



Published in final edited form as:

Curr Opin Struct Biol. 2011 April ; 21(2): 150–160. doi:10.1016/j.sbi.2011.01.011.

Alchemical free energy methods for drug discovery: Progress and challenges

John D. Chodera^a, David L. Mobley^b, Michael R. Shirts^c, Richard W. Dixon^d, Kim Branson^d, and Vijay S. Pande^e

John D. Chodera: jchodera@berkeley.edu; David L. Mobley: dlmobley@uno.edu; Michael R. Shirts: michael.shirts@virginia.edu; Richard W. Dixon: richard_dixon@vrtx.com; Kim Branson: kim_branson@vrtx.com; Vijay S. Pande: pande@stanford.edu

^aResearch Fellow, California Institute of Quantitative Biosciences (QB3), University of California, Berkeley, 260J Stanley Hall, Berkeley, California 94720, USA

^bUniversity of New Orleans, New Orleans, LA 70148

^cUniversity of Virginia, Charlottesville, VA 22904

^dVertex Pharmaceuticals, Cambridge, MA 02139

^eDepartment of Chemistry, Stanford University, Stanford, CA 94305

Abstract

Improved rational drug design methods are needed to lower the cost and increase the success rate of drug discovery and development. Alchemical binding free energy calculations, one potential tool for rational design, have progressed rapidly over the last decade, but still fall short of providing robust tools for pharmaceutical engineering. Recent studies, especially on model receptor systems, have clarified many of the challenges that must be overcome for robust predictions of binding affinity to be useful in rational design. In this review, inspired by a recent joint academic/industry meeting organized by the authors, we discuss these challenges and suggest a number of promising approaches for overcoming them.

Keywords

alchemical free energy calculations; molecular dynamics; ligand binding; free energy perturbation

1. Introduction

R&D spending in the pharmaceutical industry has risen sharply in the last decade, with real expenditures by members of the U.S. pharmaceutical trade industry PhRMA doubling to \$65.3 billion in 2009 from \$32.4 billion in 2000 (in 2009 dollars) [1]. Despite this, the number of new molecular entities (NMEs) approved by the U.S. Food and Drug Administration (FDA) from 2004–2009 was only half that of the previous five years [2], and the number of truly innovative NMEs has remained stable at 5–6 per year [3]. This situation is especially grim if one considers the continual emergence of drug-resistant strains of

© 2011 Elsevier Ltd. All rights reserved.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

viruses and bacteria, a process which actively depletes the limited repertoire of useful therapeutics, sometimes leaving few, if any, alternatives in treatment.

Drug discovery has begun to integrate *rational design* techniques, in which a drug is engineered with the help of structural biology, alongside traditional screening approaches—a shift reflected in FDA approval requirements that make it difficult to move therapeutics of unknown mechanism forward. While virtual screening methods have wide deployment within the industry and play a large role in modern drug discovery efforts, there is concern that these methods may have reached a limit in effectiveness [4]. Although undoubtedly useful in eliminating some inactive compounds, current virtual screening methods are insufficiently effective in selecting molecules that are actually bioactive against the desired target; lead optimization efforts alone still consume, on average, two years and \$146 million [3].

Given that bridges, buildings, and aircraft are now regularly designed entirely using computers [5], why is it that we cannot design small molecules of a few dozen atoms? Admittedly, design goals are often complex—potential therapeutics must not only possess high affinity to the target, but meet multiple additional criteria, such as high selectivity, low off-target activity, good solubility, and a host of bioavailability and toxicity properties collectively known as *ADME-Tox*—absorption, distribution, metabolism, excretion, and toxicity. But it is precisely complex, multi-objective design problems where a computational approach should be superior to human-guided efforts. If computational approaches are currently ineffective, it is likely because we lack good predictive models for each of the individual objectives.

How can we move beyond the limitations of current virtual screening methods? Existing approaches rely upon a variety of approximations to allow large numbers of compounds to be screened quickly, often neglecting, or considering in an *ad hoc* fashion, statistical mechanical effects (such as conformational entropy, averaging over multiple conformations or binding modes, and the discrete nature of solvent) and chemical effects (such as protonation state and tautomer distributions, and their shifts upon binding) for computational efficiency. Unfortunately, it is precisely the neglect of these effects that is likely to be responsible for the gross inaccuracies of current scoring functions when making quantitative estimates of binding interactions [4].

Free energy calculations, at least in principle, offer a way to incorporate these effects to compute quantitatively accurate binding affinities. *Alchemical* free energy methods [6] in particular show great potential in enabling the computation of binding free energies with reasonable computational resources. In an alchemical calculation, instead of simulating the binding/unbinding processes directly, which would require a simulation many times the lifetime of the complex, the ligand is alchemically transmuted into either another chemical species or a noninteracting “dummy” molecule through intermediate, possibly nonphysical stages. Because free energy is a state function, the choice of intermediates is in principle arbitrary, but in practice, can have great impact on the efficiency of the calculation [7].

These methods experienced a wave of initial enthusiasm in the late 1980s and early 1990s following their introduction, but this enthusiasm was quickly quelled when it became evident that some of these early successes were due either to luck or conformation bias [8]. In the intervening decade, numerous methodological advances (see [9, 10, 11, 12, 13] for recent reviews) have sparked a new wave of enthusiasm. But are these advances sufficient for alchemical free energy methods to finally play a role in drug discovery efforts? And if so, what barriers remain to their widespread use in industry alongside current docking and scoring virtual screening tools?

To address precisely these questions, the authors organized a recent workshop in Cambridge, MA, hosted at Vertex Pharmaceuticals [14]. Their intent was to bring together representatives of the pharmaceutical industry, lead practitioners of free energy methods from academia, and representatives of companies that build the current generation of state-of-the-art virtual screening tools to identify which problems within industry might benefit from practical forms of these tools, as well as the operational hurdles that currently prevent the application of these tools. Pharmaceutical industry representatives made it clear that multiple opportunities exist to support the traditional structure-activity relationship (SAR) driven preclinical optimization process via improved compound ranking and prioritization. Surprisingly, statistical models of prediction-guided prioritization suggest that even moderate accuracy (RMS errors of ~ 2 kcal/mol) could be sufficient to produce substantial efficiency gains in lead optimization campaigns [11]. The ability to suggest considerably less conservative structural modifications, beyond the guidance of observed SAR, would have significant impact provided robust predictions could be made for the target compounds. Selectivity optimization, in which proposed modifications are evaluated for their negative impact on binding to non-therapeutic molecular targets, would be of tremendous use, especially in designing isoform-selective inhibitors for kinase and other targets. Late-stage preclinical evaluation frequently identifies issues, such as pharmacokinetic (PK) liabilities or toxicity risk factors, sufficient to halt progression of compounds with promising activity. The ability to rescue such compounds through potentially radical alteration of the core chemical scaffold while maintaining target potency and selectivity would have great utility in such cases. Finally, the qualities required for a drug include not only target potency and metabolism but also the reliable manufacture and formulation of the compound. Successful optimization of these parameters is often driven by the physical and structural properties such as solubility [15, 16], logP [17, 18], and crystal form of the active ingredient [19] that can in principle be computed through alchemical free energy approaches.

In this review, we aim to briefly address issues relevant to these opportunities, highlighting relevant work from the recent literature. We make no attempt to be comprehensive—such a task would be daunting, given that over 3,500 papers using the most popular free energy computation approaches were published in the last decade, with the publication rate increasing $\sim 17\%$ per year. After briefly surveying recent literature to assess the current state of alchemical free energy methods, we discuss a number of challenges that remain before these approaches can provide clear utility in industrial drug discovery challenges where existing virtual screening approaches are struggling, noting recent work hinting at potential solutions where possible. Finally, we outline several steps that can be taken to clearly move the field toward the goal of producing effective tools for rational drug design.

2. The state of the art

Alchemical free energy methods can be used to compute either *absolute* binding affinities (for an individual ligand to a receptor) or *relative* binding affinities (a difference between two or more related ligands). In lead optimization efforts, where optimization through small, sequential chemical modification is of primary interest, accurate relative free energies could determine whether modifications have increased affinity and selectivity. If uncorrelated conformations can easily be sampled, relative free energy calculations (recently reviewed [12]) can be more efficient, requiring fewer alchemical states (and hence less computational effort) to bridge the phase space between two related molecules. However, free energy calculations utilizing straightforward molecular dynamics simulations generally suffer from slow exploration along many conformational degrees of freedom, which introduces difficult sampling issues for both absolute and relative free energy calculations.

If all ligands share the same binding mode and no protein conformational changes occur that modulate the protein-ligand interactions, relative free energies may benefit from fortuitous cancellation of errors, facilitating the computation of precise relative binding affinities in practical computation times. However, protein conformational changes, even at the sidechain ro-tamer level, can be far too slow to sample in molecular dynamics simulations only a few nanoseconds in length [20, 21], yet the energetics of these changes can have significant effects on binding affinities. Relative calculations avoid this problem *only* if the protein conformational change affects the binding free energy of each ligand identically, which is not likely to be the case in general [20]. Dominant ligand binding modes can be far from obvious, even given the bound structure of a closely related ligand [22, 23, 24] and it may not be possible to sample all potential binding modes in a single simulation, leading to dramatically different relative binding free energy estimates depending on the starting structure [22]. In some cases, multiple binding modes may be relevant; this has been observed by in calculations [21, 22, 23, 24, 25, 26] and in experiments in which multiple binding modes are clearly resolved [27, 28, 29, 30, 31] or in which minute changes to a ligand dramatically alter the binding mode (e.g. [31] and references therein). Even choices of alchemical intermediates, such as whether and how artificial restraining potentials are used, can introduce artificially long correlation times that frustrate sampling [32]. As a result, situations where the “cancellation of errors” assumption breaks down are almost impossible to predict ahead of time, and can lead to highly erroneous relative or absolute free energy differences that make failure to agree with experiment difficult to interpret [22]; this is likely at least partly to blame for their notorious lack of reliability. Despite this, academic lead optimization efforts relying on this assumed cancellation of errors have had some successes, though they often lack quantitative accuracy and human guidance is typically necessary (e.g. [33]).

Absolute calculations (recently reviewed [9, 11]), on the other hand, greatly simplify the ability to learn from failures and hence improve algorithms and forcefields. Unlike relative free energies, where the experimental error is a large fraction of the typical dynamic range seen in related compounds synthesized in a lead optimization effort, absolute free energies cover a much larger dynamic range of binding affinities, so that experimental error is a much smaller fraction of this range. Interpretation of failure is also easier, as it is clear which compounds differ from experiment; with relative free energies, it is often not clear whether the calculation for one or both compounds suffer from pathologies. In the end, if the goal is to produce accurate, robust, and reliable methodologies for free energy calculations, absolute and relative free energy calculations have identical challenges as far as sampling and accuracy, though errors introduced by these challenges may be somewhat smaller in relative calculations in general. Because the lessons for accuracy and reliability are often clearer, we focus our review on studies that compute absolute free energies.

Common practice for assessing the performance of free energy calculations has been comparison of the predicted free energies of binding to experimental affinity measurements, often obtained by biophysical [34] or enzymatic assays. However, experimental measurements are invariably contaminated with error, which can affect maximum possible correlation with experiment that can be achieved [35]. Further, experiments often measure proxies for the binding free energy or affinity (such as the IC₅₀ or the apparent inhibition constant K_i) which do not always provide a reliable estimate of the binding free energy except under very specific mechanistic conditions [36]. Finally, dynamic range in experimental measurements may be limited (often spanning only 3–4 kcal/mol or less), meaning that low root-mean-square (RMS) errors with experiment may not be difficult to attain with a method that provides the right order of magnitude estimate for affinity. Any measure of expected utility of free energy calculations in effectively directing drug discovery efforts will need to take these issues into account [35].

Work assessing the accuracy of absolute binding free energy calculations has largely focused on a few model receptor systems, due to the ease with which failures can yield useful methodological insights. In recent years, the most popular of these model systems has undoubtedly been a hydrophobic cavity mutant (L99A) of T4 lysozyme (also reviewed in [11]). Despite the simplicity of the small apolar binding site and relative rigidity of the protein, this system has proven surprisingly challenging for rapid virtual screening methods like docking [37], and has been nontrivial for free energy methods to quantitatively predict affinity [21]. Many ligands (of which toluene is a prototypical example) are small and reminiscent of fragment screening sets, rather than drug-like molecules, and therefore possess multiple nearly-degenerate binding orientations separated by substantial kinetic barriers, frustrating quantitative estimation of affinity [32]. This also makes it difficult to predict the experimentally-resolved binding mode, as noted elsewhere [38, 39, 31].

Slow repacking rearrangements of some side chains have been observed upon binding, requiring very long simulations or divide-and-conquer approaches to achieve convergence [40, 41]. Despite this, addressing these issues allows current-generation forcefields to obtain RMS errors in computed binding free energies of approximately 1–2 kcal/mol [32, 42, 43, 44], though we note that the dynamic range of ligand binding affinities is relatively small (3–4 kcal/mol). Introduction of an additional mutation, M102Q, creates a polar version of this binding site; RMS errors of 1–2 kcal/mol have been reported for this system (with known binders again spanning a 3–4 kcal/mol range) [22, 44].

Another popular model system has been the FK506 binding protein 12 (FKBP12). This protein, pharmaceutically interesting due to its role in suppressing immune response, binds a number of large cyclic natural products and related molecules. Several studies of this system have reported the results of alchemical calculations [45, 25, 46, 47, 48]. Notably, computed binding affinities vary between studies by up to 2–3 kcal/mol, likely an indication of long time scales leading to convergence difficulties in short simulations, as well as differences related to handling of the standard state [47]. Other factors that differ among these studies (including force field, simulation setup, simulation package, details of sampling approach) make it difficult to further interpret differences in RMS errors achieved. Several of these studies, however, have directly demonstrated that correlation times for internal ligand degrees of freedom can be tens of nanoseconds, reinforcing the importance of sufficient simulation lengths or enhanced sampling techniques. Even seemingly minor details such as the need for an inhomogeneous dispersion correction to account for the differing density of van der Waals sites in the protein and solvent can result in differences of up to 1 kcal/mol [49].

The serine protease trypsin, which has an exposed binding pocket able to accommodate relatively small positively-charged ligands (such as the prototypical inhibitor benzamidine), has also been the focus of a number of recent relative and absolute free energy calculation studies. Earlier work on this system found that predicted binding affinities generally captured experimental trends for substituted benzamidines, but the computed range of binding affinities was shifted and enlarged toward more favorable binding; computed free energies relative to unsubstituted benzamidine ranged from -2.1 to +0.17 kcal/mol, while calorimetrically determined free energies only ranged from -0.64 to +0.91 kcal/mol [50, 51]. More recent studies with the AMOEBA polarizable force field [52] reported markedly improved agreement with experiment (average error less than 0.5 kcal/mol), though at far greater computational expense [53, 54]. However, due to the small size of this study (five ligands) [53, 54], it may be premature to expect these accuracy gains from the use of polarizable forcefields will be consistently realized.

Other studies of particular interest include the calculation of absolute binding free energies of antibiotics targeting the bacterial ribosome [55], and application of absolute free energy techniques to binding to a bacterial membrane transporter [56]. In some cases, relative free energy calculations are being directly utilized in lead optimization in drug discovery efforts—notably the Jorgensen lab's work applying rapid free energy calculations in several systems [57, 58, 59, 59] and the work of Steinbrecher and collaborators [39]. Studies of tetracycline binding to the Tet repressor protein (TetR) of Gram-negative bacteria highlighted the large effect that choice of conformation and protonation state has on the computed binding free energy [60]. Finally, recent work of Michel and Essex has highlighted how free energy methods can be much more effective than docking methods at identifying ligands of the estrogen receptor [38].

Many studies have described new algorithmic advances, but there are very few thorough evaluations of free energy methods. Even fewer have been tested on the same standard benchmark systems, making it difficult to evaluate how much progress the field has made over time. As a result, we still have a very limited idea about when alchemical free energy methods can currently be expected to perform well. The largely anecdotal literature, however, does provide us with a much clearer understanding of when they can be expected to perform poorly. Conformational changes slow enough to present sampling difficulties, even at the single side chain level, can affect computed binding affinities to a significant degree [40, 22]. It is, unfortunately, nearly impossible to know when these issues will appear; for example, two newly-characterized ligands to the well-studied T4 lysozyme L99A/M102Q polar binding site surprisingly induced novel protein conformational changes, leading to large errors in the computed binding free energies in which these changes were not sampled in the simulation time scales [22].

Receptors accommodating charged ligands also appear to present additional challenges. There is of yet relatively little alchemical free energy work examining these systems, but one study on a model binding site in cytochrome C peroxidase (CCP) found alchemical techniques substantially overestimated the magnitude and span of binding free energies [61], a finding confirmed in some trypsin studies [50, 51]. There exist technical reasons (related to the treatment of long-range electrostatics) why binding free energies of charged ligands may be especially difficult to calculate with these techniques, necessitating corrections to the computed free energies [62, 63]. Hence, the published data suggest that as we move away from relatively rigid binding sites and neutral ligands, there is the potential for considerably more uncertainty in binding free energy estimates.

Even in favorable cases, care must be taken to sample all relevant ligand binding modes [21, 22], as these can sometimes change in unexpected ways when a scaffold is modified [31]. Provided any relevant long time scale motions can be overcome, errors can be in the 1–2 kcal/mol range in computed binding free energies [21, 22, 44] or occasionally even better [53, 54]. However, extremely large errors—in excess of 6 kcal/mol—are possible in some situations [64]. In some cases, the same method can yield dramatically different results across targets—yielding R^2 values near 0.8 for some targets while giving correlations near zero for others for reasons that remain unclear [65].

3. Challenges and potential solutions

The challenging aspects of a binding free energy calculation can naturally be separated into four categories: modeling and simulation setup, sampling, force field accuracy, and analysis. Researchers must model the relevant chemical species, assign force field parameters of sufficient accuracy, and choose appropriate alchemical intermediates. They must then employ some method capable of sampling the relevant configurations (and potentially,

chemical states) with the appropriate probability during each phase of the calculation, using limited computer resources and wall clock time. Finally, they must analyze the results in a way that detects sampling problems and obtain as accurate an estimate as possible. Issues with any of these aspects can lead to significant deviations between computation and experiment.

Modeling and simulation preparation

Before performing a calculation, an atomistic model of the receptor-ligand system must be constructed. This model must contain all of the chemical components essential to computing a quantitatively accurate binding affinity. Creating the model may require generating a complete atomic structure of the receptor from incomplete or inexact structural data, assigning an appropriate protonation state, constructing an atomic model of the ligand in an appropriate tautomeric and protonation state, and docking the ligand into the receptor to generate initial configurations for simulation. Salts or counterions may influence the binding affinities, as might any post-translational modifications or the presence of other bound species; cofactors such as heme and nicotinamide adenine dinucleotide (NAD⁺) are not uncommon.

Forcefield parameters for all of these chemical species present in the model must also be generated or assigned from a database. Procedurally, this process is still complex and time-consuming, which has no doubt played a role in the limited adoption of these approaches within pharma, where lead optimization cycles operate on time scales of a few weeks. Due to a lack of commonly agreed-upon best practices (despite recent efforts [66, 12, 13]), many decisions must be made that require expert knowledge to avoid errors that may have a significant impact on the computed binding affinity. While tools for automated ligand parameterization do exist [67], these often struggle with exotic chemistries, and there are numerous anecdotal reports of issues even for mundane chemistries. Tools for automatically performing ligand and complex modeling and preparation are sorely needed, though recent attempts from industry have made some progress in this direction (e.g. integration of Desmondsetup into Schrödinger's Maestro product). Alchemical intermediates must be selected to provide sufficient overlap without too much wasted effort; recent work suggests the beginnings of theory and methodology for doing this optimally [68].

Sampling

Alchemical free energy calculations require sampling from the equilibrium distribution of a number of thermodynamic states in which interactions between the ligand and its environment are modulated. This requires the equilibrium sampling scheme—such as molecular dynamics (MD) or Monte Carlo (MC)—to move away from the initial structure into a region of high equilibrium probability (*equilibration*), and to “mix” well within the equilibrium-populated regions of conformation space so that all relevant states are sampled a sufficient number of times during the simulation to obtain a precise estimate (*convergence*).

To complicate matters, the protein and/or ligand may change, or exist as a mixture of, protonation [69, 60] or tautomeric states [70] upon binding, or there may be significant populations of multiple such states during some part of the binding process. This has recently been termed the *multiple state problem* [10], and may require semigrand canonical simulation methodologies to address, such as those described in Refs. [71, 72].

As a benchmark for accessible timescales, modern eight-core Intel Core i7 processors using the popular Gromacs MD code [73] can simulate solvated dihydrofolate reductase (23,569 atoms) at a rate of ~10 ns/day. If one CPU-day is expended for each alchemical state in standard free energy calculation, all relevant conformational transitions that can affect the

receptor-ligand interactions must be sampled at time scales much shorter than 10 ns. If, on the other hand, binding a ligand induces an allosteric conformational change in the receptor, but the time scales for conformational change far exceed 10 ns, then the simulation will fail to sample the relevant conformations in proper proportion generally leading to incorrect binding affinities.

Solvent degrees of freedom generally relax on the time scale of picoseconds to nanoseconds; as a result, alchemical transformations in solvent are generally easily converged in current practical simulation time scales. Still, slow torsional transitions in small molecules (such as in carboxylic acids) can cause surprising convergence issues even in hydration free energy calculations [74]. Relaxation of protein conformational degrees of freedom can be considerably slower; side chain reorganization can occur on microsecond time scales, and large-scale allosteric conformational changes on the millisecond time scale. Simply “waiting” for these conformational changes to occur is beyond what one can expect from modern MD simulations, which can typically only reach microsecond simulation time scales with great effort—one must explicitly consider schemes to directly address time scales in excess of what can be sampled.

One way to speed sampling is simply speed up the MD or MC simulation directly with improved algorithms or hardware. Recently, new parallel force calculation algorithms, such as neutral territory decomposition [75, 76], have been incorporated into MD packages such as Desmond [77], Gromacs [78] and NAMD [79]. As commodity hardware is reaching a limit in clock speeds, groups have looked to specialized hardware [80, 81], or developing entirely new algorithms to efficiently map these calculation onto commercial graphics processing units (GPUs) [82, 83, 84, 85]. GPUs are especially attractive for their ability to inexpensively deliver a theoretical peak of ~2 TFLOP/s of computing power, and the surprising capability of the industry to double (at least for now) this figure approximately every 12 months.

One algorithmic approach to circumvent slow correlation times is to decompose the configuration space into smaller, overlapping regions along the slow degree of freedom; each region could be efficiently sampled independently, and the results from these simulations merged to recover the overall binding affinity. The approach from Roux et al. [86] does this by computing the potential of mean force along a protein-ligand approach vector, restraining the ligand to a restricted region in each simulation. If, however, multiple slow degrees of freedom must be explicitly dealt with, computation of many-dimensional dimensional PMFs becomes extremely challenging to converge. Waters located in active sites can also possess extremely long correlation times; semigrand canonical approaches can aid convergence by explicitly allowing waters to be created and destroyed through an unphysical route [87]. Similar issues can arise for slow protein side chain degrees of freedom, making explicit decomposition of these degrees of freedom a natural approach to improve sampling [40].

However, what if one does not know in advance the specific degrees of freedom which will lead to slowing of dynamics beyond what can be sampled in reasonable wall clock time? Recent approaches for constructing *Markov state models* (MSMs)—recently reviewed in [88]—suggest a more general scheme for computing expectations in the presence of slow conformational dynamics. To construct these models, numerous short simulations are used to identify metastable conformational states, in which mixing within the region is fast (compared to typical MD simulations) while transitions between the regions are slow, such that the states have lifetimes much longer than can be sampled in typical MD simulations. While the metastable states can be identified by many short parallel simulations [88], recent adaptive schemes allow these to be constructed very efficiently on a computer cluster [89,

90]. By dividing the protein conformation space into these metastable states, binding affinities could be efficiently computed restricted to individual metastable conformations, and the relative state populations (estimated from the interstate transition matrix during the metastable state identification procedure) used to reconstruct the total binding affinity. A first step in this direction was made by Jayachandran et al. [25], who defined MSM states in terms of details involving both protein conformation and ligand conformation. This allowed one to sample the system of interest quickly by starting with docked poses as initial seeds for MD trajectories and then using MSM approaches to combine the trajectories into a single model.

Alternatively, there has been success with applying *generalized ensemble* (GE) methods to protein-ligand free energy calculations. The goal of these methods is to reduce the correlation times by allowing the system to visit multiple alchemical intermediate states in a single simulation, where it is presumed correlation times are much reduced in some intermediate states, such as when the ligand is weakly interacting with its environment. This can be done either in a serial way [91] or a parallel way [92, 44]; further details can be found in a companion review in this issue [93]. However, GE approaches in alchemical space alone may not accelerate receptor conformational changes, and recent results suggest that using temperature to accelerate these transitions may not be especially effective [94, 95]; a possible solution may come from a synergistic combination of GE and MSM methods, such as suggested recently [96], or to explicitly couple in other degrees of freedom [97]. Independent of method specific details, it is clear that simply “waiting and hoping” that sampling will be sufficient will fail in an unknowable subset of challenging problems, which implies that the future of robust free energy prediction rests in some sort of active sampling scheme.

Force fields

With advanced sampling, the community can now address the questions of force field accuracy for some systems. Indeed, there are numerous examples where sampling has led to an improved understanding of the limits of force field accuracy. For example, Shirts et al. [98, 99] used large-scale distributed computing to compute amino acid solvation free energies to high precision, allowing for a direct comparison of force fields. Mobley et al. used hydration free energy calculations to identify and resolve a problem with Lennard-Jones parameters for alkynes, improving agreement with experiment [100]. Garcia and Sanbonmatsu [101] and Sorin and Pande [102] used replica exchange and distributed computing, respectively, to address force field effects in the thermodynamics and kinetics of alpha helices. More recently, Best and Hummer have used converged replica exchange data to compare the behavior of simulated helices to new experimental data in order to improve force field torsion parameters, with compelling initial results indicating its transferability [103].

While there have been several important steps taken to improve additive force fields [104, 105, 103], it is also natural to consider the fact that additive force fields may be inherently environmentally dependent. For example, they are parameterized to the condensed phase and would not be appropriate for gas phase calculations without significant corrections [106]. However, beyond this obvious failing, the power of transferability may be more widely needed, as protein-like environments are very different in dielectric, polarization, and density from aqueous environments, and thus protein-ligand binding affinities may also accumulate inaccuracies due to the neglect of this environmental dependence.

Towards this end, several groups have been working on polarizable force fields. For proteins, the AMOEBA [52] and CHARMM [107] polarizable models are natural examples to consider, with promising initial results [53, 54]. However, these force fields have not been

used as extensively as additive force fields, and thus await more exhaustive tests. Indeed, due to the greater computational costs of these more detailed force fields, efficient sampling methods are more important than ever.

Finally, a key aspect of molecular forces is the nature of the solvation model. Here too, there are numerous choices one can make, with the most common choices being either an explicit representation of the solvent in atomic detail or an implicit (i.e. continuum) representation. While it may be natural to assume that a more detailed model (i.e. explicit solvent) is more accurate, this question is not so simple. Implicit models, such as Poisson Boltzmann (PB) approaches [108] or fast analytical approximations to PB such as Generalized Born (GB) models [109], often include detail missing in many explicit solvent models; in particular, implicit models often include a model for atomic polarization via the dielectric constant of the model and thus may have some advantages of simple (e.g. non-polarizable) explicit models. Recent comparisons between the methods (e.g. [110]) suggests that while explicit solvent can be more accurate, implicit models also do very well, especially in areas of solvation free energies. Few groups, so far, have examined implicit solvent models for use in binding free energy calculations, though early results are encouraging (e.g. [44]).

Analysis

When computing any statistical quantity, such as the binding affinity of a molecule, it is important have both a statistically efficient way to compute the quantity from samples, and a good estimate of what the statistical noise in that quantity is for a given choice of the force field model. This field is relatively well developed. For example, the multistate Bennett acceptance ratio (MBAR) now provides a way to use all simulation data in an optimal way and provide good estimates of the statistical error, provided there are no sampling issues [111, 112], though in most cases the standard pairwise Bennett acceptance ratio will be almost as good. Earlier work has suggested an automatic scheme for detecting the length of the transient equilibration phase, which can be used to control simulation lengths [113]. Thermodynamic integration can be problematic as the bias due to the number of intermediate states must be monitored to avoid numerical error in the calculation, and it is usually less efficient than BAR or MBAR, as well requiring additional effort to implement analytical energy derivatives [114].

4. Outlook for progress

Over the past decade, the field has been extraordinarily productive in generating new algorithmic ideas and advancing technologies to facilitate the development of more accurate forcefields, it has failed to produce an effective set of tools for the design of small molecules. To do so, it is necessary for the field to begin a shift from a research focus to an engineering focus. This shift will require a focus on developing accepted best practices for running calculations, measuring accuracy, and improving methodologies, as well as a clearer plan for how academia and industry work together, both to share data and to find the resources required to develop better drug design tools.

Automated software pipelines

Software tools for automating the preparation of systems using “best practices” methodology are needed. These tools would not only enable the use of alchemical methods by non-experts in academia and pharma, but facilitate high-throughput use and evaluation. With automation, results will be less operator-dependent, allowing meaningful and automated assessment of performance of algorithms and forcefields. To achieve high-throughput, analysis of simulation data must also be automated, and with particular attention paid to diagnostics of convergence problems. Inline diagnostics, in which results are

continually re-evaluated “on the fly,” will also help the simulations adapt to natural correlation times within the system, or signal that it will be impossible to converge the calculations using the desired simulation protocol. Open source tools that can be adapted for different techniques and software tools would also be useful, as there is no single approach that will work for all researchers, all tools, or all systems.

Sensitivity analysis

In addition to a lack of comprehensive assessments of accuracy across multiple systems, there is a lack of literature determining which parameters have significant impact on free energies of binding for a given protein-ligand system. Considering there are numerous reports of changes in experimental conditions affecting measured binding affinities, it would seem that the same should be true of computed binding free energies. By assessing the error incurred in methodically omitting contributions from statistical mechanical effects (e.g. multiple conformations, conformational entropy, receptor flexibility) and chemical effects (e.g. protonation changes or multivalence, tautomerization), the magnitude of these effects can be assessed in highly realistic models of ligand binding, even if current forcefield models are inadequate to quantitatively capture experimental binding affinities to high accuracy. Insights from this effort would be immediately useful in improving existing virtual screening methods, as well as algorithms and forcefields over the longer term, where our limited understanding of these dominant physical determinants of binding is believed to have hindered their improvement [4].

Standardized benchmark sets

Without comprehensive benchmark evaluations, it is impossible to gauge expected predictive accuracy if current techniques were directly applied to problems in drug discovery; as a result, simply attempting to incorporate existing tools into pipelines would be a risky endeavor. To gauge progress toward the goal of deploying a viable engineering tool, it is essential to establish standardized benchmark sets of receptor-ligand systems. These sets should span a range of complexity, from simple targets where quantitative accuracy should be unproblematic to pharmaceutically relevant targets where accuracy is largely unknown at present. To make steady progress on recognized issues, such a set should include a variety of model systems that each introduce a limited number of complications—such as conformational changes upon binding, local unfolding, cryptic binding sites, and protonation state changes—and be collected under uniform, controlled conditions. In addition to existing data, new data must be continually added to the set to avoid overfitting, ensuring that improvements made to deal with known pathological cases can also deal with new ones. While several curated databases of ligand-receptor binding affinities exist, the data generally comes from a variety of laboratories making measurements under different conditions. These almost universally omit assessments of the experimental error, which will be critical in assessing actual improvement versus simply fitting the noise. A recent high-throughput crystallographic screen and biophysical binding assay of a fragment set against trypsin to provide a community dataset is a prime example of efforts that can continue to drive the field forward [115]. Computational benchmark sets with extensive conformational sampling could be used to benchmark novel sampling approaches.

Continual improvement

The ability to explain existing datasets is not sufficient; these are vulnerable to being over-fit by *ad hoc* corrections. Instead, honest evaluation of progress requires continual collection of new data to evaluate improvements, explore where methodologies break down, and discover new phenomena not previously observed. Small-scale realizations of this process have already demonstrated their utility in revealing shortcomings in algorithms and force fields [21, 22]. By organizing the community to make predictions in advance of experiment

through periodic blind challenges (such as the SAMPL challenges [116, 117]), it is possible to continually gauge performance and drive progress on a larger scale. Several pharmaceutical companies have expressed interest in providing datasets from inactive projects, but it is currently unclear what mechanism will provide appropriate incentive to go through the nontrivial process of releasing this data. Instead, by engaging in community-supported efforts in which the burden of these experiments are shared, new data could be obtained on an appropriate community-selected set of targets and timeframes. While it may be difficult to finance the synthesis of complex ligands, intriguing alternatives exist, such as screening existing libraries (e.g. [115]) or mutating the receptor in a high-throughput manner.

Acknowledgments

JDC acknowledges support from a QB3-Berkeley Distinguished Postdoctoral Fellowship. DLM acknowledges the Louisiana Board of Regents Research Competitiveness Subprogram and the Louisiana Optical Network Initiative, supported by the Louisiana Board of Regents Post-Katrina Support Fund Initiative grant LEQSF(2007-12)-ENH-PKSFI-PRS-01. VSP thanks NIH (R01-GM062868) and NSF (EF-0623664) for support. The authors thank Vertex, especially Mark Murcko and Pat Walters, for hosting the workshop which inspired this review, as well as all of the workshop attendees for their insightful contributions.

References

1. Pharmaceutical Research and Manufacturers of America, Pharmaceutical industry profile 2010. Washington, DC: PhRMA; March. 2010
2. Matthieu, MP. Parexel International Corporation, Parexel's Bio/Pharmaceutical R&D Statistical Sourcebook 2008/2009. Chicago: January. 2008
3. Paul SM, Mytelka DS, Dunwiddie CT, Persinger CC, Munos BH, Lindborg SR, Schacht AL. How to improve r&d productivity: The pharmaceutical industry's grand challenge. *Nat Rev Drug Discovery*. 2010; 9, •• A recent analysis of the cost and success rates of various stages of modern drug discovery and development efforts.
4. Schneider G. Virtual screening: an endless staircase? *Nat Rev Drug Discovery*. 2010; 9:273–276., •• An excellent perspective on some of the limitations currently holding back virtual screening methods.
5. Norris G. Boeing's seventh wonder. *IEEE Spectrum*. 1995; 32:20–23.
6. Tembe BL, McCammon JA. Ligand-receptor interactions. *Computers & Chemistry*. 1984; 8:281–283.
7. Steinbrecher T, Mobley DL, Case DA. Nonlinear scaling schemes for lennard-jones interactions in free energy calculations. *J Chem Phys*. 2007; 127(21):214108. [PubMed: 18067350]
8. Pearlman, DA. Chapter 2: Free energy calculations: Methods for estimating ligand binding affinities. In: Reddy, MR.; Erion, MD., editors. *Free Energy Calculations in Rational Drug Design*. Vol. Ch. 2. Kluwer Academic/Plenum Publishers, New York, 233 Spring Street; New York, NY 10013: 2001.
9. Deng Y, Roux B. Computations of standard binding free energies with molecular dynamics simulations. *J Phys Chem B*. 2009; 113(8):2234–2246. [PubMed: 19146384]
10. Aleksandrov A, Thompson D, Simonson T. Alchemical free energy simulations for biological complexes: Powerful but temperamental.... *J Mol Recognit*. 2009; 23:117–127. [PubMed: 19693787]
11. Shirts MR, Mobley DL, Brown SP. Free energy calculations in structure-based drug design. Kenneth DR, Merz M, Reynolds CH. *Structure Based Drug Design*. Cambridge University Press New York, NY 2010, • A recent review of binding free energy calculations covering techniques, applications, and potential uses in drug discovery.
12. Michel J, Essex JW. Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations. *J Comput Aided Mol Des*. 2010; 24:639–658. [PubMed: 20509041] , • A comprehensive review of the state of the field of free energy calculations,

including an excellent discussion of good practices. This should be read by students and others needing an introduction to the practical aspects of these calculations.

13. Christ CD, Mark AE, van Gunsteren WF. Basic ingredients of free energy calculations: A review. *J Comp Chem*. 2010; 31:1569–1582. [PubMed: 20033914] , • A review of alchemical binding free energy calculations focusing primarily on relative free energies, providing a good overview of both methods and key applications.
14. 2010 workshop on free energy methods in drug design, program and talk slides. available at <http://www.alchemistry.org/>
15. Sanz E, Vega C. Solubility of KF and NaCl in water by molecular simulation. *J Chem Phys*. 2007; 126(1):014507. [PubMed: 17212500]
16. Paluch AS, Jayaraman S, Shah JK, Maginn EJ. A method for computing the solubility limit of solids: Application to sodium chloride in water and alcohols. *J Chem Phys*. 2010; 133(12):124504. [PubMed: 20886947]
17. Garrido NM, Queimada AJ, Jorge M, Macedo EA, Economou IG. 1-Octanol/Water partition coefficients of n-Alkanes from molecular simulations of absolute solvation free energies. *J Chem Theory Comput*. 2009; 5(9):2436–2446.
18. Best SA, M KM Jr, Reynolds CH. Free energy perturbation study of octanol/water partition coefficients: Comparison with continuum GB/SA calculations. *J Phys Chem B*. 1999; 103(4):714–726.
19. Jayaraman S, Maginn EJ. Computing the melting point and thermodynamic stability of the orthorhombic and monoclinic crystalline polymorphs of the ionic liquid 1-n-butyl-3-methylimidazolium chloride. *J Chem Phys*. 2007; 127(21):214504. [PubMed: 18067361]
20. Mobley DL, Chodera JD, Dill KA. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J Chem Theory Comput*. 2007; 3:1231–1235. [PubMed: 18843379]
21. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA. Predicting absolute ligand binding free energies to a simple model site. *J Mol Biol*. 2007; 371(4):1118–1134. [PubMed: 17599350]
22. Boyce S, Mobley D, Rocklin G, Graves A, Dill KA, Shoichet BK. Predicting ligand binding affinity with alchemical free energy methods in a polar model *J Mol Biol*. 2009; 394:747–763. [PubMed: 19782087] , •• Demonstrates the predictive accuracy of alchemical free energy methodologies for the T4 lysozyme L99A/M102Q polar site, and highlights a number of sampling challenges for this system, such as protein conformational changes upon binding and slow sampling of multiple binding modes.
23. Steinbrecher T, Case DA, Labahn A. A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase. *J Med Chem*. 2006; 49:1837–1844. [PubMed: 16539369]
24. Oostenbrink C, van Gunsteren WF. Free energies of binding of poly-chlorinated biphenyls to the estrogen receptor from a single simulation. *Proteins*. 2004; 54:237–246. [PubMed: 14696186]
25. Jayachandran G, Shirts MR, Park S, Pande VS. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J Chem Phys*. 2006; 125:084901. [PubMed: 16965051]
26. Lazaridis T, Matsunov A, Gandolfo F. Contributions to the binding free energy of ligands to avidin and streptavidin. *Proteins*. 2002; 47:194–208. [PubMed: 11933066]
27. Graves AP, Brenk R, Shoichet BK. Decoys for docking. *J Med Chem*. 2005; 48:3714–3728. [PubMed: 15916423]
28. Stoll V, Stewart K, Maring C, Muchmore S, Giranda V, Gu Y, Wang G, Chen Y, Sun M, Zhao C. Influenza neuraminidase inhibitors: structure-based design of a novel inhibitor series. *Biochemistry*. 2003; 42:718–727. [PubMed: 12534284]
29. Constantine K, Mueller L, Metzler W, McDonnell P, Todderud G, Goldfarb V, Fan Y, Newitt J, Keifer S, Gao M. Multiple and single binding modes of fragment-like kinase inhibitors revealed by molecular modeling, residue type-selective protonation, and nuclear overhauser effects. *J Med Chem*. 2008; 51:6225–6229. [PubMed: 18771253]

30. Montfort WR, Perry KM, Fauman EB, Finer-Moore JS, Maley GF, Hardy L, Maley F, Stroud RM. Structure, multiple site binding, and segmental accommodation in thymidylate synthase on binding dUMP and an anti-folate. *Biochemistry*. 1990; 29:6964–6977. [PubMed: 2223754]
31. Mobley DL, Dill KA. Binding of small-molecule ligands to proteins: “what you see” is not always “what you get”, *Structure, Folding and Design*. 2009; 17(4):489–498., • This recent review surveys computational and experimental evidence supporting the ideas that multiple ligand orientations may be relevant for binding, and that even small protein conformational changes can be thermodynamically significant.
32. Mobley DL, Chodera JD, Dill KA. On the use of orientational restraints and symmetry corrections in alchemical free energy *J Chem Phys*. 2006; 125:084902. [PubMed: 16965052]
33. Jorgensen WL. Efficient lead discovery and optimization. *Acc Chem Res*. 2009; 42:724–733. [PubMed: 19317443]
34. Holdgate GA, Anderson M, Edfeldt F, Geschwinder S. Affinity-based, biophysical methods to detect and analyze ligand binding to recombinant proteins: Matching high information content with high throughput. *J Struct Biol*. 2010; 172:142–157. [PubMed: 20609391]
35. Brown SP, Muchmore SW, Hajduk PJ. Healthy skepticism: assessing realistic model performance. *Drug Discovery Today*. 2009; 14:420–427. [PubMed: 19340931] , • An insightful look into how predictive models should be realistically assessed against experimental data.
36. Duggleby RG. Determination of inhibition constants, i_{50} values and the type of inhibition for enzyme-catalyzed reactions. *Biochem Med Metabol Bio*. 1988; 40:204–212.
37. Graves AP, Shivakumar DM, Boyce SE, Jacobson MP, Case DA, Shoichet BK. Rescoring docking hit lists for model cavity sites: Predictions and experimental testing. *J Mol Biol*. 2008; 377:914–934. [PubMed: 18280498]
38. Michel J, Essex J. Hit identification and binding mode predictions by rigorous free energy simulations. *J Med Chem*. 2008; 51:6654–6664. [PubMed: 18834104]
39. Steinbrecher T, Hrenn A, Dormann K, Merfort I. Bornyl (3, 4, 5-trihydroxy)-cinnamate—an optimized human neutrophil elastase inhibitor designed by free energy calculations. *Bioorganic & Medicinal Chemistry*. 2008; 16:2385–2390. [PubMed: 18078761]
40. Mobley DL, Chodera JD, Dill KA. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J Chem Theory Comput*. 2007; 3(4): 1231–1235. [PubMed: 18843379]
41. Jiang W, Roux B. Free energy perturbation hamiltonian Replica-Exchange molecular dynamics (FEP/H-REMD) for absolute ligand binding free energy calculations. *Journal of Chemical Theory and Computation*. 2010; 6(9):2559–2565., • This work applies some enhanced sampling techniques to protein conformational changes in the T4 lysozyme model binding site, resulting in improved agreement with experiment relative to earlier work from this group.
42. Deng Y, Roux B. Calculation of standard binding free energies: Aromatic molecules in the t4 lysozyme 199a mutant. *J Chem Theory Comput*. 2006; 2:1255–1273.
43. Clark M, Guarnieri F, Shkurko I, Wiseman J. Grand canonical monte carlo simulation of ligand-protein binding. *J Chem Info Model*. 2006; 46(1):231–242.
44. Gallicchio E, Lapelosa M, Levy RM. Binding energy distribution analysis method (bedam) for estimation of protein-ligand binding affinities. *J Chem Theory Comput*. 2010; 6:2961–2977. [PubMed: 21116484] , •• Replica-exchange alchemical binding free energy calculations in implicit-solvent demonstrate both the ability to overcome sampling limitations and good agreement with experiment for T4 lysozyme model cavity systems.
45. Shirts, MR. PhD dissertation. Stanford; January. 2005 Calculating precise and accurate free energies in biomolecular systems.
46. Wang J, Deng Y, Roux B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophysical journal*. 2006; 91(8):2798–2814. [PubMed: 16844742]
47. Fujitani H, Tanida Y, Matsuura A. Massively parallel computation of absolute binding free energy with well-equilibrated states. *Phys Rev E*. 2009; 79(2):21914.
48. Ytreberg F. Absolute fkbp binding affinities obtained via nonequilibrium unbinding simulations. *J Chem Phys*. 2009; 130:164906. [PubMed: 19405629]

49. Shirts MR, Mobley DL, Chodera JD, Pande VS. Accurate and efficient corrections for missing dispersion interactions in molecular simulations. *J Phys Chem B*. 2007; 111:13052–13063. [PubMed: 17949030]
50. Talhout R, Villa A, Mark AE, Engberts JBFN. Understanding binding affinity: A combined isothermal titration calorimetry/molecular dynamics study of the binding of a series of hydrophobically modified benzamidium chloride inhibitors to trypsin. *J Am Chem Soc*. 2003; 125:10570–10579. [PubMed: 12940739]
51. Villa A, Zangi R, Pieffet G, Mark AE. Sampling and convergence in free energy calculations of protein-ligand interactions: the binding of triphenoxypyridine derivatives to factor xa and trypsin. *J Comput Aided Mol Des*. 2003; 23:673–686. [PubMed: 15068366]
52. Ponder JW, Wu C, Ren P, Pande VS, Chodera JD, Schneiders MJ, Haque I, Mobley DL, Lambrecht DS, DiStasio RA Jr, Head-Gordon M, Clark GNI, Johnson ME, Head-Gordon T. Current status of the AMOEBA polarizable force field. *J Phys Chem B*. 2010; 114:2549–2564. [PubMed: 20136072]
53. Jiao D, Golubkov PA, Darden TA, Ren P. Calculation of protein-ligand binding free energy by using a polarizable potential. *Proc Nat Acad Sci USA*. 2008; 105(17):6290–6295. [PubMed: 18427113]
54. Jiao D, Zhang J, Duke RE, Li G, Schnieders MJ, Ren P. Trypsin-ligand binding free energies from explicit and implicit solvent simulations with polarizable potential. *J Comp Chem*. 2009; 30(11): 1701–1711. [PubMed: 19399779]
55. Ge X, Roux B. Absolute binding free energy calculations of spar-somycin analogs to the bacterial ribosome. *The Journal of Physical Chemistry B*. 2010; 114:9525–9539. [PubMed: 20608691]
56. Zhao C, Caplan DA, Noskov SY. Evaluations of the absolute and relative free energies for antidepressant binding to the amino acid membrane transporter LeuT with free energy simulations. *Journal of Chemical Theory and Computation*. 2010; 6(6):1900–1914.
57. Zeevaart JG, Wang L, Thakur VV, Leung CS, Tirado-Rives J, Bailey CM, Domaoal RA, Anderson KS, Jorgensen WL. Optimization of azoles as Anti-Human immunodeficiency virus agents guided by Free-Energy calculations. *Journal of the American Chemical Society*. 2008; 130(29):9492–9499. [PubMed: 18588301]
58. Leung CS, Zeevaart JG, Domaoal RA, Bollini M, Thakur VV, Spasov KA, Anderson KS, Jorgensen WL. Eastern extension of azoles as non-nucleoside inhibitors of HIV-1 reverse transcriptase; cyano group alternatives. *Bioorganic & Medicinal Chemistry Letters*. 2010; 20(8): 2485–2488. [PubMed: 20304641]
59. Leung SS, Tirado-Rives J, Jorgensen WL. Vancomycin resistance: Modeling backbone variants with d-Ala-d-Ala and d-Ala-d-Lac peptides. *Bioorganic & Medicinal Chemistry Letters*. 2009; 19(4):1236–1239. [PubMed: 19128968]
60. Aleksandrov A, Proft J, Hinrichs W, Simonson T. Protonation patterns in tetracycline:Tet repressor recognition: Simulations and experiments. *ChemBioChem*. 2007; 8:675–685. [PubMed: 17361981]
61. Banba S, Guo Z, B CL III. Efficient sampling of ligand orientations and conformations in free energy calculations using the [lambda]-dynamics method. *J Chem Phys*. 2000; 104:69036910.
62. Kastholz M, Hünenberger P. Computation of methodology-independent ionic solvation free energies from molecular simulations. i. the electrostatic potential in molecular liquids. *J Chem Phys*. 2006; 124:124106. [PubMed: 16599661]
63. Kastholz M, Hünenberger P. 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. [PubMed: 16784292]
64. Dolenc J, Oostenbrink C, Koller J, Gunsteren WV. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an aaaa dna binding site. *Nucleic acids research*. 2005; 33(2):725. [PubMed: 15687382]
65. Michel J, Verdonk M, Essex J. Protein-ligand binding affinity predictions by implicit solvent simulations: a tool for lead optimization? *J Med Chem*. 2006; 49(25):7427–7439. [PubMed: 17149872]

66. Pohorille A, Jarzynski C, Chipot C. Good practices in free-energy calculations. *J Phys Chem B*. 2010; 114:10235–10253. [PubMed: 20701361] , • An excellent review of many of the sources and manifestations of statistical error in alchemical free energy calculations.
67. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanics calculations. *J Mol Graph Model*. 2006; 25:247–260. [PubMed: 16458552]
68. Shenfeld DK, Xu H, Eastwood MP, Dror RO, Shaw DE. Minimizing thermodynamic length to select intermediate states for free-energy calculations and replica-exchange simulations. *Phys Rev E*. 2009; 80:046705., •• This work lays the foundation for a theory of how to choose optimal alchemical intermediates for free energy calculations.
69. Czodrowski P, Sotriffer CA, Klebe G. Protonation changes upon ligand binding to trypsin and thrombin: Structural interpretation based on pKa calculations and ITC experiments. *J Mol Biol*. 2007; 367:1347–1356. [PubMed: 17316681]
70. Martin YC. Let's not forget tautomers. *J Comput Aid Mol Des*. 2009; 23:693–704., • A reminder of the importance of tautomers in molecular design.
71. Mongan J, Case DA. Biomolecular simulation at constant pH. *Curr Opin Struct Biol*. 2005; 15:157–163. [PubMed: 15837173]
72. Stern HA. Molecular simulation with variable protonation states at constant pH. *J Chem Phys*. 2007; 126:164112. [PubMed: 17477594]
73. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008; 4:435–447.
74. Klimovich PV, Mobley DL. Predicting hydration free energies using all-atom molecular dynamics simulations and multiple starting conformations. *J Comput Aid Mol Des*. 2010; 24:307.
75. Shaw DE. A fast, scalable method for the parallel evaluation of distance-limited pairwise particle interactions. *J Comp Chem*. 2005; 26(13):1318–1328. [PubMed: 16013057]
76. Bowers KJ, Dror RO, Shaw DE. Zonal methods for the parallel execution of range-limited n-body simulations. *J Comp Phys*. 2007; 221(1):303–329.
77. Chow, E.; Rendleman, CA.; Bowers, KJ.; Dror, RO.; Hughes, DH.; Gullingsrud, J.; Sacerdoti, FD.; Shaw, DE. Tech Rep TR–2008-01. D. E. Shaw Research; July. 2008 Desmond performance on a cluster of multicore processors.
78. Hess B, Kutzner C, van der Spoel D, Lindahl E. GROMACS 4: Algorithms for highly efficient, load-balanced, and scalable molecular simulation. *J Chem Theory Comput*. 2008; 4:435–447.
79. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable molecular dynamics with NAMD. *J Comput Chem*. 2005; 26:1781–1802. [PubMed: 16222654]
80. Toyoda S, Miyagawa H, Kitamura K, Amisaki T, Hashimoto E, Ikeda H, Kusumi A, Miyakawa N. Development of MD Engine: High-speed accelerator with parallel design for molecular dynamics simulations. *J Comput Chem*. 1999; 20:185–189.
81. Shaw DE, Deneroff MM, Dror RO, Kuskin JS, Larson RH, Salmon JK, Young C, Batson B, Bowers KJ, Chao JC, Eastwood MP, Gagliardo J, Grossman JP, Ho CR, Ierardi DJ, Kolossvy I, Klepeis JL, Layman T, McLeavey C, Moraes MA, Mueller R, Priest EC, Shan Y, Spengler J, Theobald M, Towles B, Wang SC. Anton, a special-purpose machine for molecular dynamics simulation. *Commun ACM*. 2008; 51(7):91–97.
82. Eastman P, Pande VS. Efficient nonbonded interactions for molecular dynamics on a graphics processing unit. *J Comp Chem*. 2010; 31(6):1268–1272. [PubMed: 19847780]
83. Friedrichs MS, Eastman P, Vaidyanathan V, Houston M, Legrand S, Beberg AL, Ensign DL, Bruns CM, Pande VS. Accelerating molecular dynamic simulation on graphics processing units. *Journal of Computational Chemistry*. 2009; 30(6):864–872. [PubMed: 19191337]
84. Stone JE, Hardy DJ, Ufimtsev IS, Schulten K. GPU-accelerated molecular modeling coming of age. *J Mol Graph Model*. 2010; 29:116–125. [PubMed: 20675161]
85. Harvey MJ, Giupponi G, Fabritiis GD. ACEMD: Accelerating biomolecular dynamics in the microsecond time scale. *Journal of Chemical Theory and Computation*. 2009; 5(6):1632–1639.
86. Ge X, Roux B. Calculation of the standard binding free energy of spar-somycin to the ribosomal peptidyl-transferase P-site using molecular dynamics simulations with restraining potentials. *J Mol Recognit*. 2010; 23:128–141. [PubMed: 20151411]

87. Deng Y, Roux B. Computation of binding free energy with molecular dynamics and grand canonical Monte Carlo simulations. *J Chem Phys.* 2008; 128:115103. [PubMed: 18361618]
88. Pande VS, Beauchamp K, Bowman GR. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods.* 2010; 52:99–105. [PubMed: 20570730]
89. Singhal N, Pande VS. Error analysis and efficient sampling in Markovian state models for molecular dynamics. *J Chem Phys.* 2005; 123:204909. [PubMed: 16351319]
90. Bowman GR, Ensign DL, Pande VS. Enhanced modeling via network theory: Adaptive sampling of markov state models. *Journal of Chemical Theory and Computation.* 2010; 6(3):787–794.
91. Li H, Fajer M, Yang W. Simulated scaling method for localized enhanced sampling and simultaneous “alchemical” free energy simulations: A general method for molecular mechanical, quantum mechanical, and quantum mechanical/molecular mechanical simulations. *J Chem Phys.* 2007; 126:024106. [PubMed: 17228942]
92. Jiang W, Hodoscek M, Roux B. Computation of absolute hydration and binding free energy with free energy perturbation distributed Replica-Exchange molecular dynamics. *Journal of Chemical Theory and Computation.* 2009; 5(10):2583–2588.
93. Gallicchio E, Levy RM. Advances in all atom sampling methods for modeling protein-ligand binding affinities. *Current Opinion in Structural Biology.*
94. Rhee YM, Pande VS. Multiplexed-replica exchange molecular dynamics method for protein folding simulation. *Biophys J.* 2003; 84:775–786. [PubMed: 12547762]
95. Zheng W, Andrec M, Gallicchio E, Levy RM. Simulating replica exchange simulations of protein folding with a kinetic network model. *Proc Natl Acad Sci U S A.* 2007; 104:15340–15345. [PubMed: 17878309]
96. Huang X, Bowman GR, Bacallado S, Pande VS. Rapid equilibrium sampling initiated from nonequilibrium data. *Proc Natl Acad Sci U S A.* 2009; 106:19765–19769. [PubMed: 19805023]
97. Zheng L, Chen M, Yang W. Random walk in orthogonal space to achieve efficient free-energy simulation of complex systems. *Proc Natl Acad Sci USA.* 2008; 105:20227–20232. [PubMed: 19075242]
98. Shirts MR, Pitner JW, Swope WC, Pande VS. Extremely precise free energy calculations of amino acid side chain analogs: Comparison of common molecular mechanics force fields for proteins. *J Chem Phys.* 2003; 119(11):5740–5761.
99. Shirts MR, Pande VS. Solvation free energies of amino acid side chains for common molecular mechanics water models. *J Chem Phys.* 2005; 122:134508. [PubMed: 15847482]
100. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA. Small molecule hydration free energies in explicit solvent: An extensive test of fixed-charge atomistic simulations. *J Chem Theory Comput.* 2009; 5(2):350–358. [PubMed: 20150953]
101. Garcia AE, Sanbonmatsu KY. Alpha-helical stabilization by side chain shielding of backbone hydrogen bonds. *Proc Natl Acad Sci U S A.* 2002; 99:2782–2787. [PubMed: 11867710]
102. Sorin EJ, Pande VS. Exploring the helix-coil transition via all-atom equilibrium ensemble simulations. *Biophys J.* 2005; 88:2472–2493. [PubMed: 15665128]
103. Best RB, Hummer G. Optimized molecular dynamics force fields applied to the helix-coil transition of polypeptides. *J Phys Chem B.* 2009; 113:9004–9015. [PubMed: 19514729]
104. Hornak V, Abel R, Okur A, Strockbine B, Roitberg A, Simmerling C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins.* 2006; 65(3):712–25. [PubMed: 16981200]
105. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE. Improved side-chain torsion potentials for the Amber ff99SB protein force field. *Proteins.* 2010; 78:1950–1958. [PubMed: 20408171]
106. Swope WC, Horn HW, Rice JE. Accounting for polarization when using fixed charge force fields. II. method and application for computing effect of polarization cost on free energy of hydration. *J Phys Chem B.* 2010; 114:8631–8645. [PubMed: 20540502]
107. Baker CM, Lopes PEM, Zhu X, Roux B, Alexander J, MacKerell D. Accurate calculation of hydration free energies using pair-specific Lennard-Jones parameters in the CHARMM Drude polarizable force field. *J Chem Theory Comput.* 2010; 6:1181–1198. [PubMed: 20401166]

108. Sharp KA, Nicholls A, Friedman R, Honig B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry*. 1991; 30:9686–9697. [PubMed: 1911756]
109. Still WC, Tempczyk A, Hawley RC, Hendrickson T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J Am Chem Soc*. 1990; 112:6127–6129.
110. Nicholls A, Mobley DL, Guthrie JP, Chodera JD, Bayly CI, Cooper MD, Pande VS. Predicting small-molecule solvation free energies: an informal blind test for computational chemistry. *J Med Chem*. 2008; 51:769–779. [PubMed: 18215013]
111. Shirts MR, Chodera JD. Statistically optimal analysis of samples from multiple equilibrium states. *J Chem Phys*. 2008; 129:124105. [PubMed: 19045004] , •• This paper describes a general, optimal replacement for previous methods of analyzing alchemical free energy calculations (such as BAR and WHAM) along with reliable estimates of statistical error.
112. Fajer M, Swift R, McCammon JA. Using multistate free energy techniques to improve the efficiency of replica exchange accelerated molecular dynamics. *J Comput Chem*. 2009; 30:1719–1725. [PubMed: 19421994]
113. Yang W, Bitetti-Putzer R, Karplus M. Free energy simulations: Use of reverse cumulative averaging to determine the equilibrated region and the time required for convergence. *J Chem Phys*. 2004; 120:2618–2628. [PubMed: 15268405]
114. Shirts MR, Pande VS. Comparison of efficiency and bias of free energies computed by exponential averaging, the bennett acceptance ratio, and thermodynamic integration. *J Chem Phys*. 2005; 122:144107. [PubMed: 15847516]
115. Newman J, Fazio VJ, Caradoc-Davies TT, Branson K, Peat TS. Practical aspects of the SAMPL challenge: Providing an extensive experimental data set for the modeling community. *J Biomol Screen*. 2009; 14:1245–1250. [PubMed: 19822883]
116. Guthrie JP. A blind challenge for computational solvation free energies: Introduction and overview. *J Phys Chem B*. 2009; 113:4501–4507. [PubMed: 19338360]
117. Skillman AG, Geballe MT, Nicholls A. SAMPL2 challenge: prediction of solvation energies and tautomer ratios. *J Comput Aided Mol Des*. 2010; 24:257–258. [PubMed: 20425136]