go beyond the free-energy methods mentioned herein (*97*, *98*). One of the remarkable properties of the algorithms utilized in free-energy calculations is that they can be combined seamlessly to effectively enhance sampling. A class of algorithms, often referred to as Hamiltonian tempering (*99*), improves configurational sampling through a modification of the underlying potential energy function. This class of algorithms includes accelerated MD (*100*) (aMD) and its Gaussian variant (*101*) (GaMD), which can be associated to geometric free-energy calculations, using, for instance, an adaptive biasing force. To enhance conformational sampling, the microscopic isothermal MD propagator utilized to evolve the atomic configurations is replaced periodically by a different Hamiltonian, like aMD, by means of non-equilibrium switches, and acceptance of the proposed moves is determined by a Metropolis–Hastings criterion (*102*). A symmetric two-end momentum reversal ensures that the algorithm obeys microscopic detailed balance, and supplies the expected equilibrium Boltzmann distribution (*103*, *104*). It is also possible to enhance the sampling via hybrid nonequilibrium MD guided by a coarse-grained model (*105*). Another Hamiltonian-tempering algorithm, known as replica-exchange solute tempering (*106*) (REST2), and also the alchemical enhanced sampling (ACES) method discussed in the present book by Lee et al., and often used in conjunction with FEP calculations, enhances sampling by scaling the interaction of the solute with its environment to lower the barriers separating conformational states. Assuming proper post-hoc reweighting of the trajectories, or generation of the latter within a replica-exchange scheme, Hamiltonian-tempering algorithms can generate Boltzmann-distributed configurations. Closely related to replica-exchange strategies (*107*, *108*), multiple-walker schemes address sampling nonuniformity, which, most commonly, arises from poor timescale separation, and misrepresentation of the reaction coordinate by means of a naive set of collective variables. Under these premises, a number of walkers are spawned to populate different regions of the free-energy landscape, exchanging periodically information about the latter, like the locally measured gradients in multiple-walker adaptive biasing force (MW-ABF) scheme (*109*). This approach may be brought to a higher level of sophistication through Darwinian selection, cloning good walkers that cover large stretches of the transition path, while eliminating less efficient, kinetically trapped ones (*110*). Similar strategies, leaning on multiple-copy algorithms available in a number of MD engines, include, but are not limited to, multi-canonical temperature and Hamiltonian tempering replica-exchange MD (*111–113*) (REMD), replica-exchange umbrella sampling (*108*), and bias-exchange umbrella sampling (*114*) (BEUS). The basic

idea of replica-exchange schemes, swapping either temperatures or windows, is to make high-temperature—or high-coupling-parameter—configurations available to the ensemble of states generated at low temperature—or at low coupling parameter. This powerful strategy may, however, prove insufficient under certain circumstances, whereby in the course of an alchemical transformation conformational states of a solute remain marginally sampled due to insuperable barriers. Under these circumstances, it might be beneficial to preface the alchemical transformation by a geometrical free-energy calculation, and use the resulting PMF as a boost potential (*113*) to populate adequately all conformational states as the solute is coupled reversibly to its environment. An alternate road towards enhanced sampling consists in associating importance-sampling algorithms, such as metadynamics (*115*) and an extended-Lagrangian version of ABF (*116*, *117*) (eABF), exploiting the aggressive exploration of the free-energy landscape by the former, and the accurate determination of the local gradients by the latter. This seamless combination of two popular algorithms, referred to as meta-eABF (*118*, *119*), allows a coarser stratification scheme to be employed, and maps complex free-energy landscapes between 3 to 6 times faster than eABF alone, with virtually no computational overhead. Other schemes for enhanced sampling calculations include blue moon sampling (*120*), conformational flooding (*121*), weighted ensemble path sampling (WESTPA) (*122*), and local elevation (*123*).

In the sections that follow, we will summarize recent examples of the current state of the art, and describe Challenges and Limitations.

Examples of State of the Art Studies

A growing body of evidence supports that free-energy methods have reached a state of maturity where they can be successfully applied to real-world applications and impact the progression of drug discovery projects. However, as readers of this book are likely aware, there is a publication bias where positive results are more frequently presented than negative results (*124*). Experts in the field have undoubtedly experienced the pains and challenges associated with obtaining accurate and reliable free-energy simulations. Days, weeks, or even months might be needed to develop a protocol that has the desired reliability and accuracy for a given system of interest. Similar to developing an experimental assay for a new target, a great deal of care and expertise is needed when developing a free-energy simulation protocol for a new system. With free-energy simulations, an incorrect parameter or misplaced atom can render the simulation results meaningless, with valuable time, money, and computational resources wasted, not to mention the impact on the environment caused by the massive power consumption of modern computational resources. Nonetheless, with careful attention to necessary simulation elements and expertise in the field, it is generally possible to develop robust *in silico* binding assays using the free-energy methods described above and throughout the chapters of this book. The numerous published retrospective and prospective success examples, some of which are highlighted below, lend credibility to value of free-energy simulations in drug discovery. That being said, the inexperienced scientist might find it uncomfortable, frustrating, or even unapproachable to embark into the free-energy simulation world, where a "one size fits all" or "push-button" approach is still far from being a reality. Careful assessment of the system to be studied, careful preparation of the free-energy simulation protocol, monitoring of the simulations, thorough post-processing of the results and assessment of convergence all constitute critical elements to reliable free-energy calculations such as the recent applications presented below.

Recent Retrospective Applications

G protein-coupled receptors (GPCRs) are the most important, yet also one of the most challenging class of drug targets for modeling, owing to their plasticity, allosterism, dynamic configurational ensemble with side-chain orientations, backbone rearrangements, and dynamic water positioning and displacement, thus making accurate binding free energy calculations more challenging than for globular proteins. In a combined work, Sosei Heptares and Janssen used free-energy simulations to predict the RBFE for ligands of two different GPCRs, highlighting key elements for the successful outcome of the calculations such as the water molecule positioning, determining amino acid ionization states, equilibrating the system with known ligands, and enhanced sampling (*125*). The team showed that they could not yield meaningful results by following a standard FEP protocol, but results could be improved in a systematic way by tailoring their protocol to the specific system. This example illustrates the importance of protocol development by experts. Due to the rigorous thermodynamic nature of free-energy methods, it is generally possible to systematically improve simulation results by identifying problems, which typically relate to 1) incorrect system setup, 2) poor force field, or 3) insufficient sampling and reach sufficient accuracies to have a positive impact on decision making in drug discovery projects.

Large and flexible ligands also challenge the applicability of free-energy simulations for ligand optimization, where protocols were originally developed for less flexible small molecules. In one example, academic researchers partnering with an Astrazeneca team reported on the structure-based design of peptide macro-cycles targeting the protein binding site of human adaptor protein 14-3-3 (*126*). The researchers observed large activity cliffs, where large changes in experimental binding affinities were observed for relatively small variations in the chemical substituent size. Free-energy simulations were performed to rationalize observed trends. To account for insufficient convergence of the large, flexible ligands, restrained calculations were performed and complemented with extensive enhanced sampling of the ligands. These calculations revealed that changes in affinity originate both from altered direct protein-ligand interactions as well as conformational changes of the free ligand in solution. The study predicted a specific interaction that was important to explain some of the steep SAR, which was verified for one of the high affinity ligands by X-ray crystallography (*126*).

Cyclic nucleotide phosphodiesterases (PDEs) are metalloenzymes that play a key role in regulating the signaling molecules cAMP and cGMP. Metal atom interactions with organic and biomolecules present another challenge in free energy simulations, namely, poor force field treatment of the directional valence orbitals of metal atoms. PDE's share a highly conserved catalytic site (about 50%), thus presenting significant challenges to the design of selective drug candidates with classical structure-based design approaches. In one retrospective study, researchers applied free-energy simulations to predict the selectivity of inhibitors that bind two pairs of closely related PDE families: PDE9/1 and PDE5/6, where only one co-crystallized structure per pair was publicly available at the time (*127*). The authors demonstrated that free-energy simulations of homology models constructed for these metalloenzymes could accurately reproduce experimentally observed selectivity profiles. Moreover, using free energy calculations they could determine the protonation state of pKa of studied inhibitors. Based on these data, the authors could show that free-energy simulations are capable of producing robust predictions of affinity and selectivity for this challenging system, although significant time, expertise, and novel method development was needed to obtain the published level of accuracy.

Recent Prospective Applications

Industrial and academic groups are embracing relative and absolute binding free energy calculations (RBFE and ABFE), as evidenced by a number of prospective applications. A robust *in silico* binding assay allows drug discovery teams to explore a larger swath of chemical space more rapidly than can be done with experiments alone, which should facilitate more rapid progression of projects to the clinic. Indeed, there is a growing body of literature demonstrating the value of free-energy simulations in prospective drug discovery applications. Below, a small number of recent examples is presented. While there is no intention to be exhaustive, the below selection represents a diverse set of targets and challenges that should provide the reader with a general sense of the opportunities that accurate and reliable free-energy simulations can provide within the context of drug discovery projects.

In one example, researchers studied HIV-1 reverse transcriptase (HIV-RT), which is the enzyme responsible for converting the HIV RNA genome into DNA, an essential step in retroviral replication. HIV-RT possesses a unique and highly selective hydrophobic allosteric pocket in which many diverse allosteric inhibitors called non-nucleoside reverse transcriptase inhibitors (NNRTIs) have been designed, six of them now being FDA-approved. In a computer-aided NNRTI design journey that nearly lasted 20 years (*128*), lead optimization with free-energy simulations led to the discovery of NNRTIs, where they optimized micro-molar binders to picomolar leads. The optimized molecules include the most potent NNRTIs reported to date (*129–134*), including compounds with picomolar activity for the Y181C and Y181C/K103N HIV-RT mutant proteins (*135–137*).

In another study, a series of triazolopyrimidines binding to the embryonic ectoderm development (EED) subunit of the Polycomb repressive complex 2 (PRC2) complex were explored using free-energy simulations to inform compound designs for potential anticancer therapeutics (*138*). The authors performed a large number of free-energy simulations to rapidly evaluate structural modifications in a previously unexplored region of the EED binding site, which resulted in a series of novel triazolopyrimidine EED ligands with improved physicochemical properties and which inhibit PRC2 methyltransferase activity in a cancer-relevant G401 cell line.

In a large-scale assessment of binding free energy calculations in active drug discovery projects, researchers at Merck KGaA started a large initiative in 2016 to prospectively assess the prediction accuracy of RBFE calculations. In this study, 12 targets and 23 chemical series were tested performing over 35,000 individual perturbation calculations with the subsequent synthesis and testing of 400 blindly predicted and novel molecules (*139*). The authors conclude that the pre-requisites for a successful RBFE calculation is the choice of an accurate small molecule and protein force field, a high quality crystal structure, a known binding mode, known IC50 data with 3 log unit range, and the initial validation of the RBFE method/protocol/software to be used in the study by retrieving a root mean squared error (RMSE) of <1.3 kcal/mol for the known data before embarking in prospective lead optimization using RBFE methods. Moreover, the method is still limited in the type of transformations that can converge to the desired accuracy, although these transformations were of high interest to the project teams (e.g., from aromatic ring systems to aliphatic chains, charge changes, addition of new groups via flexible linkers). The team also found that, contrary to what was expected, RBFE calculations had more impact in their hit-to-lead campaigns rather than lead or fragment optimization.

Using a computationally empowered workflow, a team from Bayer identified novel covalent allosteric binders for the KRAS G12C isoform (*140*). Due to its frequent mutations in multiple lethal cancers, KRAS is one of the most-studied anticancer targets nowadays. Since the discovery of the

druggable allosteric binding site containing a G12C mutation, KRASG12C has been the focus of attention in oncology research. The free-energy simulation workflow involved initial enumeration of virtual molecules tailored for the KRAS allosteric binding site, and pharmacophore modeling, docking, and finally free-energy simulations were used to prioritize the compounds with the best profiles. The synthesized naphthyridinone scaffold showed the ability to react with G12C and inhibit KRASG12C. Analogues were prepared to establish structure-activity relationships, while molecular dynamics simulations and crystallization of the inhibitor-KRASG12C complex highlighted an unprecedented binding mode (*141*).

More recently, free-energy simulations were also employed in the discovery of new inhibitors of the main protease (Mpro) of SARS-CoV-2, the virus responsible for COVID-19 (*142*). Starting from 14 known drugs as inhibitors of the main protease (Mpro), the weak hit FDA-approved drug perampanel was redesigned to yield multiple noncovalent, nonpeptidic inhibitors with an IC50 of 20 nM. FEP calculations for Mproligand complexes provided valuable guidance on beneficial modifications that rapidly delivered the potent analogues. The design efforts were confirmed by high-resolution X-ray crystal structures for five analogues bound to Mpro. Results of cell-based antiviral assays further demonstrated the potential of the compounds for treatment of COVID-19. In addition to the possible therapeutic significance, the work demonstrates the power of computational chemistry for drug discovery, especially free-energy-guided lead optimization, to rapidly progress weak hits to potent leads.

In one of the largest free-energy simulation studies on a GPCR, a collection of 3,4-dihydropyrimidin-2(1H)-ones was enumerated and computationally screened with free-energy simulations against the A2B adenosine receptor (A2BAR) (*143*). Many of the compounds were found to bind selectively to A2BAR, with a number of potent and selective antagonists further confirmed by functional cyclic adenosine monophosphate experiments, and an accurate model of the structure-activity relationship of this chemotype was obtained. The assessment of the effect in representative cytochromes (CYP3A4 and CYP2D6) demonstrated insignificant inhibitory activity, while *in vitro* experiments in three prostate cancer cells demonstrated that this pair of compounds exhibits a pronounced antimetastatic effect.

In another study of a viral protein, the RNA-dependent RNA polymerase (RdRp) of norovirus was studied as a target of antiviral agents aimed at providing protection against norovirus-associated gastroenteritis (*144*). Free-energy simulations were performed on the crystal structure of norovirus RdRp in complex with several known binders to determine binding free energies of these molecules relative to the natural nucleotide substrates. Using free-energy simulations, a virtual nucleotide library containing 121 molecules was screened and two novel molecules were successfully identified with *in vitro* activity.

As a final prospective example presented here, a team of researchers at Janssen recently described the hit-to-lead exploration of a [1,2,4]triazolo[1,5-a]pyrimidine phosphodiesterase 2A (PDE2A) inhibitor arising from high-throughput screening (*145*). Starting from a co-crystal structure of the target in complex with a lead compound, rounds of lead optimization using relative binding free-energy simulations helped prioritize chemically-diverse substituents to explore the chemical space. FEP calculations were performed for 265 putative PDE2A inhibitors, and 100 compounds were synthesized and tested in binding assays, representing a relatively large prospective application. The screen provided unexpectedly active molecules with IC50 values of some molecules achieving sub-nanomolar potency. One of the top compounds prioritized from the free-energy simulations showed a PDE2A IC50 of 1.3 +/-0.39 nM, 100-fold selectivity versus other PDE enzymes, a clean

cytochrome P450 profile, *in vivo* target occupancy, and promise for further lead optimization. More examples can be found in (*146*, *147*).

These recent examples only represent a small sample of the many retrospective and prospective applications of free energy methods in drug discovery. They demonstrate that free-energy calculations can be applied to active drug discovery projects and are being increasingly adopted by the pharmaceutical industry and the academic community. However, significant challenges still remain and expertise is required to address many 'hidden' caveats of the target and lead series when performing free-energy calculations.

Challenges and Limitations

As explained above, there are many possible methodological variants to compute binding free energies within a rigorous thermodynamic framework. Indeed, with proper system setup, sufficient sampling, and an accurate force field all of the above approaches should produce accurate binding free energies, although the rate of convergence (and therefore simulation time) will depend on the details of the method. However, even with a rigorous thermodynamic framework, the setup and analysis of free energy methods is elaborate and requires a high degree of technical expertise, especially in challenging cases. There are a number of important factors beyond force field and sampling that must be carefully considered in order to consistently compute accurate binding free energies. Below, we touch on a number of the most common challenges encountered in drug discovery applications of free-energy simulations. While this list is not exhaustive, it is intended to give a flavor of the complexities involved and the importance of the details. In short, every atom matters in free-energy simulations and therefore it is prudent to take special care throughout the entire process (preparing, running, and analyzing results). Excellent recent discussions on best practices for alchemical free energy calculations cover many of the topics below in detail (*75*, *148*).

Force Field

Having a proper model of the Born-Oppenheimer energy surface is paramount to the accuracy of binding free energy calculations (and molecular simulations in general). Without an accurate molecular mechanical force field, even the most sophisticated free-energy approaches will converge to the wrong answer. Though not entirely free of approximations, a rigorous quantum mechanical (QM) treatment could ultimately be the desired approach to model the energetics of biological systems, but it is currently impractical to treat an entire protein-ligand complex in explicit water with quantum mechanics. As such, the potential energy of the atoms in the system is typically approximated via a classical molecular mechanics force field that is parameterized based on quantum mechanics and experimental data. The force field representation allows for much faster energy evaluations while still representing the underlying physics in an approximate manner. There are many ways in which a force field can be generated to achieve accurate energetics at a fraction of the time of a QM calculation. However, all approximations will come at a cost in terms of accuracy relative to quantum mechanical calculations. The calculated free energies integrate contributions from the protein and ligand force field as well as the solvent model, all of which are prone to introducing errors (*149*). The primary interest in drug discovery efforts is striking the right balance between computational throughput and accuracy. It is beyond the scope of this introductory chapter to cover force fields extensively, but a few of the most significant challenges and limitations will be discussed.

For ligands, there are two common ways to obtain force field parameters: 1) Lookup tables based on similar chemical moieties that have previously been parameterized or 2) bespoke parameterization

based on quantum mechanics. In some cases a quantum mechanical/molecular mechanical (QM/MM) or quantum mechanical force field (QMFF) approach (*150*) can be used for the ligand and ligand-protein interaction, although at considerable computational expense relative to more conventional MM force fields. One attractive procedure for including these types of QM interactions in alchemical free energy simulations is to use a so-called "book-ending" or "reference potential" approach that only requires consideration of free energy corrections at the real state endpoints, and not along the entire alchemical transformation pathway (*88*, *91*, *92*).

For the solvent model, usually water is represented explicitly in relative and absolute free energy calculations, although the presence of explicit water molecules is particularly computationally demanding. Recently implicit solvent models have been also employed in alchemical free-eenrgy calculations (*151*). The single-decoupling method (SDM) (*152*), for example, is based on progressively turning on the effective interaction between the ligand and the receptor with an implicit representation of the solvent. Replica-exchange FEP simulations with a GB continuum model (REFEP-GB) have also been employed to combine FEP with a GB implicit solvent model to calculate relative binding free energies (*153*), which can significantly speed the calculation time reaching convergence in less than 1 ns per lambda step reaching reasonable accuracy compared to experimental data.

A particularly promising strategy for improvement of molecular simulation force fields is to use machine learning (ML) or deep learning potentials (*154–170*). ML-based potentials afford a promising solution to the development of next-generation molecular simulation force fields with the efficiency comparable to that of MM force fields, and accuracy that has started to approach that of high-level QM methods (*171*). More recently, ML-based forces fields have generated encouraging results for energy calculations of small molecules (*155*), but applications to free energy simulations have not been published at the time of this writing. Most force fields rely on fixed charge models that do not treat polarization explicitly. Ideally, a force field would adapt to the environment (e.g. solvent, protein, lipid) to account for polarizability, as would be the case with a quantum mechanical potential. While polarizable force fields do exist, such as AMOEBA (*172*), applications to binding free-energy simulations have been limited, likely due to the challenges associated with the parameterization process. Interestingly, the most commonly used force fields use the same general functional form, which has not been updated in decades (*173–177*), although other functional forms have been presented (*178*, *179*). As such, the commonly used force fields for free-energy simulations still perform poorly in situations with close contacts that cannot be explained with a simple Lennard-Jones potential and fixed charges. Recent augmentations of force fields have added charges off of atom centers in order to more accurately reproduce non-spherical charge distributions, such as those observed with lone pairs (*180*), halogen bonds (*181*), and other sigma hole interactions (*182*). Another significant limitation of most modern force fields is the reliance on atom types, which have been constructed to simplify the process of assigning bonded and non-bonded parameters to atoms that should have similar properties, but this generalization creates a lack of precision on the specific parameters applied to each atom. Recent efforts by the Open Force Field Consortium have overcome the atom typing limitation, although results have not yet been shown to improve compared to force fields with traditional atom typing procedures (*183*).

Importantly, any framework developed to address the above problems should be automated in such a way that it can be run robustly in industrial applications where hundreds of ligands need to be processed in a single iteration of hit-to-lead or lead optimization. Finally, building a "good enough" force field for free-energy simulations in drug discovery will likely need to be done iteratively using accurate binding free energy data, which is still sparse. The protein and water force field parameters

are also essential in relative binding free energy calculations to accurately model protein motions. While high-quality protein force fields are broadly available due to the limited chemical space (20 amino acids and one water molecule) and the large number of groups working on them. Nonetheless, there are still inaccuracies in the protein and water force fields and multiple groups continue working on force field improvements.

Optimization of the Alchemical Transformation Pathways

The free energy is a state function, and thus the free energy difference between states is formally independent of pathway. However, the practical computation of the free energy is in fact very sensitive to the pathway that is chosen, and on the degree of phase space overlap along the pathway. There has been much effort devoted to the choice of pathways that are the most amenable to stable, precise computation. Some of the most commonly encountered issues that can arise in alchemical transformations that limit the overall reliability and statistical precision of free energy estimates include: 1) the end-point catastrophe, 2) the particle collapse, and 3) the large gradient-jump problems. Each of these problems arise from issues related to phase space overlap along the alchemical transformation pathway that can result in instabilities or "phase-change-like" behavior making sampling and/or thermodynamic averaging extremely difficult. One of the methods to help mitigate these problems and facilitate stable alchemical transformations is to "soften" certain types of interatomic interactions in the potential energy function as certain chemical moieties are created and others are annihilated as one ligand transforms into another. These so-called "softcore" potentials have been instrumental in addressing the stability of alchemical transformations, and are an ongoing topic of research. The transformation pathways can be further enhanced through developing more advanced λ-scheduling (*184*); i.e., the specific schedule of how individual energy terms of transformed as a function of the λ transformation progress variable. Examples of outstanding challenges in alchemical transformations include those that involve changes in charge state and scaffold/core hopping (*185*, *186*).

Equilibration

Equilibration protocols are employed after system preparation and before the free energy simulations. The objective of the equilibration is to have the system adequately relaxed such that the production simulation is stable. Systems that have not been sufficiently equilibrated in binding free-energy simulations can lead to large errors. For RBFE simulations, the equilibration can be challenging, since ligands with diverse functional groups might clash with the initial protein structure. In such systems that are not carefully equilibrated, the protein can locally unfold or the ligand can unbind due to the large initial forces, leading to suboptimal results. It is therefore important to sufficiently equilibrate to provide the best starting point for the free-energy simulation. Standard equilibration protocols have been developed for most MD engines, but free-energy simulations offer an additional challenge, since each protein-ligand system could have different levels of clashes and therefore require different equilibration approaches. While tedious, this step is extremely important for accurate free-energy results.

Slowly Exchanging Buried/Trapped Waters

In addition to sampling protein and ligand molecules, many biological systems contain buried waters that cannot freely exchange with bulk solvent during the time course of free-energy simulation (~5 ns). Introducing ligand modifications into these regions of buried waters require changes in the

number of buried water molecules in order to accurately predict binding energies. Neglect of these buried waters will result in inaccurate structures and energies. Approaches to handle this critical challenge include Grand Canonical Monte Carlo (GCMC) (*187*, *188*) and mixed Monte Carlo Molecular Dynamics (MCMD) (*189*), which have both been shown to improve the convergence and accuracy of free-energy simulations. It should be noted that 5 ns, as mentioned above, is a sampling time per lambda used in a typical RBFE calculation. With this sampling time, one ligand perturbation can be calculated per day per GPU (*65*), providing the appropriate throughput in a large scale project (*139*). In challenging cases, the sampling time per lambda-window may be extended in order to obtain convergence provided that time and computing resources allow it (*66*).

Quality of Initial Structure

The starting protein structure (from experiments or modeling) has to be of sufficiently high quality and no major conformational changes should be expected during the simulations as large-scale protein movements cannot be usually sampled within the timeframe of FEP calculations (*65*, *139*, *190*). Moreover, to predict free-energy changes upon ligand binding, accurate bound poses are necessary. If no experimental structures for the protein-ligand complex exist, several strategies are reported to predict the bound pose in advance of free-energy calculation such as generalized replica exchange with solute tempering (gREST) and FEP, cross docking and calculating interaction fingerprints and others (*190–192*).

Treatment of Protein Flexibility

One of the great strengths of free-energy simulations using molecular dynamics is the ability to capture conformational changes, making it possible to treat biologically relevant motions underlying biomolecular recognition. However, most of the current implementations of RBFE in drug discovery run a few nanoseconds of MD simulation per lambda window, which limits the scope of conformational changes that can be assessed with high confidence. One of the great challenges of RBFE simulations is understanding the conformational free energy landscape associated with the target of interest and developing bespoke workflows based on that landscape. When high-energy barriers separate relevant states, biased simulations like umbrella sampling can be used (*193*). The trade off between computational time and simulation accuracy represents one of the fundamental challenges in free-energy simulations in drug discovery. Assess convergence is particularly challenging with high-energy barriers, since statistical methods cannot infer what has not been observed in the simulations. As such, target-specific knowledge is often required to choose the right enhanced sampling method and associated degrees of freedom for the collective variables of interest (*113*). Understanding what enhanced sampling method to apply for a given problem requires expertise in the sampling method as well.

Movement of even a single side chain can significantly influence RBFE results, as seen in a study of a series of pyrazine PDE2 inhibitors (*194*). The single underpredicted compound in the series was lacking a large side chain at a specific position that was filled by all other ligands in the series. As a result, the compound induced a conformational change in Leu770 that filled the vacant pocket left from the smaller ligand substituents, as verified by experimental X-ray structures. While it was previously known that different ligands can induce either an open or closed conformation based on the rotamer of Leu770, the conformational transition took approximately 30 ns of MD simulation. While the appropriate conformational transition was observed with a longer simulations, without a priori knowledge of the two states it might not have been known that additional simulation time was

needed. T4 lysosyme L99A offers another case where a single side-chain rotamer can influence the binding free energy predictions, as described by Jiang and Roux (*113*).

Backbone flexibility is also important to consider. For example, a fragment optimization study of Janus Kinase 2 (JAK2) shows that even for very small molecules, protein backbone conformational changes can influence results (*195*). Two protein X-ray crystal structures, each with different small molecules bound, were available for this study. Both receptor structures were independently used as inputs and produced comparable results (R^2-values of 0.8 between the results from each structure). However, the large ligands in site with the smaller co-crystallized ligand required constraints in the docking calculations to ensure that reasonable poses were obtained. In a similar example from the same paper (this time with the protein target HSP90), two X-ray crystal structures with ligands of different size were available. Comparison of the two structures revealed a helix partially unwinding with a subset of ligand substituents, which substantially changed the shape of the binding pocket. This time, only the receptor structure with the large co-crystallized ligand was used, since all members of the ligand series could be reliably placed in that receptor binding pocket but many ligands were too large to fit into the structure that was co-crystallized with the smaller ligand.

Finally, from that same work, the authors showed fragments binding to p38α MAP kinase. Again, there were multiple receptor cocrystal structures available and the protein exhibited backbone conformational changes in the binding pocket. In this case, the authors removed a subset of ligands that could not be accurately docked. While the binding free energy results were accurate for the subset of ligands used, it highlights some limitations associated with standard RBFE protocols where significant conformational changes might be needed to accommodate the ligands in the series. As seen in these examples, even moderate protein backbone conformational changes can introduce challenges into RBFE predictions, which in these cases stemmed primarily from the inability to generate viable input poses from docking. In one case, the authors were able to overcome the challenge by imposing restraints on the ligands and in another case a single viable receptor structure was used. However, in the final case no suitable solution was found and a subset of the ligands had to be removed from the data set. Based on these examples, it can be seen that more robust ways of treating protein backbone flexibility in RBFE workflows are needed. In the meantime, users of RBFE calculations should carefully examine each system on a case-by-case basis and apply the appropriate system setup, simulation protocol, and analysis steps to overcome the specific challenges at hand. It is not trivial to detect possible protein conformational changes *a priori* and can be even harder to implement robust solutions. Fortunately, free energy methods are built on the foundations of molecular physics and statistical mechanics, so solutions should exist for all problems, although they may require significant human time and expertise to uncover.

Treatment of Ligands

Ligand Alignment and Atom Mapping

In addition to the issues common to molecular dynamics simulations (as described above), there are additional steps specifically associated with the setup of alchemical simulations. Incorrect treatment at this stage can result in slower convergence at best, and more commonly wrong results. This becomes particularly challenging and important when automating a large number of ligand perturbations. As such, care should be taken to appropriately prepare the ligands. Several reviews cover the details associated with setting up alchemical free energy simulations (*65*, *139*, *148*, *196*, *197*), which will be briefly highlighted below.

For RBFE simulations, the common part of the perturbation pairs should be aligned (typically in Cartesian space) to minimize the thermodynamic path length between endpoints in order to improve convergence and overlap between adjacent lambda windows. This step can be done via a variety of methods, including docking with core constraints, shape-based alignment, pharmacophore overlap, or some form of manual overlay. The need for highly accurate alignments to facilitate atom mapping, especially when the precise alignment of atoms in unknown, makes this a challenge, especially for complex drug-like molecules where a one-to-one mapping is not always obvious. Manually inspection of the results from automated alignment protocols is recommended, since each approach has inherent limitations that depend on the ligand series being explored. Additionally, given that most RBFE pose placement tools do not account for protein flexibility, it may be necessary to perform conformational sampling and/or energy minimization of the binding site (or whole protein) to check whether clashes can be alleviated by maintaining the ligand and receptor structures in the final pose (*190*). In addition, as discussed above, a careful equilibration protocol should then be applied to get the ligand, protein, and waters into a stable state with minimal energetic clashes while maintaining the appropriate ligand alignments. An important consideration is the degree of movement that is acceptable during the equilibration, since sufficient energy overlap between adjacent lambda windows typically requires a degree of structural overlap.

Even with high quality ligand alignments (each pair of matching atoms within 0.5 Å), atom mapping can still present challenges and ambiguous solutions. The most commonly used approach to atom mapping involves maximum common substructure (MCS) search with atomic graphs regardless of atom type and bond valences, thereby maximizing their topological overlap. Even in cases where two molecules do not differ in the number or connectivity of atoms, there still might be challenges in atom mapping because the ligand modifications may engage in different interactions which then induce different conformations in the binding site. When there is a difference in topology of the two molecules of interest (i.e. there is not a sufficient MCS between the molecules), atoms not present in the atomic graph of one molecule can be introduced as dummy atoms to match the topology of the other molecule.

Connecting Perturbations through Graphs

Modifications are typically introduced based on an initial lead ligand or set of ligands. Toward that end, a network, or thermodynamic "graph" is created where the "edges" represent thermodynamic transformations between individual ligands that form the nodes (intersection of the edges) of the graph. Redundant pathways are often included such that the network contains free energy cycles. While this increased the computational costs (more ligand perturbations), it can improve the accuracy and reliability of RBFE calculations, since each closed thermodynamic cycle can be used as a convergence test and also to correct outliers. In addition, the graphs may contain edges that have already been measured experimentally to improve the overall predictions (*198*). There exist a number of methods, some of which are formally equivalent, to estimate the free energy of a single edge, such as the Bennett acceptance ratio method (BAR) (*199*), multistate BAR (MBAR) (*200*), and the unbinned WHAM (UWHAM) method (*201*, *202*), as well as variational methods for their solution (*203*). More recently, a variational approach for network-wide analysis of free energy simulations has been introduced to include experimental or high-precision reference data (*204*).

Multiple Binding Poses

Ligand placement may result in multiple plausible poses that cannot be unambiguously distinguished. In such case it is possible to perform RBFE calculations on each of the poses. If the conformations interconvert during the RBFE simulations, then the same computed $\Delta\Delta G$ will be generated for all of the input ligand conformations. However, if the different poses are separated by high-energy barriers then they will not interconvert during the simulations and must be treated specially. The nature of the energy barrier and whether it is associated with both the bound and unbound states or just the bound state will dictate the best approach. One option is to perform independent free-energy simulations on each of the binding modes, taking the free energy of the more favorable of the ligand binding modes (or using a Boltzmann-weighted energy) (*65*, *205*). The ligand should sampling the different conformations in the unbound state in order for the energies to be comparable. Additionally, the overall entropic contribution from having multiple poses can be computed, which can be as much as 0.6 kcal/mol (kT ln(2)) when two states are isoenergetic, and larger if more than two such states exist (*206*).

Tautomers and Ionization States

Exchanges of hydrogen atoms (tautomerism and ionization) involve making and breaking of bonds, which is not handled with traditional molecular mechanics force fields. Properly treating these states therefore requires special treatment, which is critical because each of the tautomers and protomers generally have different energetic preferences in solution and in the binding site. Indeed, it has been estimated that 25% of drug-like molecules have multiple energetically accessible tautomers (*207*). There are a number of methods that can account for dynamic sampling of protonation and/or tautomerization during MD simulations, but the reference value of the different states is generally not known, thus accurate estimation of the free energy involving changes in protonation state remains highly challenging. While tautomers play a critical role in binding, tools to accurately predict tautomer energies are not common. A recent publication of a tautomer database may provide an opportunity to improve tautomer prediction accuracy (*208*). Within the current paradigm of molecular mechanics force fields, whenever multiple tautomers exist, each should be considered as a distinct ligand, with the tatuomeric energy being included in the final free energy value, similar to the above discussion about multiple binding modes. Changes in protonation states can be even more challenging for RBFE calculations, since it is involves molecules with different charges (see next paragraph). Methods like constant-pH molecular dynamics can be used in cases of tautomer and ionization states (*209*, *210*), which would remove the need for pre-generating the relevant tautomers, although validation of this approach in the context of free-energy simulations has been limited and still requires knowledge of a reference value for the pKa or tautomer energy.

Charged Mutations

Computing binding affinity differences for molecules with different formal charges remains a challenge for free-energy simulations. Favorable results have been published (*211*), although other studies have suggested otherwise (*212*). There are technical challenges associated with introducing or removing a charge during free-energy calculations using the particle-mesh Ewald (PME) approach for long-range electrostatics (*213–215*). Additionally, charged groups produce much larger contributions to the total energy, which can add noise to the calculations when the charge is not conserved across the ligands. Even errors in the solvation step (perturbing ligand A to B in solution)

could result in many kcal/mol of error for charge changes, which might be compounded in the bound state. Finally, more computational time is needed for the system to relax around different formal charges due to the long-range nature of the electrostatic forces. Additionally, the overlap between adjacent lambda windows will likely be lower than for charge-preserving changes. Encouraging works have demonstrated success in ligand and protein residue mutations with different formal charges (*216–218*), although there have not been additional publications using this approach. As such, when molecules with different formal charges are of interest, it is advised that at least one reference molecule for each formal charge state be synthesized and assayed to provide a reference energy for subsequent RBFE simulations of congeneric molecules with the same formal charge.

Covalently-Bound and Metal-Bound Ligands

The commonly used molecular mechanics force fields do not account for the formation or breaking of covalent bonds. However, binding free-energy simulations can still be used in such situations by analyzing only the non-covalent interactions, assuming that the energetics of the covalent bond is constant throughout the ligand series. In this same way, the relative affinity of metal-bound ligands can be computed if the contribution coming from the formation of the ligand-metal bond is constant across the ligands in the series, although such assumptions must be made with caution. Quantum mechanical calculations can be performed to assess the geometries and energetics of the different ligands. When substitutions are far from the covalent/metal interaction site and the geometry of the ligands are preserved throughout the series, then this approximation is more likely to be valid. Performing covalent or metal-binding RBFE simulations requires a thermodynamic cycle unique from a traditional non-covalent binding calculation, as has been described (*219*, *220*) and validated in a successful prospective application (*221*). When the energetics of the covalent or metal interaction are suspected to be different across the ligands of interest, it should be possible to perform a separate quantum mechanical calculation to assess the energy of the covalent linkage, which can then be used to augment the RBFE calculations (*222*), although these calculations are challenging and generally require a high level of expertise in the application of quantum mechanics to biological systems.

Convergence

Convergence of free-energy simulations depends on the nature of the free energy surface and specifically the height of the energetic barriers separating biologically relevant states. In general, large-scale protein conformational changes cannot be sampled sufficiently within the time frame of standard free-energy simulations (1-10 ns per lambda). Even small conformational changes, such as the 180-degree flip of a ligand ring in a binding site or amide bond, might have sufficiently high free-energy barrier to not exchange during the simulations. Improving convergence by running simulations longer is often the simplest approach, but not practical for most cases due to the vast increase needed to overcome many barriers (*223*). For example, extending a simulation by a factor of ten would only allow it to cross energetic barriers 1.4 kcal/mol higher than at the shorter simulation. As such, overcoming a barrier that is 5 kcal/mol higher than those being crossed during a certain simulation time would require more than 1,000 longer simulations. Despite rapid increases in available computing power, extending MD simulation times is likely not the best approach to improve reliability of free energy calculations. Instead, enhanced sampling methods have been developed to improve convergence of slow degrees of freedom, although expertise and system knowledge is generally needed to control which degrees of freedom to apply the enhanced sampling.

For example, a method that works well in one case might cause the protein to denature in another situation, as a result of differences in free energy surfaces. There are many enhanced sampling approaches, which are covered through many excellent reviews and associated primary sources (*224*–*227*).

Analysis

Post-simulation analysis is an important step that should be taken before making decisions based on the free energy results. Careful analysis can identify problems with the simulations and/or build confidence in predictions before making final recommendations for chemical synthesis. Traditional tools for MD analysis, such as computing energy fluctuations and atomic movements can be informative, although it is more complex in alchemical free energy simulations due to the multiple lambda windows. Uncertainties in free energies can be calculated with the block bootstrapping method, where the data are divided into blocks of observations and then sampled randomly with replacement to compute uncertainties that arise from the finite nature of the simulation. However, this approach only assesses uncertainties within the conformational ensemble that has been sampled and cannot account for conformational states that have been missed. Running multiple independent simulations from different random seeds and/or starting configurations is a more robust way to estimate sampling errors, although it takes more computational resources and still does not guarantee that all relevant conformational states have been sampled. Another way to detect possible convergence problems is by performing the same perturbation in the reverse direction (i.e. preparing the system with ligand B and perturbing to ligand A rather than the reverse) (*228*).

Workflows and Automation

In addition to methodology development, the integration of methods is critical to the success of real-world free energy calculations (*43*, *64*, *64*, *71*, *97*, *228*–*235*). Furthermore, there are many tools to help with post-simulation analysis (*200*, *202*, *203*, *234*, *236*–*240*). Workflows are an important part of large-scale applications of alchmical binding free energy simulations in drug discovery projects (*139*) to help to ensure best practices (*148*) and enable high throughput with reduced human effort and error (*241*–*246*). In this way, scientific advances can be more broadly exposed to academic, government and industry labs to guide the design and facilitate discovery of new therapeutics. Examples of such efforts include commercially available automated FEP/TI pipelines such as the FEP+ module of Schrödinger Suite (*72*), the Molecular Operating Environment (MOE) offered by the Chemical Computing Group (CCG) (*247*, *248*) providing a pipeline for performing relative binding free energy calculations using the AMBER TI platform (*75*), BIOVIA Discovery Studio and Pipeline Pilot (*249*) offering a workflow for setting up and implementing relative FEP calculations on GPUs using CHARMM (*28*), QuantumBio's movable type method (*250*), and FEP in Flare of the Cresset software (*245*) that uses open-source tools such as AMBER tools (*251*), OpenMM (*83*), LOMAP (*252*), Sire (*253*), and BioSimSpace (*254*). Non-commercial workflows include the Free Energy Workflow tool (*241*) available for AMBER (*251*, *255*), QligFEP (*256*) for the Q open-source MD package (*257*), FESetup (*242*) for AMBER and GROMACS (*258*), BioSimSpace (*254*) that automates the creation of the required input files for the implementation of relative binding free energy simulations using SOMD (*253*) or GROMACS (*259*), BFEE2 (*246*) for absolute free energy calculations generating configurational files in NAMD (*184*) and GROMACS formats, and three

web-servers, the "Alchemical Assistant" of the LigParGen server (*260*), FEPrepare (*261*), and the CHARMM-GUI FEP calculator (*262*).

Data Sharing and Open Science

In real world drug design projects, structural information and availability of ligand series with large affinity spread are not always available to the scientific community. Open, findable, accessible, interoperable, and reusable public data help to ensure the reproducibility of published results can support further method development, comparative benchmarking of free energy calculations, and advance the field as a whole. For example, in machine learning and artificial intelligence there have been significant improvements in the field due to data sharing and open science challenges. Open Science represents a new approach to the scientific process based on cooperative work and new ways of diffusing knowledge by using digital technologies and new collaborative tools. With Open Science we can make the primary outputs of publicly funded research results – publications and the research data – publicly accessible in digital format with no or minimal restriction. As such, Open Science is about extending the principles of openness to the whole research cycle, fostering sharing and collaboration as early as possible thus entailing a systemic change to the way science and research is done. Open Science will be crucial in providing datasets and tools freely to the scientific community and the general public, which will enable cooperative work for the maximum output of harnessing big data to develop more efficient computer-aided drug design methodologies. In this respect, several initiatives have been launched that foster the concept of open data in global drug discovery efforts. The drug design data resource (D3R) organized challenges against blinded experimental data to prospectively test computational methodologies as an opportunity for improved methods and algorithms to emerge (*263*, *264*). SAMPL (Statistical Assessment of the Modeling of Proteins and Ligands) is another set of community-wide blind challenges aimed to advance computational techniques as standard predictive tools in rational drug design (*264*, *265*). Both of these challenges have been based on blinded data such as binding affinity and hydration free energy data, that later became openly available to the community to test the latest modeling methods and force fields.

The need to produce open, reproducible scientific output has led to discussions for the standardization of molecular simulation file formats, streamlining molecular simulation data, best practices for sharing data and workflows used to produce and analyze molecular simulations, code sharing as well as strategies to enhance the reproducibility of such data (*258*, *266–269*). This trend has been accelerated by the unprecedented pandemic crisis, which brought together the molecular modeling community in a rapid manner to collaborate in a global and timely fashion using open research practices, sharing of research outputs, data and code, thereby facilitating research and research reproducibility and timely collaboration beyond borders (*270*, *271*).

Conclusions

The chapters of this book present the latest advances in free-energy methods and applications in the context of drug discovery applications from a diverse set of academic and industry researchers. The many successes of free-energy methods applied in drug discovery projects have sparked the interest of applying similar protocols beyond affinity calculations. Although *in vivo* biology is not only dictated by ligand-protein binding and operates under other processes such as kinetics, allostery, and biological phenomena operate beyond thermodynamic equilibrium, the concept of calculating free energy changes within a thermodynamic cycle can be still be applied not only to ligand-protein

binding but also to predicting solubility, protein-protein binding, protein stability, crystal packing preference and other properties that are crucial in the drug discovery and development process. At the moment, free-energy simulations still require experts to obtain accurate and reliable predictions that can be confidently used to make decisions in drug discovery campaigns. While challenges remain for free-energy methods in facilitating drug design, tremendous progress continues to be made, and it is hoped that the combined efforts of leading research teams in the field such as those presented in the chapters that follow will forge the next generation of methods that enable breakthroughs in drug discovery.

References

1. Beveridge, D. L.; DiCapua, F. M. Free Energy via Molecular Simulation: Applications to Chemical and Biomolecular Systems. *Annu. Rev. Biophys. Biomol. Struct.* **1989**, *18*, 431–492.
2. Straatsma, T. P.; McCammon, J. A. Computational Alchemy. *Annu. Rev. Phys. Chem.* **1992**, *43*, 407–435.
3. Kollman, P. Free Energy Calculations: Applications to Chemical and Biochemical Phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
4. Simonson, T.; Archontis, G.; Karplus, M. Free Energy Simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.
5. Jorgensen, W. L. Free Energy Calculations: a Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
6. Postma, J. P. M.; Berendsen, H. J. C.; Haak, J. R. Thermodynamics of Cavity Formation in Water—A Molecular-Dynamics Study. *Faraday Symp. Chem. Soc.* **1982**, *17*, 55–67.
7. Jorgensen, W. L.; Ravimohan, C. Monte-Carlo Simulation of Differences in Free-Energies of Hydration. *J. Chem. Phys.* **1985**, *83*, 3050–3054.
8. Warshel, A. Dynamics of Reactions in Polar-Solvents -Semi-Classical Trajectory Studies of Electron-Transfer and Proton-Transfer Reactions. *J. Phys. Chem.* **1982**, *86*, 2218–2224.
9. Tember, B. L.; McCammon, J. A. Ligand Receptor Interactions. *Comp. Chem.* **1984**, *8*, 281–283.
10. Warshel, A. Simulating the Energetics and Dynamics of Enzymatic Reactions. *Pontif. Acad. Sci. Scr. Varia* **1982**, *55*, 59–81.
11. Lybrand, T. P.; McCammon, J. A.; Wipff, G. Theoretical Calculation of Relative Binding-Affinity in Host Guest Systems. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 833–835.
12. Wong, C. F.; McCammon, J. A. Dynamics and Design of Enzymes and Inhibitors. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
13. Hermans, J.; Subramaniam, S. The Free Energy of Xenon Binding to Myoglobin from Molecular Dynamics Simulation. *Isr. J. Chem.* **1986**, *27*, 225–227.
14. Cournia, Z.; Allen, B. K.; Beuming, T.; Pearlman, D. A.; Radak, B. K.; Sherman, W. Rigorous Free Energy Simulations in Virtual Screening. *J. Chem. Inf. Model.* **2020**, *60*, 4153–4169.
15. De Donder, T. *L'affinité*; Gauthier-Villars: Paris, 1927.
16. Peierls, R. E. On the Theory of the Diamagnetism of Conduction Electrons. *Z. Phys.* **1933**, *80*, 763–791.
17. Landau, L. D. *Statistical Physics*; Clarendon: Oxford, 1938.

18. Landau, L. D.; Lifshitz, E. M. *Course of Theoretical Physics, Vol. 5: Statistical Physics*; Pergamon Press: Oxford, 1958.
19. Kirkwood, J. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.* **1935**, *3*, 300–313.
20. Zwanzig, R. W. High Temperature Equation of State by a Perturbation Method. *J. Chem. Phys.* **1954**, *22*, 1420–1426.
21. Torrie, G. M.; Valleau, J. P. Monte Carlo Free Energy Estimates Using Non-Boltzmann Sampling: Application to the Sub-Critical Lennard-Jones Fluid. *Chem. Phys. Lett.* **1974**, *28*, 578–581.
22. Torrie, G. M.; Valleau, J. P. Nonphysical Sampling Distributions in Monte Carlo Free-Energy Estimation: Umbrella Sampling. *J. Comp. Phys.* **1977**, *23*, 187–199.
23. Valleau, J. P.; Card, D. N. Monte Carlo Estimation of the Free Energy by Multistage Sampling. *J. Chem. Phys.* **1972**, *57*, 5457–5462.
24. Jarzynski, C. Nonequilibrium Equality for Free Energy Differences. *Phys. Rev. Lett.* **1997**, *78*, 2690–2693.
25. Crooks, G. E. Path-Ensemble Averages in Systems driven far from Equilibrium. *Phys. Rev. E* **2000**, *61*, 2361–2366.
26. Roux, B. The Calculation of the Potential of Mean Force Using Computer Simulations. *Comput. Phys. Comm.* **1995**, *91*, 275–282.
27. Knight, J. L.; Brooks, C. L. Lambda-dynamics free energy simulation methods. *J. Comput. Chem.* **2009**, *30*, 1692–1700.
28. Raman, E. P.; Paul, T. J.; Hayes, R. L.; Brooks, C. L. Automated, Accurate, and Scalable Relative Protein-Ligand Binding Free-Energy Calculations Using Lambda Dynamics. *J. Chem. Theory Comput.* **2020**, *16*, 7895–7914.
29. Vilseck, J. Z.; Armacost, K. A.; Hayes, R. L.; Goh, G. B.; Brooks, C. L., III. Predicting Binding Free Energies in a Large Combinatorial Chemical Space Using Multisite Dynamics. *J. Phys. Chem. Lett.* **2018**, *9*, 3328–3332.
30. Ferrenberg, A. M.; Swendsen, R. H. New Monte Carlo Technique for Studying Phase Transitions. *Phys. Rev. Lett.* **1989**, *63*, 1195–1198.
31. Kumar, S.; Bouzida, D.; Swendsen, R.; Kollman, P.; Rosenberg, J. The Weighted Histogram Analysis Method for free-energy calculations on biomolecules. I. The method. *J. Comp. Chem.* **1992**, *13*, 1011–1021.
32. Roux, B.; Nina, M.; Pomès, R.; Smith, J. C. Thermodynamic Stability of Water Molecules in the Bacteriorhodopsin Proton Channel: A Molecular Dynamics Free Energy Perturbation Study. *Biophys. J.* **1996**, *71*, 670–681.
33. Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The Statistical-Thermaldynamic Basis for Computation of Binding Affinities: A Critical Review. *Biophys. J.* **1997**, *72*, 1047–1069.
34. Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute Binding Free Energies: A Quantitative Approach for Their Calculation. *J. Phys. Chem. B* **2003**, *107*, 9535–9551.
35. Deng, Y.; Roux, B. Calculation of Standard Binding Free Energies: Aromatic Molecules in the T4 Lysozyme L99A Mutant. *J. Chem. Theory Comput.* **2006**, *2*, 1255–1273.
36. Wang, J.; Deng, Y.; Roux, B. Absolute Binding Free Energy Calculations Using Molecular Dynamics Simulations with Restraint Potentials. *Biophys. J.* **2006**, *91*, 2798–2814.

37. Carter, E.; Ciccotti, G.; Hynes, J. T.; Kapral, R. Constrained Reaction Coordinate Dynamics for the Simulation of Rare Events. *Chem. Phys. Lett.* **1989**, *156*, 472–477.
38. den Otter, W. K.; Briels, W. J. The Calculation of Free-Energy Differences by Constrained Molecular Dynamics Simulations. *J. Chem. Phys.* **1998**, *109*, 4139–4146.
39. den Otter, W. K. Thermodynamic Integration of the Free Energy along a Reaction Coordinate in Cartesian Coordinates. *J. Chem. Phys.* **2000**, *112*, 7283–7292.
40. Darve, E.; Pohorille, A. Calculating Free Energies Using Average Force. *J. Chem. Phys.* **2001**, *115*, 9169–9183.
41. Comer, J.; Gumbart, J. C.; Hénin, J.; Lelièvre, T.; Pohorille, A.; Chipot, C. The Adaptive Biasing Force Method: Everything you Always Wanted to Know, but were Afraid to Ask. *J. Phys. Chem. B* **2015**, *119*, 1129–1151.
42. Woo, H.; Roux, B. Calculation of Absolute Protein-Ligand Binding Free Energy from Computer Simulations. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6825–6830.
43. Gumbart, J. C.; Roux, B.; Chipot, C. Standard Binding Free Energies from Computer Simulations: What Is the Best Strategy? *J. Chem. Theory Comput.* **2013**, *9*, 794–802.
44. Gumbart, J. C.; Roux, B.; Chipot, C. Efficient Determination of Protein-Protein Standard Binding Free Energies from First Principles. *J. Chem. Theory Comput.* **2013**, *9*, 3789–3798.
45. Chandler, D. *Introduction to modern statistical mechanics*; Oxford University Press, 1987.
46. Izrailev, S.; Stepaniants, S.; Isralewitz, B.; Kosztin, D.; Lu, H.; Molnar, F.; Wriggers, W.; Schulten, K. Steered Molecular Dynamics. In *Computational Molecular Dynamics: Challenges, methods, ideas*; Deuflhard, P., Hermans, J., Leimkuhler, B., Mark, A. E., Skeel, R., Reich, S. , Eds.; Lecture Notes in Computational Science and Engineering; Springer Verlag: Berlin, 1998; Vol. 4, pp 39–65.
47. Park, S.; Khalili-Araghi, F.; Tajkhorshid, E.; Schulten, K. Free Energy Calculation from Steered Molecular Dynamics Simulations Using Jarzynski's Equality. *J. Chem. Phys.* **2003**, *119*, 3559–3566.
48. Lybrand, T.; McCammon, A.; Wipff, G. Theoretical calculation of relative binding-affinity in host guest systems. *Proc. Natl. Acad. Sci. USA* **1986**, *83*, 833–835.
49. Wong, C. F.; McCammon, J. A. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *J. Am. Chem. Soc.* **1986**, *108*, 3830–3832.
50. Bash, P. A.; Singh, C.; Brown, K.; Langridge, R.; Kollman, A. Dynamics and design of enzymes and inhibitors. *Science* **1987**, *235*, 574–6.
51. Rao, N.; Singh, C.; Bash, A.; Kollman, A. Dynamics and design of enzymes and inhibitors. *Nature* **1987**, *328*, 551–554.
52. Merz, K. M., Jr.; Kollman, A. P. Free Energy Perturbation Simulations of the Inhibition of Thermolysin: Prediction of the Free Energy of Binding of a New Inhibitor. *J. Am. Chem. Soc.* **1989**, *111*, 5649–5658.
53. Chandrasekhar, J.; Smith, S. F.; Jorgensen, W. L. Theoretical Examination of the SN2 Reaction Involving Chloride Ion and Methyl Chloride in the Gas Phase and Aqueous Solution. *J. Am. Chem. Soc.* **1985**, *107*, 154–163.
54. Warshel, A.; Levitt, M. Theoretical studies of enzymic reactions: Dielectric, electrostatic and steric stabilization of the carbonium ion in the reaction of lysozyme. *J. Mol. Biol.* **1976**, *103*, 227–249.

55. Jorgensen, W. L. Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution. *Acc. Chem. Res.* **1989**, *22*, 184–189.
56. van Gunsteren, W. F.; Daura, X.; Mark, A. E. Computation of Free Energy. *Helv. Chim. Acta* **2002**, *85*, 3113–3129.
57. Chipot, C.; Pohorille, A. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer: New York, 2007.
58. Shirts, M. R.; Mobley, D. L.; Chodera, J. D. Alchemical Free Energy Calculations: Ready for Prime Time? *Annu. Rep. Comput. Chem.* **2007**, *3*, 41–59.
59. Jorgensen, W. L.; Thomas, L. L. Perspective on free energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869–876.
60. Deng, Y. Q.; Roux, B. Computations of Standard Binding Free Energies with Molecular Dynamics Simulations. *J. Phys. Chem. B* **2009**, *113*, 2234–2246.
61. Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569–1582.
62. Chodera, J. D.; Mobley, D. L.; Shirts, M. R.; Dixon, R. W.; Branson, K.; Pande, V. S. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
63. Chipot, C. Frontiers in Free Energy Calculations of Biological Systems. *WIREs Comput. Mol. Sci.* **2014**, *4*, 71–89.
64. Hansen, N.; van Gunsteren, W. F. Practical Aspects of Free Energy Calculations: A Review. *J. Chem. Theory Comput.* **2014**, *10*, 2632–2647.
65. Cournia, Z.; Allen, B.; Sherman, W. Relative Binding Free Energy Calculations in Drug Discovery: Recent Advances and Practical Considerations. *J. Chem. Inf. Model.* **2017**, *57*, 2911–2937.
66. Rizzi, A.; Jensen, T.; Slochower, D. R.; Aldeghi, M.; Gapsys, V.; Ntekoumes, D.; Bosisio, S.; Papadourakis, M.; Henriksen, M.; de Groot, B. L.; Cournia, Z.; Dickson, A.; Michel, J.; Gilson, M. K.; Shirts, M. R.; Mobley, D. L.; Chodera, J. D. The SAMPL6 SAMPLing Challenge: Assessing the Reliability and Efficiency of Binding Free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 601–633.
67. Coskun, D. Computational approaches for protein pKa calculations. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 143–159.
68. van Gunsteren, W. F.; Berendsen, H. J. C. Computer Simulation of Molecular Dynamics: Methodology, Applications and Perspectives in Chemistry. *Agnew. Chem., Int. Ed. Engl.* **1990**, *29*, 992–1023.
69. Karplus, M.; McCammon, J. A. Molecular Dynamics Simulations of Biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–653.
70. Simonson, T.; Archontis, G.; Karplus, M. Free Energy Simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430–437.
71. Jorgensen, W. L. Efficient Drug Lead Discovery and Optimization. *Acc. Chem. Res.* **2009**, *42*, 724–733.

72. Abel, R.; Wang, L.; Harder, E. D.; Berne, B. J.; Friesner, R. A. Advancing Drug Discovery through Enhanced Free Energy Calculations. *Acc. Chem. Res.* **2017**, *50*, 1625–1632.
73. Armacost, K. A.; Riniker, S.; Cournia, Z. Novel Directions in Free Energy Methods and Applications. *J. Chem. Inf. Model.* **2020**, *60*, 1–5.
74. Song, L. F.; Merz, K. M., Jr. Evolution of Alchemical Free Energy Methods in Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5308–5318.
75. Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; McGee, T. D., Jr.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; Tsai, H.-C.; Xu, H.; Sherman, W.; York, D. M. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595–5623.
76. Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K. Accelerating Molecular Modeling Applications with Graphics Processors. *J. Comput. Chem.* **2007**, *28*, 2618–2640.
77. Anderson, J. A.; Lorenz, C. D.; Travesset, A. General Purpose Molecular Dynamics Simulations Fully Implemented on Graphics Processing Units. *J. Comput. Phys.* **2008**, *227*, 5342–5359.
78. Hardy, D. J.; Stone, J. E.; Schulten, K. Multilevel Summation of Electrostatic Potentials Using Graphics Processing Units. *Parallel Comput.* **2009**, *35*, 164–177.
79. Harvey, M. J.; De Fabritiis, G. An Implementation of the Smooth Particle Mesh Ewald Method on GPU Hardware. *J. Chem. Theory Comput.* **2009**, *5*, 2371–2377.
80. Göetz, A.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 1. Generalized Born. *J. Chem. Theory Comput.* **2012**, *8*, 1542.
81. Salomon-Ferrer, R.; Götz, A. W.; Poole, D.; Le Grand, S.; Walker, R. C. Routine Microsecond Molecular Dynamics Simulations with AMBER on GPUs. 2. Explicit Solvent Particle Mesh Ewald. *J. Chem. Theory Comput.* **2013**, *9*, 3878–3888.
82. Lee, T.-S.; Hu, Y.; Sherborne, B.; Guo, Z.; York, D. M. Toward Fast and Accurate Binding Affinity Prediction with pmemdGTI: An Efficient Implementation of GPU-Accelerated Thermodynamic Integration. *J. Chem. Theory Comput.* **2017**, *13*, 3077–3084.
83. Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T.; Zhao, Y.; Beauchamp, K. A.; Wang, L. P.; Simmonett, A. C.; Harrigan, M. P.; Stern, C. D.; Wiewiora, R. P.; Brooks, B. R.; Pande, V. S. OpenMM 7: Rapid Development of High Performance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
84. Mermelstein, D. J.; Lin, C.; Nelson, G.; Kretsch, R.; McCammon, J. A.; Walker, R. C. Fast and Flexible GPU Accelerated Binding Free Energy Calculations within the Amber Molecular Dynamics Package. *J. Comput. Chem.* **2018**, *39*, 1354–1358.
85. Giese, T. J.; York, D. M. A GPU-Accelerated Parameter Interpolation Thermodynamic Integration Free Energy Method. *J. Chem. Theory Comput.* **2018**, *14*, 1564–1582.
86. Lee, T.-S.; Cerutti, D. S.; Mermelstein, D.; Lin, C.; LeGrand, S.; Giese, T. J.; Roitberg, A.; Case, D. A.; Walker, R. C.; York, D. M. GPU-Accelerated Molecular Dynamics and Free Energy Methods in Amber18: Performance Enhancements and New Features. *J. Chem. Inf. Model.* **2018**, *58*, 2043–2050.

87. Song, L. F.; Lee, T.-S.; Zhu, C.; York, D. M.; Merz, K. M., Jr. Using AMBER18 for Relative Free Energy Calculations. *J. Chem. Inf. Model.* **2019**, *59*, 3128–3135.
88. Boresch, S.; Woodcock, H. L. Convergence of single-step Free Energy Perturbation. *Mol. Phys.* **2017**, *115*, 1200–1213.
89. Gapsys, V.; Michielssens, S.; Peters, J. H.; de Groot, B. L.; Leonov, H. Calculation of Binding Free Energies. *Methods Mol. Biol.* **2015**, *1215*, 173–209.
90. Jeong, D.; Andricioaei, I. Reconstructing Equilibrium Entropy and Enthalpy Profiles from Non-Equilibrium Pulling. *J. Chem. Phys.* **2013**, *138*, 114110.
91. Kearns, F. L.; Hudson, P. S.; Woodcock, H. L.; Boresch, S. Computing Converged Free Energy Differences between Levels of Theory via Nonequilibrium Work Methods: Challenges and Opportunities. *J. Comput. Chem.* **2017**, *38*, 1376–1388.
92. Giese, T. J.; York, D. M. Development of a Robust Indirect Approach for MM → QM Free Energy Calculations That Combines Force-Matched Reference Potential and Bennett's Acceptance Ratio Methods. *J. Chem. Theory Comput.* **2019**, *15*, 5543–5562.
93. Åqvist, J.; Luzhkov, V. B.; Brandsdal, B. O. Ligand Binding Affinities from MD Simulations. *Acc. Chem. Res.* **2002**, *35*, 358–365.
94. Lu, N.; Kofke, D. A. Accuracy of Free-Energy Perturbation Calculations in Molecular Simulation. II. Heuristics. *J. Chem. Phys.* **2001**, *115*, 6866–6875.
95. Wu, D.; Kofke, D. A. Phase-Space Overlap measures. II. Design and implementation of staging methods for free-energy calculations. *J. Chem. Phys.* **2005**, *123*, 84109.
96. Kofke, D. A. On the sampling requirements for exponential-work free-energy calculations. *Mol. Phys.* **2006**, *104*, 3701–3708.
97. Chipot, C., Pohorille, A., Eds. *Free Energy Calculations: Theory and Applications in Chemistry and Biology*; Springer Series in Chemical Physics; Springer: New York, 2007; Vol. 86.
98. Lelièvre, T.; Stoltz, G.; Rousset, M. *Free Energy Computations: A Mathematical Perspective*; Imperial College Press, 2010.
99. Swendsen, R. H.; Wang, J. S. Replica Monte Carlo Simulation of Spin Glasses. *Phys. Rev. Lett.* **1986**, *57*, 2607–2609.
100. Voter, A. F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Phys. Rev. Lett.* **1997**, *78*, 3908–3911.
101. Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2017**, *13*, 9–19.
102. Suh, D.; Radak, B. K.; Chipot, C.; Roux, B. Enhanced Configurational Sampling with Hybrid Non-Equilibrium Molecular Dynamics/Monte Carlo Propagator. *J. Chem. Phys.* **2018**, *148*, 014101.
103. Chen, Y.; Roux, B. Efficient Hybrid Non-Equilibrium Molecular Dynamics—Monte Carlo Simulations with Symmetric Momentum Reversal. *J. Chem. Phys.* **2014**, *141*, 114107.
104. Chen, Y.; Roux, B. Generalized Metropolis Acceptance Criterion for Hybrid Non-Equilibrium Molecular Dynamics—Monte Carlo Simulations. *J. Chem. Phys.* **2015**, *142*, 024101.
105. Chen, Y.; Roux, B. Enhanced Sampling of an Atomic Model with Hybrid Nonequilibrium Molecular Dynamics: Monte Carlo Simulations Guided by a Coarse-Grained Model. *J. Chem. Theory Comput.* **2015**, *11*, 3572–3583.

106. Wang, L.; Friesner, R. A.; Berne, B. J. Replica Exchange with Solute Scaling: A more Efficient Version of Replica Exchange with Solute Tempering (REST2). *J. Phys. Chem. B* **2011**, *115*, 9431–9438.
107. Sugita, Y.; Okamoto, Y. Replica-Exchange Molecular Dynamics Method for Protein Folding. *Chem. Phys. Lett.* **1999**, *314*, 141–151.
108. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional Replica-Exchange Method for Free-Energy Calculations. *J. Comput. Chem.* **2000**, *113*, 6042–6051.
109. Comer, J.; Phillips, J.; Schulten, K.; Chipot, C. Multiple-Replica Strategies for Free-Energy Calculations in NAMD: Multiple-Walker Adaptive Biasing Force and Walker Selection Rules. *J. Chem. Theory Comput.* **2014**, *10*, 5276–5285.
110. Minoukadeh, K.; Chipot, C.; Lelièvre, T. Potential of Mean Force Calculations: A Multiple-Walker Adaptive Biasing Force Approach. *J. Chem. Theory Comput.* **2010**, *6*, 1008–1017.
111. Woods, C. J.; Essex, J. W.; King, M. A. The Development of Replica-Exchange-Based Free-Energy Methods. *J. Phys. Chem. B* **2003**, *107*, 13703–13710.
112. Jiang, W.; Hodoscek, M.; Roux, B. Computation of Absolute Hydration and Binding Free Energy with Free Energy Perturbation Distributed Replica-Exchange Molecular Dynamics. *J. Chem. Theory Comput.* **2009**, *5*, 2583–2588.
113. Jiang, W.; Roux, B. Free Energy Perturbation Hamiltonian Replica-Exchange Molecular Dynamics (FEP/H-REMD) for Absolute Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2010**, *6*, 2559–2565.
114. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *J. Chem. Phys.* **2000**, *113*, 6042–6051.
115. Laio, A.; Parrinello, M. Escaping Free Energy Minima. *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 12562–12565.
116. Fu, H.; Shao, X.; Chipot, C.; Cai, W. Extended Adaptive Biasing Force Algorithm. An On-The-Fly Implementation for Accurate Free-Energy Calculations. *J. Chem. Theory Comput.* **2016**, *12*, 3506–3513.
117. Lesage, A.; Lelièvre, T.; Stoltz, G.; Hénin, J. Smoothed Biasing Forces Yield Unbiased Free Energies with the Extended-System Adaptive Biasing Force Method. *J. Phys. Chem. B* **2017**, *121*, 3676–3685.
118. Fu, H.; Zhang, H.; Chen, H.; Shao, X.; Chipot, C.; Cai, W. Zooming across the Free-Energy Landscape: Shaving Barriers, and Flooding Valleys. *J. Phys. Chem. Lett.* **2018**, *9*, 4738–4745.
119. Fu, H.; Shao, X.; Cai, W.; Chipot, C. Taming Rugged Free-Energy Landscapes Using an Average Force. *Acc. Chem. Res.* **2019**, *52*, 3254–3264.
120. Ciccotti, G.; Kapral, R.; Vanden-Eijnden, E. Blue Moon Sampling, Vectorial Reaction Coordinates, and Unbiased Constrained Dynamics. *ChemPhysChem* **2005**, *6*, 1809–1814.
121. Grubmüller, H. Predicting slow structural transitions in macromolecular systems: Conformational flooding. *Phys. Rev. E Sep.* **1995**, *52*, 2893–2906.
122. Zuckerman, D. M.; Chong, L. T. Weighted Ensemble Simulation: Review of Methodology, Applications, and Software. *Annu. Rev. Biophys.* **2017**, *46*, 43–57.
123. Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput.-Aided Mol. Des.* **1994**, *8*, 695–708.

124. Olson, C. M.; Rennie, D.; Cook, D.; Dickersin, K.; Flanagin, A.; Hogan, J. W.; Zhu, Q.; Reiling, J.; Pace, B. Publication Bias in Editorial Decision Making. *JAMA* **2002**, *287*, 2825–2828.
125. Deflorian, F.; Perez-Benito, L.; Lenselink, B.; Congreve, M.; van Vlijmen, H. W. T.; Mason, J. S.; de Graaf, C.; Tresadern, G. Accurate Prediction of GPCR Ligand Binding Affinity with Free Energy Perturbation. *J. Chem. Inf. Model.* **2020**, *60*, 5563–5579.
126. Wallraven, K.; Holmelin, F. L.; Glas, A.; Hennig, S.; Frolov, A. I.; Grossmann, T. N. Adapting Free Energy Perturbation Simulations for Large Macrocyclic Ligands: How to Dissect Contributions from Direct Binding and Free Ligand Flexibility. *Chem. Sci.* **2020**, *11*, 2269–2276.
127. Moraca, F.; Negri, A.; de Oliveira, C.; Abel, R. Application of Free Energy Perturbation (FEP+) to Understanding Ligand Selectivity: A Case Study to Assess Selectivity Between Pairs of Phosphodiesterases (PDE's). *J. Chem. Inf. Model.* **2019**, *59*, 2729–2740.
128. Jorgensen, W. L. Computer-Aided Discovery of Anti-HIV Agents. *Bioorg. Med. Chem.* **2016**, *24*, 4768–4778.
129. Lee, W. G.; Chan, A. H.; Spasov, K. A.; Anderson, K. S.; Jorgensen, W. L. Design, Conformation, and Crystallography of 2-Naphthyl Phenyl Ethers as Potent Anti-HIV Agents. *ACS Med. Chem. Lett.* **2016**, *7*, 1156–1160.
130. Chan, A. H.; Lee, W. G.; Spasov, K. A.; Cisneros, J. A.; Kudalkar, S. N.; Petrova, Z. O.; Buckingham, A. B.; Anderson, K. S.; Jorgensen, W. L. Covalent Inhibitors for Eradication of Drug-Resistant HIV-1 Reverse Transcriptase: From Design to Protein Crystallography. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, 9725–9730.
131. Kudalkar, S. N.; Beloor, J.; Quijano, E.; Spasov, K. A.; Lee, W. G.; Cisneros, J. A.; Saltzman, W. M.; Kumar, P.; Jorgensen, W. L.; Anderson, K. S. From In Silico Hit to Long-Acting Late-Stage Preclinical Candidate to Combat HIV-1 Infection. *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *115*, E802–E811.
132. Sasaki, T.; Gannam, Z. T. K.; Kudalkar, S. N.; Frey, K. M.; Lee, W. G.; Spasov, K. A.; Jorgensen, W. L.; Anderson, K. S. Molecular and Cellular Studies Evaluating a Potent 2-Cyanoindolizine Catechol Diether NNRTI Targeting Wildtype and Y181C Mutant HIV-1 Reverse Transcriptase. *Bioorg. Med. Chem. Lett.* **2019**, *29*, 2182–2188.
133. Dodda, L. S.; Tirado-Rives, J.; Jorgensen, W. L. Unbinding Dynamics of Non-Nucleoside Inhibitors from HIV-1 Reverse Transcriptase. *J. Phys. Chem. B.* **2019**, *123*, 1741–1748.
134. Duong, V. N.; Ippolito, J. A.; Chan, A. H.; Lee, W. G.; Spasov, K. A.; Jorgensen, W. L.; Anderson, K. S. Structural Investigation of 2-Naphthyl Phenyl Ether Inhibitors Bound to WT and Y181C Reverse Transcriptase Highlights Key Features of the NNRTI Binding Site. *Protein Sci.* **2020**, *29*, 1902–1910.
135. Frey, K. M.; Puleo, D. E.; Spasov, K. A.; Bollini, M.; Jorgensen, W. L.; Anderson, K. S. Structure-Based Evaluation of Non-Nucleoside Inhibitors with Improved Potency and Solubility that Target HIV Reverse Transcriptase Variants. *J. Med. Chem.* **2015**, *58*, 2737–2745.
136. Lee, W.-G.; Frey, K. M.; Gallardo-Macias, R.; Spasov, K. A.; Bollini, M.; Anderson, K. S.; Jorgensen, W. L. Picomolar Inhibitors of HIV-1 Reverse Transcriptase: Design and Crystallography of Naphthyl Phenyl Ethers. *ACS Med. Chem. Lett.* **2014**, *5*, 1259–1262.

137. Nichols, S. E.; Domaoal, R. A.; Thakur, V. V.; Tirado-Rives, J.; Anderson, K. S.; Jorgensen, W. L. Discovery of Wild-Type and Y181C Mutant Non-Nucleoside HIV-1 Eeverse Transcriptase Inhibitors Using Virtual Screening with Multiple Protein Structures. *J. Chem. Inf. Model.* **2009**, *49*, 1272–1279.
138. O'Donovan, D. H.; Gregson, C.; Packer, M. J.; Greenwood, R.; Pike, K. G.; Kawatkar, S.; Bloecher, A.; Robinson, J.; Read, J.; Code, E.; Hsu, J. H.-R.; Shen, M.; Woods, H.; Barton, P.; Fillery, S.; Williamson, B.; Rawlins, P. B.; Bagal, S. K. Free Energy Perturbation in the Design of EED Ligands as Inhibitors of Polycomb Repressive Complex 2 (PRC2) Methyltransferase. *Bioorg. Med. Chem. Lett.* **2016**, *39*, 127904.
139. Schindler, C. E. M.; Baumann, H.; Blum, A.; Böse, D.; Buchstaller, H.-P.; Burgdorf, L.; Cappel, D.; Chekler, E.; Czodrowski, P.; Dorsch, D.; Eguida, M. K. I.; Follows, B.; Fuchß, T.; Grädler, U.; Gunera, J.; Johnson, T.; Jorand Lebrun, C.; Karra, S.; Klein, M.; Knehans, T.; Koetzner, L.; Krier, M.; Leiendecker, M.; Leuthner, B.; Li, L.; Mochalkin, I.; Musil, D.; Neagu, C.; Rippmann, F.; Schiemann, K.; Schulz, R.; Steinbrecher, T.; Tanzer, E.-M.; Unzue Lopez, A.; Viacava Follis, A.; Wegener, A.; Kuhn, D. Large-Scale Assessment of Binding Free Energy Calculations in Active Drug Discovery Projects. *J. Chem. Inf. Model.* **2020**, *60*, 5457–5474.
140. Mortier, J.; Friberg, A.; Badock, V.; Moosmayer, D.; Schroeder, J.; Steigemann, P.; Siegel, F.; Gradl, S.; Bauser, M.; Hillig, R.; Briem, H.; Eis, K.; Bader, B.; Nguyen, D.; Christ, C. D. Computationally Empowered Workflow Identifies Novel Covalent Allosteric Binders for $KRAS^{G12C}$. *ChemMedChem* **2020**, *15*, 827–832.
141. Meier, K.; Bluck, J. P.; Christ, C. D. Use of Free Energy Methods in the Drug Discovery Industry. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 39–66.
142. Zhang, C.-H.; Stone, E. A.; Deshmukh, M.; Ippolito, J. A.; Ghahremanpour, M. M.; Tirado-Rives, J.; Spasov, K. A.; Zhang, S.; Takeo, Y.; Kudalkar, S. N.; Liang, Z.; Isaacs, F.; Lindenbach, B.; Miller, S. J.; Anderson, K. S.; Jorgensen, W. L. Potent Noncovalent Inhibitors of the Main Protease of SARS-CoV-2 from Molecular Sculpting of the Drug Perampanel Guided by Free Energy Perturbation Calculations. *ACS Cent. Sci.* **2021**, *7*, 467–475.
143. Majellaro, M.; Jespers, W.; Crespo, A.; Núñez, J.; Novio, S.; Azuaje, J.; Prieto-Díaz, R.; Gioé, C.; Alispahic, B.; Brea, J.; Loza, I.; Freire-Garabal, M.; Garcia-Santiago, C.; Rodríguez-García, C.; García-Mera, X.; Caamaño, O.; Fernandez-Masaguer, C.; Sardina, F.; Stefanachi, A.; El Maatougui, A.; Mallo-Abreu, A.; Åqvist, J.; Gutiérrez-de Terán, H.; Sotelo, E. 3,4-Dihydropyrimidin-2(1H)-ones as Antagonists of the Human A2B Adenosine Receptor: Optimization, Structure-Activity Relationship Studies, and Enantiospecific Recognition. *J. Med. Chem.* **2021**, *64*, 458–480.
144. Freedman, H.; Kundu, J.; Tchesnokov, P.; Law, L. M.; Nieman, J. A.; Schinazi, F.; Tyrrell, L.; Gotte, M.; Houghton, M. Application of Molecular Dynamics Simulations to the Design of Nucleotide Inhibitors Binding to Norovirus Polymerase. *J. Chem. Inf. Model.* **2020**, *60*, 458–480.
145. Tresadern, G.; Velter, I.; Trabanco, A.; Van den Keybus, F.; Macdonald, J.; Somers, V. F.; Vanhoof, G.; Leonard, M.; Lamers, M. A. C.; Van Roosbroeck, E. M.; Buijnsters, J. J. A.

[1,2,4]Triazolo[1,5-a]pyrimidine Phosphodiesterase 2A Inhibitors: Structure and Free-Energy Perturbation-Guided Exploration. *J. Med. Chem.* **2020**, *63*, 12887–12910.
146. Ozen, A.; Perola, E.; Brooijmans, N.; Kim, J. Prospective Application of Free Energy Methods in Drug Discovery Programs. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 127–141.
147. Knight, J. L.; Leswing, K.; Bos, P. H.; Wang, L. Impacting Drug Discovery Projects with Large-scale Enumerations, Machine Learning Strategies, and Free-energy Predictions. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 205–226.
148. Mey, A. S. J. S.; Allen, B. K.; Bruce McDonald, H. E.; Chodera, J. D.; Hahn, D. F.; Kuhn, M.; Michel, J.; Mobley, D. L.; Naden, L. N.; Prasad, S.; Rizzi, A.; Scheen, J.; Shirts, M. R.; Tresadern, G.; Xu, H. Best Practices for Alchemical Free Energy Calculations [Article v1.0]. *Living J. Comput. Mol. Sci.* **2020**, *2*, 18378.
149. Gapsys, V.; Pérez-Benito, L.; Aldeghi, M.; Seeliger, D.; van Vlijmen, H.; Tresadern, G.; de Groot, B. L. Large scale relative protein ligand binding affinities using non-equilibrium alchemy. *Chem. Sci.* **2020**, *11*, 1140–1152.
150. Giese, T. J.; York, D. M. Quantum Mechanical Force Fields for Condensed Phase Molecular Simulations. *J. Phys. Condens. Matter* **2017**, *29*, 383002.
151. Decherchi, S.; Masetti, M.; Vyalov, I.; Rocchia, W. Implicit solvent methods for free energy estimation. *Eur. J. Med. Chem.* **2015**, *91*, 27–42.
152. Gallicchio, E.; Lapelosa, M.; Levy, R. M. Binding Energy Distribution Analysis Method (BEDAM) for Estimation of ProteinLigand Binding Affinities. *J. Chem. Theory Comput.* **2010**, *6*, 2961–2977.
153. Ostermeir, K.; Zacharias, M. Rapid alchemical free energy calculation employing a generalized born implicit solvent model. *J. Phys. Chem. B* **2015**, *119*, 968–975.
154. Behler, J.; Parrinello, M. Generalized Neural-Network Representation of High-Dimensional Potential-Energy Surfaces. *Phys. Rev. Lett.* **2007**, *98*, 146401–146404.
155. Smith, J. S.; Isayev, O.; Roitberg, A. E. ANI-1: an Extensible Neural Network Potential with DFT Accuracy at Force Field Computational Cost. *Chem. Sci.* **2017**, *8*, 3192–3203.
156. Smith, J. S.; Roitberg, A. E.; Isayev, O. Transforming Computational Drug Discovery with Machine Learning and AI. *ACS Med. Chem. Lett.* **2018**, *9*, 1065–1069.
157. Smith, J.; Nebgen, B.; Zubatyuk, R.; Lubbers, N.; Devereux, C.; Barros, K.; Tretiak, S.; Isayev, O.; Roitberg, A. Approaching Coupled Cluster Accuracy with a General-purpose Neural Network Potential Through Transfer Learning. *Nat. Commun.* **2019**, *10*, 2903.
158. Nebgen, B.; Lubbers, N.; Smith, J. S.; Sifain, A. E.; Lokhov, A.; Isayev, O.; Roitberg, A. E.; Barros, K.; Tretiak, S. Transferable Dynamic Molecular Charge Assignment Using Deep Neural Networks. *J. Chem. Theory Comput.* **2018**, *14*, 4687–4698.
159. Unke, O. T.; Meuwly, M. A Reactive, Scalable, and Transferable Model for Molecular Energies from a Neural Network Approach based on Local Information. *J. Chem. Phys.* **2018**, *148*, 241708.

160. Schütt, K.; Sauceda, H.; Kindermans, P.; Tkatchenko, A.; Müller, K. SchNet—A Deep Learning Architecture for Molecules and Materials. *J. Chem. Phys.* **2018**, *148*, 241722.
161. Unke, O.; Meuwly, M. PhysNet: A Neural Network for Predicting Energies, Forces, Dipole Moments, and Partial Charges. *J. Chem. Theory Comput.* **2019**, *15*, 3678–3693.
162. Xie, T.; Grossman, J. C. Crystal Graph Convolutional Neural Networks for an Accurate and Interpretable Prediction of Material Properties. *Phys. Rev. Lett.* **2018**, *120*, 145301.
163. Zhang, Y.; Hu, C.; Jiang, B. Embedded Atom Neural Network Potentials: Efficient and Accurate Machine Learning with a Physically Inspired Representation. *J. Phys. Chem. Lett.* **2019**, *10*, 4962–4967.
164. Bartók, A.; Csányi, G. Gaussian Approximation Potentials: A Brief Tutorial Introduction. *Int. J. Quantum Chem.* **2015**, *115*, 1051–1057.
165. Shao, Y.; Hellstrom, M.; Mitev, P. D.; Knijff, L.; Zhang, C. PiNN: A Python Library for Building Atomic Neural Networks of Molecules and Materials. *J. Chem. Inf. Model.* **2020**, *60*, 1184–1193.
166. Khorshidi, A.; Peterson, A. A. Amp: A Modular Approach to Machine Learning in Atomistic Simulations. *Comput. Phys. Commun.* **2016**, *207*, 310–324.
167. Zhang, L.; Han, J.; Wang, H.; Car, R.; Weinan, E. Deep Potential Molecular Dynamics: a Scalable Model with the Accuracy of Quantum Mechanics. *Phys. Rev. Lett.* **2018**, *120*, 143001.
168. Chen, Y.; Zhang, L.; Wang, H.; Weinan, E. DeePKS: A Comprehensive Data-Driven Approach toward Chemically Accurate Density Functional Theory. *J. Chem. Theory Comput.* **2021**, *17*, 170–181.
169. Lindsey, R. K.; Fried, L. E.; Goldman, N.; Bastea, S. Active Learning for Robust, High-Complexity Reactive Atomistic Simulations. *J. Chem. Phys.* **2020**, *153*, 134117.
170. Gastegger, M.; McSloy, A.; Luya, M.; Schutt, K. T.; Maurer, R. J. A Deep Neural Network for Molecular Wave Functions in Quasi-Atomic Minimal Basis Representation. *J. Chem. Phys.* **2020**, *153*, 044123.
171. Mueller, T.; Hernandez, A.; Wang, C. Machine Learning for Interatomic Potential Models. *J. Chem. Phys.* **2020**, *152*, 050902.
172. 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.
173. Rappé, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A., III; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.
174. Momany, F. A.; Rone, R. Validation of the General Purpose QUANTA® 3.2/CHARMm® force field. *J. Comput. Chem.* **1992**, *13*, 888–900.
175. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M., Jr.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A Second Generation Force Field for the Simulation of Proteins, Nucleic Acids, and Organic Molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179–5197.
176. Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17*, 490–519.

177. Damm, W.; Frontera, A.; Tirado-Rives, J.; Jorgensen, W. L. OPLS all-atom force field for carbohydrates. *J. Comput. Chem.* **1997**, *18*, 1955–1970.
178. Warshel, A.; Levitt, M.; Lifson, S. Consistent Force Field for Calculation of Vibrational Spectra and Conformations of some Amides and Lactam Rings. *J. Mol. Spectrosc.* **1970**, *33*, 84–99.
179. Senftle, T. P.; Hong, S.; Islam, M. M.; Kylasa, S. B.; Zheng, Y.; Shin, Y. K.; Junkermeier, C.; Engel-Herbert, R.; Janik, M. J.; Aktulga, H. M.; Verstraelen, T.; Grama, A.; van Duin, A. C. T. The ReaxFF Reactive Force-Field: Development, Applications and Future Directions. *npj Comput. Mater.* **2016**, *2*, 1–14.
180. Lii, J.-H.; Allinger, N. L. The Important Role of Lone-Pairs in Force Field (MM4) Calculations on Hydrogen Bonding in Alcohols. *J. Phys. Chem. A* **2008**, *112*, 11903–11913.
181. Jorgensen, W. L.; Schyman, P. Treatment of Halogen Bonding in the OPLS-AA Force Field: Application to Potent Anti-HIV Agents. *J. Chem. Theory Comput.* **2012**, *8*, 3895–3901.
182. El Hage, K.; Bereau, T.; Jakobsen, S.; Meuwly, M. Impact of Quadrupolar Electrostatics on Atoms Adjacent to the Sigma-Hole in Condensed-Phase Simulations. *J. Chem. Theory Comput.* **2016**, *12*, 3008–3019.
183. Mobley, D. L.; Bannan, C. C.; Rizzi, A.; Bayly, C. I.; Chodera, J. D.; Lim, V. T.; Lim, N. M.; Beauchamp, K. A.; Shirts, M. R.; Gilson, M. K.; et al. Open Force Field Consortium: Escaping Atom Types Using Direct Chemical Perception with SMIRNOFF v0. 1. *bioRxiv* **2018**, 286542.
184. Jiang, W.; Chipot, C.; Roux, B. Computing Relative Binding Affinity of Ligands to Receptor: An Effective Hybrid Single-Dual-Topology Free-Energy Perturbation Approach in NAMD. *J. Chem. Inf. Model.* **2019**, *59*, 3794–3802.
185. Wang, L.; Deng, Y.; Wu, Y.; Kim, B.; LeBard, D. N.; Wandschneider, D.; Beachy, M.; Friesner, R. A.; Abel, R. Accurate Modeling of Scaffold Hopping Transformations in Drug Discovery. *J. Chem. Theory Comput.* **2017**, *13*, 42–54.
186. Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346–6358.
187. Deng, Y. Q.; Roux, B. Computation of Binding Free Energy with Molecular Dynamics and Grand Canonical Monte Carlo Simulations. *J. Chem. Phys.* **2008**, *128*, 115103.
188. Ross, G. A.; Russell, E.; Deng, Y.; Lu, C.; Harder, E. D.; Abel, R.; Wang, L. Enhancing water sampling in free energy calculations with grand canonical Monte Carlo. *J. Chem. Theory Comput.* **2020**, *16*, 6061–6076.
189. Ben-Shalom, I. Y.; Lin, Z.; Radak, B. K.; Lin, C.; Sherman, W.; Gilson, M. K. Accounting for the Central Role of Interfacial Water in Protein–Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2020**, *16*, 7883–7894.
190. Athanasiou, C.; Vasilakaki, S.; Dellis, D.; Cournia, Z. Using Physics-Based Pose Predictions and Free Energy Perturbation Calculations to Predict Binding Poses and Relative Binding Affinities for FXR Ligands in the D3R Grand Challenge 2. *J. Comput.-Aided Mol. Des.* **2018**, *32*, 21–44.
191. Oshima, H.; Re, S.; Sugita, Y. Prediction of Protein–Ligand Binding Pose and Affinity Using the gREST+FEP Method. *J. Chem. Inf. Model.* **2020**, *60*, 5382–5394.

192. Niitsu, A.; Re, S.; Oshima, H.; Kamiya, M.; Sugita, Y. De Novo Prediction of Binders and Nonbinders for T4 Lysozyme by gREST Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 3879–3888.
193. Lau, A. Y.; Roux, B. The Hidden Energetics of Ligand Binding and Activation in a Glutamate Receptor. *Nat. Struct. Mol. Biol.* **2011**, *18*, 283–287.
194. Buijnsters, P.; De Angelis, M.; Langlois, X.; Rombouts, F. J.; Sanderson, W.; Tresadern, G.; Ritchie, A.; Trabanco, A. A.; VanHoof, G.; Roosbroeck, Y. V.; et al. Structure-Based Design of a Potent, Selective, and Brain penetrating PDE2 Inhibitor with Demonstrated Target Engagement. *ACS Med. Chem. Lett.* **2014**, *5*, 1049–1053.
195. Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model.* **2015**, *55*, 2411–2420.
196. Wang, L.; Chambers, J.; Abel, R. Protein-Ligand Binding Free Energy Calculations with FEP+. *Biomolecular Simulations*; Springer, 2019; pp 201–232.
197. Lee, T.-S.; Allen, B. K.; Giese, T. J.; Guo, Z.; Li, P.; Lin, C.; Dwight, M. T.; Pearlman, D. A.; Radak, B. K.; Tao, Y.; et al. Alchemical Binding Free Energy Calculations in AMBER20: Advances and Best Practices for Drug Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 5595–5623.
198. Wang, L.; Wu, Y.; Deng, Y.; Kim, B.; Pierce, L.; Krilov, G.; Lupyan, D.; Robinson, S.; Dahlgren, M. K.; Greenwood, J.; Romero, D. L.; Masse, C.; Knight, J. L.; Steinbrecher, T.; Beuming, T.; Damm, W.; Harder, E.; Sherman, W.; Brewer, M.; Wester, R.; Murcko, M.; Frye, L.; Farid, R.; Lin, T.; Mobley, D. L.; Jorgensen, W. L.; Berne, B. J.; Friesner, R. A.; Abel, R. Accurate and Reliable Prediction of Relative Ligand Binding Potency in Prospective Drug Discovery by way of a Modern Free-Energy Calculation Protocol and Force Field. *J. Am. Chem. Soc.* **2015**, *137*, 2695–2703.
199. Bennett, C. H. Efficient Estimation of Free Energy Differences from Monte Carlo Data. *J. Comput. Phys.* **1976**, *22*, 245–268.
200. Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.
201. Souaille, M.; Roux, B. Extension to the Weighted Histogram Analysis Method: Combining Umbrella Sampling with Free Energy Calculations. *Comput. Phys. Commun.* **2001**, *135*, 40–57.
202. Tan, Z.; Gallicchio, E.; Lapelosa, M.; Levy, R. M. Theory of Binless Multi-State Free Energy Estimation with Applications to Protein-Ligand Binding. *J. Chem. Phys.* **2012**, *136*, 144102.
203. Ding, X.; Vilseck, J. Z.; Brooks, C. L., III. Fast Solver for Large Scale Multistate Bennett Acceptance Ratio Equations. *J. Chem. Theory Comput.* **2019**, *15*, 799–802.
204. Giese, T. J.; York, D. M. Variational Method for Networkwide Analysis of Relative Ligand Binding Free Energies with Loop Closure and Experimental Constraints. *J. Chem. Theory Comput.* **2021**, *17*, 1326–1336.
205. Kaus, J. W.; Harder, E.; Lin, T.; Abel, R.; McCammon, J. A.; Wang, L. How to Deal with Multiple Binding Poses in Alchemical Relative Protein-Ligand Binding Free Energy Calculations. *J. Chem. Theory Comput.* **2015**, *11*, 2670–2679.
206. Mobley, D. L.; Chodera, J. D.; Dill, K. A. On the Use of Orientational Restraints and Symmetry Corrections in Alchemical Free Energy Calculations. *J. Chem. Phys.* **2006**, *125*, 084902.

207. Martin, Y. C. Let's not Forget Tautomers. *J. Comput.-Aided Mol. Des.* **2009**, *23*, 693–704.
208. Wahl, O.; Sander, T. Tautobase: an Open Tautomer Database. *J. Chem. Inf. Model.* **2020**, *60*, 1085–1089.
209. Mongan, J.; Case, D. A.; McCammon, J. A. Constant pH Molecular Dynamics in Generalized Born Implicit Solvent. *J. Comput. Chem.* **2004**, *25*, 2038–2048.
210. Khandogin, J.; Brooks, C. L. Constant pH Molecular Dynamics with Proton Tautomerism. *Biophys. J.* **2005**, *89*, 141–157.
211. Zhou, R.; Das, P.; Royyuru, A. K. Single Mutation Induced H3N2 Hemagglutinin Antibody Neutralization: a Free Energy Perturbation Study. *J. Phys. Chem. B* **2008**, *112*, 15813–15820.
212. Gkeka, P.; Eleftheratos, S.; Kolocouris, A.; Cournia, Z. Free Energy Calculations Reveal the Origin of Binding Preference for Aminoadamantane Blockers of Influenza A/M2TM Pore. *J. Chem. Theory Comput.* **2013**, *9*, 1272–1281.
213. Kastenholz, M. A.; Hünenberger, P. H. Computation of Methodology-Independent Ionic Solvation Free Energies from Molecular Simulations. I. The Electrostatic Potential in Molecular Liquids. *J. Chem. Phys.* **2006**, *124*, 124106.
214. Kastenholz, M. A.; Hünenberger, P. H. Computation of Methodology-Independent Ionic Solvation Free Energies from Molecular Simulations. II. The Hydration Free Energy of the Sodium Cation. *J. Chem. Phys.* **2006**, *124*, 224501.
215. Lin, Y. L.; Aleksandrov, A.; Simonson, T.; Roux, B. An Overview of Electrostatic Free Energy Computations for Solutions and Proteins. *J. Chem. Theory Comput.* **2014**, *10*, 2690–2709.
216. Chen, W.; Deng, Y.; Russell, E.; Wu, Y.; Abel, R.; Wang, L. Accurate Calculation of Relative Binding Free Energies between Ligands with Different Net Charges. *J. Chem. Theory Comput.* **2018**, *14*, 6346–6358.
217. Clark, A. J.; Negron, C.; Hauser, K.; Sun, M.; Wang, L.; Abel, R.; Friesner, R. A. Relative Binding Affinity Prediction of Charge-Changing Sequence Mutations with FEP in Protein-Protein Interfaces. *J. Mol. Biol.* **2019**, *431*, 1481–1493.
218. Ekimoto, T.; Yamane, T.; Ikeguchi, M. Elimination of Finite-Size Effects on Binding Free Energies via the Warp-Drive Method. *J. Chem. Theory Comput.* **2018**, *14*, 6544–6559.
219. Zhang, H.; Jiang, W.; Chatterjee, P.; Luo, Y. Ranking Reversible Covalent Drugs: from Free Energy Perturbation to Fragment Docking. *J. Chem. Inf. Model.* **2019**, *59*, 2093–2102.
220. Lameira, J.; Bonatto, V.; Cianni, L.; dos Reis Rocho, F.; Leitão, A.; Montanari, C. A. Predicting the Affinity of Halogenated Reversible Covalent Inhibitors through Relative Binding Free Energy. *Phys. Chem. Chem. Phys.* **2019**, *21*, 24723–24730.
221. Kuhn, B.; Tichy, M.; Wang, L.; Robinson, S.; Martin, R. E.; Kuglstatter, A.; Benz, J.; Giroud, M.; Schirmeister, T.; Abel, R.; et al. Prospective Evaluation of Free Energy Calculations for the Prioritization of Cathepsin L Inhibitors. *J. Med. Chem.* **2017**, *60*, 2485–2497.
222. Chatterjee, P.; Botello-Smith, W. M.; Zhang, H.; Qian, L.; Alsamarah, A.; Kent, D.; Lacroix, J. J.; Baudry, M.; Luo, Y. Can Relative Binding Free Energy Predict Selectivity of Reversible Covalent Inhibitors? *J. Am. Chem. Soc.* **2017**, *139*, 17945–17952.
223. Shirts, M. R.; Chodera, J. D. Statistically Optimal Analysis of Samples from Multiple Equilibrium States. *J. Chem. Phys.* **2008**, *129*, 124105.

224. Tiwary, P.; van de Walle, A. A Review of Enhanced Sampling Approaches for Accelerated Molecular Dynamics. *Multiscale Materials Modeling for Nanomechanics*; Springer, 2016; pp 195–221.
225. Roe, D. R.; Bergonzo, C.; Cheatham, T. E., III. Evaluation of Enhanced Sampling provided by Accelerated Molecular Dynamics with Hamiltonian Replica Exchange Methods. *J. Phys. Chem. B* **2014**, *118*, 3543–3552.
226. Cavalli, A.; Spitaleri, A.; Saladino, G.; Gervasio, F. L. Investigating Drug-Target Association and Dissociation Mechanisms Using Metadynamics-Based Algorithms. *Acc. Chem. Res.* **2015**, *48*, 277–285.
227. Bernardi, R. C.; Melo, M. C.; Schulten, K. Enhanced Sampling Techniques in Molecular Dynamics Simulations of Biological Systems. *Biochim. Biophys. Acta, Gen. Subj.* **2015**, *1850*, 872–877.
228. Pohorille, A.; Jarzynski, C.; Chipot, C. Good Practices in Free-Energy Calculations. *J. Phys. Chem. B* **2010**, *114*, 10235–10253.
229. Mobley, D. L.; Dill, K. A. Binding of Small-Molecule Ligands to Proteins: What You See Is Not Always What You Get. *Structure* **2009**, *17*, 489–498.
230. Christ, C. D.; Mark, A. E.; van Gunsteren, W. F. Basic Ingredients of Free Energy Calculations: A Review. *J. Comput. Chem.* **2010**, *31*, 1569–1582.
231. Foloppe, N.; Hubbard, R. Towards Predictive Ligand Design With Free-Energy Based Computational Methods? *Curr. Med. Chem.* **2006**, *13*, 3583–3608.
232. Michel, J.; Essex, J. W. Prediction of Protein-Ligand Binding Affinity by Free Energy Simulations: Assumptions, Pitfalls and Expectations. *J. Comput.-Aided Mol. Des.* **2010**, *24*, 639–658.
233. Gallicchio, E.; Levy, R. M. Advances in all Atom Sampling Methods for Modeling Protein-Ligand Binding Affinities. *Curr. Opin. Struct. Biol.* **2011**, *21*, 161–166.
234. Chodera, J. D.; Mobley, D.; Shirts, M.; Dixon, R.; Branson, K.; Pande, V. Alchemical Free Energy Methods for Drug Discovery: Progress and Challenges. *Curr. Opin. Struct. Biol.* **2011**, *21*, 150–160.
235. de Ruiter, A.; Oostenbrink, C. Advances in the Calculation of Binding Free Energies. *Curr. Opin. Struct. Biol.* **2020**, *61*, 207–212.
236. Klimovich, P. V.; Shirts, M. R.; Mobley, D. L. Guidelines for the Analysis of Free Energy Calculations. *J. Comput.-Aided Mol. Des.* **2015**, *29*, 397–411.
237. Ding, X.; Vilseck, J. Z.; Hayes, R. L.; Brooks, C. L., III. Gibbs Sampler-Based λ-Dynamics and Rao-Blackwell Estimator for Alchemical Free Energy Calculation. *J. Chem. Theory Comput.* **2017**, *13*, 2501–2510.
238. Zhang, B. W.; Xia, J.; Tan, Z.; Levy, R. M. A Stochastic Solution to the Unbinned WHAM Equations. *J. Phys. Chem. Lett.* **2015**, *6*, 3834–3840.
239. Cui, D.; Zhang, B. W.; Tan, Z.; Levy, R. M. Ligand Binding Thermodynamic Cycles: Hysteresis, the Locally Weighted Histogram Analysis Method, and the Overlapping States Matrix. *J. Chem. Theory Comput.* **2020**, *16*, 67–79.
240. Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling Local Structural Rearrangements Using FEP/REST: Application to Relative

Binding Affinity Predictions of CDK2 Inhibitors. *J. Chem. Theory Comput.* **2013**, *9*, 1282–1293.

241. Homeyer, N.; Gohlke, H. FEW: A Workflow Tool for Free Energy Calculations of Ligand Binding. *J. Comput. Chem.* **2013**, *34*, 965–973.
242. Loeffler, H. H.; Michel, J.; Woods, C. FESetup: Automating Setup for Alchemical Free Energy Simulations. *J. Chem. Inf. Model.* **2015**, *55*, 2485–2490.
243. Gathiaka, S.; Liu, S.; Chiu, M.; Yang, H.; Stuckey, J. A.; Kang, Y. N.; Delproposto, J.; Kubish, G.; Dunbar, J. B.; Carlson, H. A.; Burley, S. K.; Walters, W. P.; Amaro, R. E.; Feher, V. A.; Gilson, M. K. D3R Grand Challenge 2015: Evaluation of Protein-Ligand Pose and Affinity Predictions. *J. Comput.-Aided Mol. Des.* **2016**, *30*, 651–668.
244. Fu, H.; Gumbart, J. C.; Chen, H.; Shao, X.; Cai, W.; Chipot, C. BFEE: A User-Friendly Graphical Interface Facilitating Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2018**, *58*, 556–560.
245. Kuhn, M.; Firth-Clark, S.; Tosco, P.; Mey, A. S. J. S.; Mackey, M.; Michel, J. Assessment of Binding Affinity via Alchemical Free-Energy Calculations. *J. Chem. Inf. Model.* **2020**, *60*, 3120–3130.
246. Fu, H.; Chen, H.; Cai, W.; Shao, X.; Chipot, C. BFEE2: Automated, Streamlined and Accurate Absolute Binding Free-Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61*, 2116–2123.
247. Chemical Computing Group ULC, 1010 Sherbooke St. West, Suite 910, Montreal, QC, Canada, H3A 2R7. *Molecular Operating Environment (MOE)*, 2021.
248. Labute, P.; Ebert, M. Optimizing Simulations Protocols for Relative Free Energy Calculations. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 227–245.
249. Biovia; Dassault Systèmes. *Discovery Studio Modeling Environment*, 2017.
250. Westerhoff, L. M.; Zheng, Z. Fast, Routine Free Energy of Binding Estimation Using MovableType. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 247–265.
251. Case, D. A.; Aktulga, H. M.; Belfon, K.; Ben-Shalom, I. Y.; Brozell, S. R.; Cerutti, D. S.; Cheatham, T. E., III; Cruzeiro, V. W. D.; Darden, T. A.; Duke, R. E.; Giambasu, G.; Gilson, M. K.; Gohlke, H.; Goetz, A. W.; Harris, R.; Izadi, S.; Izmailov, S. A.; Jin, C.; Kasavajhala, K.; C., K. M.; King, E.; Kovalenko, A.; Kurtzman, T.; Lee, T. S.; LeGrand, S.; Li, P.; Lin, C.; Liu, J.; T., L.; Luo, R.; Machado, M.; Man, V.; Manathunga, M.; Merz, K. M., Jr.; Miao, Y.; Mikhailovskii, O.; Monard, G.; Nguyen, H.; O'Hearn, K.; Onufriev, A.; Pan, F.; Pantano, S.; Qi, R.; Rahnamoun, A.; Roe, D. R.; Roitberg, A.; Sagui, C.; Schott-Verdugo, S.; Shen, J.; Simmerling, C. L.; Skrynnikov, N.; Smith, J.; Swails, J.; Walker, R. C.; Wang, J.; Wei, H.; Wolf, R.; Wu, X.; Xue, Y.; York, D. M.; Zhao, S.; Kollman, P. A. *Amber* **2021**, *2021*.
252. Liu, S.; Wu, Y.; Lin, T.; Abel, R.; Redmann, J. P.; Summa, C. M.; Jaber, V. R.; Lim, N. M.; Mobley, D. L. Lead Optimization Mapper: Automating Free Energy Calculations for Lead Optimization. *J. Comput.-Aided Mol. Des.* **2013**, *27*, 755–770.
253. Woods, C. J.; Mey, A. S. J. S.; Calabro, G.; Julien, M. *Sire Molecular Simulations Framework*, 2016.

254. Hedges, L. O.; Mey, A. S. J. S.; Laughton, C. A.; Gervasio, F. L.; Mulholland, A. J.; Woods, C. J.; Michel, J. BioSimSpace: An Interoperable Python Framework for Biomolecular Simulation. *J. Open Source Software* **2019**, *4*, 1831.
255. Lee, T.-S.; Tsai, H.-C.; Ganguly, A.; Giese, T. J.; York, D. M. Robust, efficient and automated methods for accurate prediction of protein-ligand binding affinities in AMBER Drug Discovery Boost. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 161–204.
256. Jespers, W.; Esguerra, M.; Åqvist, J.; Gutiérrez-de Terán, H. QligFEP: An Automated Workflow for Small Molecule Free Energy Calculations in Q. *J. Cheminf.* **2019**, *11*, 1–16.
257. Marelius, J.; Kolmodin, K.; Feierberg, I.; Åqvist, J. Q: A Molecular Dynamics Program for Free Energy Calculations and Empirical Valence Bond Simulations in Biomolecular Systems. *J. Mol. Graphics Modell.* **1998**, *16*, 213–225.
258. Abraham, M.; Apostolov, R.; Barnoud, J.; Bauer, P.; Blau, C.; Bonvin, A. M.; Chavent, M.; Chodera, J. D.; Čondić Jurkić, K.; Delemotte, L.; Grubmüller, H.; Howard, R. J.; Jordan, E. J.; Lindahl, E.; Ollila, O. H. S.; Selent, J.; Smith, D. G. A.; Stansfeld, P. J.; Tiemann, J. K. S.; Trellet, M.; Woods, C.; Zhmurov, A. Sharing Data from Molecular Simulations. *J. Chem. Inf. Model.* **2019**, *59*, 4093–4099.
259. Abraham, M. J.; Murtola, T.; Schulz, R.; Páll, S.; Smith, J. C.; Hess, B.; Lindahl, E. GROMACS: High Performance Molecular Simulations Through Multi-Level Parallelism from Laptops to Supercomputers. *SoftwareX* **2015**, *1*, 19–25.
260. Dodda, L. S.; Cabeza de Vaca, I.; Tirado-Rives, J.; Jorgensen, W. L. LigParGen web server: an automatic OPLS-AA parameter generator for organic ligands. *Nucleic Acids Res.* **2017**, *45*, W331–W336.
261. Zavitsanou, S.; Tsengenes, A.; Papadourakis, M.; Amendola, G.; Chatzigoulas, A.; Dellis, D.; Cosconati, S.; Cournia, Z. FEPrepare: A Web-Based Tool for Automating the Setup of Relative Binding Free Energy Calculations. *J. Chem. Inf. Model.* **2021**, *61*, 4131–4138.
262. Kim, S.; Oshima, H.; Zhang, H.; Kern, N. R.; Re, S.; Lee, J.; Roux, B.; Sugita, Y.; Jiang, W.; Im, W. CHARMM-GUI Free Energy Calculator for Absolute and Relative Ligand Solvation and Binding Free Energy Simulations. *J. Chem. Theory Comput.* **2020**, *16*, 7207–7218.
263. Parks, C. D.; Gaieb, Z.; Chiu, M.; Yang, H.; Shao, C.; Walters, W. P.; Jansen, J. M.; McGaughey, G.; Lewis, R. A.; Bembenek, S. D.; Ameriks, M. K.; Mirzadegan, T.; Burley, S. K.; Amaro, R. E.; Gilson, M. K. D3R Grand Challenge 4: Blind Prediction of Protein-Ligand Poses, Affinity Rankings, and Relative Binding Free Energies. *J. Comput.-Aided Mol. Des.* **2020**, *34*, 99–119.
264. Tielker, N.; Eberlein, L.; Beckstein, O.; Güssregen, S.; Iorga, B. I.; Kast, S. M.; Liu, S. Perspective on the SAMPL and D3R blind prediction challenges for physics-based free energy Methods. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D. , Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 67–107.
265. Yin, J.; Henriksen, N. M.; Slochower, D. R.; Shirts, M. R.; Chiu, M. W.; Mobley, D. L.; Gilson, M. K. Overview of the SAMPL5 Host–Guest Challenge: Are we Doing Better? *J. Comput.-Aided Mol. Des.* **2017**, *31*, 1–19.

266. Walters, W. P. Code Sharing in the Open Science Era. *J. Chem. Inf. Model.* **2020**, *60*, 4417–4420.
267. Hildebrand, P. W.; Rose, A. S.; Tiemann, J. K. S. Bringing Molecular Dynamics Simulation Data into View. *Trends Biochem. Sci.* **2019**, *44*, 902–913.
268. Merz, K. M., Jr.; Amaro, R.; Cournia, Z.; Rarey, M.; Soares, T.; Tropsha, A.; Wahab, H. A.; Wang, R. Editorial: Method and Data Sharing and Reproducibility of Scientific Results. *J. Chem. Inf. Model.* **2020**, *60*, 5868–5869.
269. Suruzhon, M.; Samways, M. L.; Essex, J. W. On the Issues Impacting Reproducibility of Alchemical Free Energy Calculations. In *Free Energy Methods in Drug Discovery: Current State and Future Directions*; Armacost, K., Thompson, D., Eds.; ACS Symposium Series; American Chemical Society: Washington, DC, 2021; Vol. 1397, pp 109–125.
270. Amaro, R. E.; Mulholland, A. J. A Community Letter Regarding Sharing Biomolecular Simulation Data for COVID-19. *J. Chem. Inf. Model.* **2020**, *60*, 2653–2656.
271. Austin, C. C.; Bernier, A.; Bezuidenhout, L.; Bicarregui, J.; Biro, T.; Cambon-Thomsen, A.; Carroll, S. R.; Cournia, Z.; Dabrowski, P. W.; Diallo, G.; Duflot, T.; Garcia, L.; Gesing, S.; Gonzalez-Beltran, A.; Gururaj, A.; Harrower, N.; Lin, D.; Medeiros, C.; Méndez, E.; Meyers, N.; Mietchen, D.; Nagrani, R.; Nilsonne, G.; Parker, S.; Pickering, B.; Pienta, A.; Polydoratou, P.; Psomopoulos, F.; Rennes, S.; Rowe, R.; Sansone, S.-A.; Shanahan, H.; Sitz, L.; Stocks, J.; Tovani-Palone, M. R.; Uhlmansiek, M. Research Data Alliance, Fostering Global Data Sharing: Highlighting the Recommendations of the Research Data Alliance COVID-19 Working Group. *Wellcome Open Res.* **2020**, *5*.