J Comput Aided Mol Des (2010) 24:639–658
DOI 10.1007/s10822-010-9363-3

PERSPECTIVE

Prediction of protein–ligand binding affinity by free energy simulations: assumptions, pitfalls and expectations

Julien Michel · Jonathan W. Essex

Received: 18 January 2010 / Accepted: 3 May 2010 / Published online: 28 May 2010

Abstract Many limitations of current computer-aided drug design arise from the difficulty of reliably predicting the binding affinity of a small molecule to a biological target. There is thus a strong interest in novel computational methodologies that claim predictions of greater accuracy than current scoring functions, and at a throughput compatible with the rapid pace of drug discovery in the pharmaceutical industry. Notably, computational methodologies firmly rooted in statistical thermodynamics have received particular attention in recent years. Yet free energy calculations can be daunting to learn for a novice user because of numerous technical issues and various approaches advocated by experts in the field. The purpose of this article is to provide an overview of the current capabilities of free energy calculations and to discuss the applicability of this technology to drug discovery.

Keywords Structure-based drug design · Protein–ligand binding affinity · Free energy calculations · Molecular simulations

J. Michel (✉)
Institute of Structural and Molecular Biology, The University of Edinburgh, Edinburgh EH9 3JR, UK
e-mail: mail@julienmichel.net

J. Michel
Department of Chemistry, Yale University, New Haven, CT 06520, USA

J. W. Essex
School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Introduction

Structure-based drug design has become an established paradigm to search for novel drug-like molecules able to bind to biomolecules of important pharmaceutical interest [1]. Various computational chemistry techniques are often used to assist the structure based drug design process [2]. A typical example is the virtual screen by docking, whereby the contents of databases of drug-like molecules are docked into the binding site of a protein of interest. Plausible modes of interaction between the ligand and its receptor are generated and the binding affinity of the resulting complex is estimated [1, 3, 4]. The low costs and speed at which docking studies can be conducted make the procedure valuable to prioritize compounds to assay in a drug discovery program, or to assist the design of improved ligands.

Because of the size of most databases of drug-like molecules, docking programs must be able to process efficiently thousands of compounds per day. This speed requirement limits drastically the complexity of the algorithms used to predict the binding mode and affinity of each small molecule. If the conformation of the binding site is known, modern docking programs can often predict the binding mode of a ligand in the cognate complex to within 2–3 Å RMSD, [5, 6] and progress is made in acknowledging and accounting for small conformational changes of the receptor upon binding [7–10].

Unfortunately, obtaining accurate estimates of binding affinities from docked ligands has proven much more difficult. In general, scoring functions based on a single ligand pose/single receptor conformation have not proven sufficiently accurate for individual predictions to be interpreted with confidence. As a result, successful scoring functions are mainly used to enrich the top of a database with active

compounds, such that novel active compounds would be identified more rapidly by assaying the top ranked compounds rather than randomly selecting compounds from the database [6]. There are many aspects of drug discovery that could be improved if binding affinities could be predicted more accurately. Some examples include: lead optimization where small modifications to a given scaffold are attempted in order to optimize affinity and drug-like properties; the selectivity profile of a ligand targeting a member of a family of related proteins; the identification of targets for a ligand with an unknown mode of action.

The physics of ligand binding can be understood at a microscopic level with statistical thermodynamics [11, 12]. The resulting equations that govern the free energy of binding are complex and their solutions can be estimated numerically through computer simulations. The free energy calculation methodology has attracted interest because of its formal rigor and demonstrated accuracy in a number of simpler problems such as the prediction of free energies of solvation or the prediction of the binding affinity of small host/guest systems. However, while the first applications of free energy simulations have been reported over two decades ago, the technology had until now a limited impact on drug discovery. While the approach is in principle accurate, a number of approximations have to be taken to obtain predictions at reasonable pace. Thus, the methodology has often drawn criticisms because of its limited scope, high computational cost, difficult and time consuming setup and analysis of results. Therefore, a legitimate question is when and where would it be most useful to make use of free energy simulations to assist drug design?

The purpose of this article is to present the current methodologies of free energy calculations and the common pitfalls and issues a novice user would be likely to face in trying to use these techniques to predict binding affinities. In our opinion, successful application of free energy simulations to a binding free energy study is not trivial because of the large number of technical issues that are encountered. It is often unclear to a novice user which particular approach should be adopted, in part because of a lack of consensus among experts. For instance, and to name just a few examples, in recent years free energy studies have been conducted using relative or absolute binding free energy calculations, [13, 14] explicit or implicit solvent models,[15, 16] different coupling schemes, [17] different treatment of electrostatic interactions, [14, 18] Monte Carlo or molecular dynamics sampling methodologies [19, 20].

Therefore the emphasis of this article will be to provide an up to date overview of current methodologies. The article will be concerned only with so called "rigorous" free energy calculation methodologies. By rigorous we mean methodologies that are theoretically exact (according to statistical thermodynamics) and only make approximations in their implementation. Alternative popular techniques such as the Linear Interaction Energy (LIE) or Molecular Mechanics Poisson Boltzmann (MM-PBSA) will not be covered and for which we refer the reader to other reviews available in the literature [21, 22].

"Overview of the recent literature" will present work that has been reported in the literature in recent years. This section is not meant to be exhaustive but rather to inform about the current capabilities of the field. The key equations that allow computation of binding affinities from statistical thermodynamics as free energy changes will be next introduced in "Statistical thermodynamics expressions for the computation of binding affinities". In "Computation of free energy changes", the main techniques to compute these free energy changes will be presented. "Implementation of free energy calculations" will cover the issues encountered in implementing free energy calculations on a computer. The interpretation and analysis of the results of a binding free energy calculation are discussed in "Quality and reliability of results". Finally, "Realistic expectations: which approach will work best?" will provide guidelines and recommendations to help decide whether and how to conduct a binding free energy study. The review will conclude with a perspective on the main challenges to address for the field in the coming years.

Overview of the recent literature

Free energy calculations have been used for over two decades to rationalise binding affinities. Because of limitations in computing power, early studies were typically concerned with a very small number of compounds and restricted to the computation of relative binding affinities for structurally similar compounds. In the last 5 years, advances in protocols and computing power have allowed the study of larger ligand datasets and the computation of absolute binding affinities for structurally diverse ligands. Blind predictions and rational ligand design guided by free energy calculations have also become more common.

Deng and Roux computed the absolute free energy of binding of 11 ligands of the T4 Lyzozyme L99A mutant. Agreement with experiment was good for smaller aromatic ligands, but large deviations of the computed binding affinities were observed in some cases for larger ligands, and the computed binding affinity was sensitive to the choice of the initial protein structure [18]. Mobley et al. have shown that the conformation of the side chain of Valine 111 in this protein varies between apo and holo structures and proposed a method to account for the contribution of this protein conformational change to the binding affinity of the ligands [23]. Using their own protocols, Mobley et al. then reported a RMS error of

1.9 kcal mol^{-1} for predicting the absolute free energy of binding of 11 known ligands and 2 non binders of T4 Lyzozyme L99A mutant. This was followed by a series of blind predictions for 5 new ligands, 4 compounds were predicted to bind and one to be a decoy. These results were confirmed by thermal denaturation and ITC assays. Interestingly, docking calculations failed to discriminate the decoy from the binders [14]. This work was recently extended to the T4 Lyzozyme L99A/M102Q mutant where the additional mutation introduces hydrogen bonding interactions in the binding site and permits binding of polar ligands. The binding affinities of seven known ligands and two known decoys were first computed a posteriori. A RMS error of 1.2 kcal mol^{-1} was achieved. The rank ordering by predicted potency was good but not perfect: 6 of the 7 ligands were correctly identified as binders but one decoy was predicted to bind more strongly than a known binder. This was followed by blind predictions for several new compounds using a mixture of absolute and relative binding free energy calculations. Though quantitative agreement was not achieved, 11 out of 14 compounds were correctly identified as binder or decoys from their computed absolute free energy of binding [24]. Fujitani et al. computed the absolute binding affinity of 8 ligands of the protein FKBP. These ligands were more typical of complex drug-like small molecules, and adequate sampling of the protein and ligand degrees of freedom presented a computational challenge that was addressed through the use of a dedicated supercomputer. Though the computed binding affinities systematically overestimated the measured binding affinities, they were highly correlated and could be fitted to the measurements with a RMS error of 0.4 kcal mol^{-1} only by shifting the calculated affinities by a constant offset of 3.2 kcal mol^{-1} [15]. In a more recent study using a different protocol, Tanida et al. computed the absolute free energy of binding for 6 ligands bound to a RNA aptamer. The computed binding affinities were seen to again systematically overestimate the measured affinities by a constant offset, but were otherwise highly correlated [25]. Oostenbrink and van Gunsteren developed an efficient methodology to compute from a single simulation the relative binding affinities of 17 closely related biphenyls ligand of the oestrogen receptor alpha. They achieved a mean unsigned error of about 1 kcal mol^{-1} [26]. Michel et al. compared the ability of different free energy calculation protocols and empirical functions to rank-order by potency 10 inhibitors of cyclo-oxygenase-2, 10 inhibitors of neuraminidase, and 18 inhibitors of cyclin-dependent kinase-2. The relative binding affinity of the cyclo-oxygenase-2 inhibitors could be reproduced to within a mean error of ca. 1 kcal mol^{-1} and excellent rank-ordering was achieved. Though there were outliers which precluded quantitative predictions of binding affinities, excellent rank-ordering of the neuraminidase ligands was achieved by the free energy methods. These predictions were more accurate than those achieved using empirical scoring functions. However, none of the techniques considered achieved meaningful correlation for the dataset of 18 cyclin-dependent kinase-2 inhibitors [16]. Another study of Michel and Essex contrasted the ability of different free energy calculation protocols and scoring functions to discriminate binders and non binders in a dataset of 16 structurally diverse ligands of the oestrogen receptor alpha. The most rigorous and time-consuming simulation protocols could correctly identify 5 of the 6 known binders, but the accuracy degraded when simpler protocols were used [19].

The preceding body of work was mainly concentrated on methodological efforts, but a number of workers have started using free energy calculations to guide structure-based drug design efforts. Steinbrecher et al. used thermodynamic integration to design an improved neutrophil elastate inhibitor based on a bornyl caffeate scaffold. The improved analogue was predicted to be 500 times more potent than the parent compound. Subsequent organic synthesis and assaying revealed that the improved analogue was indeed more potent, but by three fold only [27]. Pan et al. used grand canonical ensemble free energy calculations to improve by ca. 10 fold the potency of a small molecule inhibitor of the HIV-1/Tat and p300/CREB protein–protein interaction [28]. The Jorgensen lab has reported the use of free energy calculations to guide the optimization of inhibitors since 2006. Efforts have focused on the development of novel non-nucleoside inhibitors (NNRTIs) of HIV-1 reverse transcriptase. An initial 30 μM diarylamine lead was optimized to achieve an activity of 10 nM in a cell assay [29, 30]. Extensive free energy calculations coupled with medicinal chemistry efforts have subsequently led to the development of pyrimidinyl, [31] triazinyl-amines, [31] bicyclic heterocycles, [32] and azoles, [33] as novel potent NNRTIs with improved pharmacological properties. This body of work has been recently reviewed [34]. Free energy calculations have also been used in conjunction with empirical scoring functions to computationally screen a library of β-peptides inhibitors and identify improved dual inhibitors of the protein–protein interactions p53/hDM2 and p53/hDMX [35].

Statistical thermodynamics expressions for the computation of binding affinities

Definition of a relative binding free energy

The reversible binding of two ligands L1 and L2 to a protein P can be written as:

$$P + L1 \overset{\Delta G^0_{\text{bind}}(L1)}{\rightarrow} PL1 \tag{1}$$

$$P + L2 \overset{\Delta G^0_{\text{bind}}(L2)}{\rightarrow} PL2 \tag{2}$$

Where $\Delta G^0_{\text{bind}}(L1)$ is the standard binding free energy of ligand L1. Subtracting (2) from (1) yields an expression for the relative binding free energy $\Delta\Delta G_{\text{bind}}$ of L1 and L2.

$$L1 + PL2 \overset{\Delta\Delta G_{\text{bind}}}{\rightarrow} PL1 + L2 \tag{3}$$

The free energy change of this reaction is related to the equilibrium constant K_{eq}.

$$\Delta\Delta G_{\text{bind}} = -kT \ln K_{\text{eq}} = -kT \ln \frac{[PL1][L2]}{[L1][PL2]} \tag{4}$$

where the brackets denote concentrations. This quantity can be computed by a statistical thermodynamics approach, [11]

$$\begin{aligned}\Delta\Delta G_{\text{bind}} &= -kT \ln \frac{Z_{PL1,\text{solv}} Z_{L2,\text{solv}}}{Z_{L1,\text{solv}} Z_{PL2,\text{solv}}} \\ &= (-kT \ln \frac{Z_{PL1,\text{solv}}}{Z_{PL2,\text{solv}}}) - (-kT \ln \frac{Z_{L1,\text{solv}}}{Z_{L2,\text{solv}}}) \\ &= \Delta G(PL2 \rightarrow PL1) - \Delta G(L2 \rightarrow L1)\end{aligned} \tag{5}$$

where configuration integrals Z are shown instead of partition functions and pressure dependent terms are not shown for simplicity. $Z_{\text{L1,solv}}$ represents the configuration integral of the solvated ligand L1, $Z_{\text{PL1,solv}}$ the configuration integral of the solvated protein–ligand complex PL1. ΔG(PL2→PL1) is the free energy change for converting L2 into L1 in the protein and ΔG(L2→L1) the free energy change for converting L2 into L1 in the solvent. Equation 5 shows that the relative binding free energy is the difference of two ratios of configuration integrals. Expressions for the configuration integrals are given below (Eq. 11).

Definition of an absolute binding free energy

It is also possible to derive an expression for the standard binding free energy of ligand L1 [36].

$$\Delta G^0_{\text{bind}}(L1) = -kT \ln \frac{Z_{PL1,\text{solv}} Z_{\text{solv}} V}{Z_{L1,\text{solv}} Z_{P,\text{solv}} V_0} \tag{6}$$

where Z_{solv} represents the configuration integral of the solvent; V_0 the standard volume, and V the volume available to one molecule of bound ligand which depends on the way the absolute binding free energy calculation is implemented (see "Additional considerations for absolute binding free energy calculations"). Equation 6 depends on the choice of a standard state volume, in general 1 $\text{dm}^3\ \text{mol}^{-1}$ for a solute.

One of the most commonly used approaches to compute a standard binding free energy is to use the double decoupling methodology (see ref [36] for a description of the technique and refs [18, 37] for various implementations) where the interactions of the ligand with the surrounding medium are coupled/decoupled (turned on/off) in different stages. In this technique, Eq. 6 is computed as a sum of 3 equations. First the ligand L1 is converted from an ideal state, i.e. it is not interacting with the solvent or protein, to an "ideal-constrained" state (described by the symbol ∘∘ in Eqs 7 and 8) where it is constrained to stay in the vicinity of the protein P (Eq. 7 and $\Delta G_{\text{constraint}}$(L1) term in Eq. 10). This "ideal-constrained" state does not have a clearly interpretable physical meaning but should be seen as a convenient intermediate for the computations. In the absence of constraints, the non interacting ligand would drift off from the binding site, which dramatically slows down the convergence of the computed free energy changes. Next the ideal-constrained ligand is converted to a fully interacting ligand and the constraints are released (Eq. 8 and ΔG_{dec}(L1) term in Eq. 10). Finally, the solvated ligand L1 is converted into a non interacting ligand (Eq. 9 and $-\Delta G_{\text{solv}}$(L1) term in Eq. 10). The free energy change corresponds to the negative of the solvation free energy of L1 [38]

$$P + L1^{\text{id}} \rightarrow P \circ\circ L1^{\text{id}} \tag{7}$$

$$P \circ\circ L1^{\text{id}} \rightarrow PL1 \tag{8}$$

$$L1 \rightarrow L1^{\text{id}} \tag{9}$$

The absolute binding free energy can be written in term of configuration integrals as:

$$\begin{aligned}\Delta G^0_{\text{bind}}(L1) &= (-kT \ln \frac{Z_{P\circ\circ L1^{\text{id}},\text{solv}}}{Z_{P,\text{solv}} Z_{L1,\text{ideal}}}) + (-kT \ln \frac{Z_{PL1,\text{solv}}}{Z_{P\circ\circ L1^{\text{id}},\text{solv}}}) \\ &\quad - (-kT \ln \frac{Z_{L1,\text{solv}}}{Z_{L1,\text{ideal}} Z_{\text{solv}}}) \\ &= \Delta G_{\text{constraint}}(L1) + \Delta G_{\text{dec}}(L1) - \Delta G_{\text{solv}}(L1)\end{aligned} \tag{10}$$

Computation of free energy changes

The previous section has made apparent that the computation of relative (Eq. 5) or absolute (Eq. 10) binding affinities require the determination of a series of free energy changes ΔG. These free energy changes are related to configuration integrals, for instance the configuration integral $Z_{\text{PL1,solv}}$ is

$$Z_{PL1,\text{solv}} = \int dr^{\text{dofs}} \exp(-\beta U(r)) \tag{11}$$

where dofs denote the internal degrees of freedom of the molecules of solvent, the protein P, the ligand L1 and rotational and translational degrees of freedom, U the potential energy function, and β is the inverse temperature, i.e. $\beta = 1/k_{\text{b}}T$ where k_{b} is the Boltzmann constant. Owing to the large number of degrees of freedom in a protein–

ligand complex, the dimensionality of Eq. 11 is high and further, the functional form of U(r) is complicated. Consequently a direct evaluation of eq 11 is usually impossible. However, it is much easier to solve eq 5 or eq 10 by computing ratios of related partition functions (for example $Z_{PL1,solv}/Z_{PL2,solv}$ or equivalently ΔG(PL2→PL1)) and most free energy calculation techniques take this route.

Methods based on perturbation formulae

A common approach relies on the use perturbation theory to compute free energy changes by exponential averaging. The technique was introduced by Zwanzig in 1954 [39, 40].

$$\Delta G^{\mathrm{EXP}}(PL1 \rightarrow PL2) = -kT \ln \langle \exp(-\beta[U_{PL2}(r) - U_{PL1}(r)]) \rangle_{PL1} \quad (12)$$

The angular brackets <> in Eq. 12 indicate that the bracketed quantity is averaged over all the configurations PL1 can adopt and weighted by their Boltzmann probabilities. In other words Eq. 12 shows that the free energy difference between PL2 and PL1 can be computed as the exponential of the potential energy difference between PL2 and PL1 weighted by the probability distribution of the possible configurations of PL1. In practice this probability distribution is seldom directly calculated, but instead the configurations of PL1 are generated using a molecular simulation of some type, with the appropriate Boltzmann weights. A simple average of the exponential term is then performed over these configurations.

A recurring problem with Eq. 12 is the rather slow convergence of the ensemble average. The issue is that the configurations that contribute the most to this ensemble average are those where the energy of PL2 is lower than PL1, yet they could be observed infrequently during a simulation of PL1 if the two ligands L1 and L2 differ too much in their structure and phase-space overlap. An interesting consequence is that the computed free energy change ΔG^{EXP}(PL2→PL1) is usually not merely the opposite of ΔG^{EXP}(PL1→PL2). In fact, it has been known for some time that two free energy calculations performed in forwards and backwards directions provide two different values of a free energy change that bracket the correct value [41]. Thus an old strategy to estimate the precision of a free energy change is to compute the free energy in both directions. One has to be cautious not to conclude that the correct free energy change can be obtained from the average of forwards and backwards run because depending on the type of perturbation, one of the two values may be fairly close to the correct free energy change. This happens because there is an asymmetry in the convergence properties of Eq. 12, i.e. it can be advantageous to select PL1 or PL2 as a reference state. It has been argued that the reference state should be the state of higher entropy because the low energy configurations of the perturbed state are more likely to be a subset of the low energy configurations of the reference state, but unfortunately it is not easy to determine a priori which protein–ligand complex PL1 or PL2 has the higher entropy [42, 43].

A frequent strategy to alleviate this problem is to multistage the calculation of the free energy difference.

$$\Delta G^{\mathrm{EXP}}(PL1 \rightarrow PL2) = \sum_{k=0}^{k=n-1} \Delta G^{\mathrm{EXP}}(\lambda_k \rightarrow \lambda_{k+1}) \quad (13)$$

In Eq. 13 a coupling parameter λ has been introduced to allow the definition of a continuum of systems that describe the two protein–ligand complexes of interest, with PL1 and PL2 as the limiting cases. By convention λ is defined over the interval [0…1] and PL1 is defined at $\lambda_0 = 0.0$ and PL2 at $\lambda_n = 1.0$. Because the intermediate systems can be made arbitrarily close to each other, the problems of convergence and hysteresis of Eq. 12 can be largely eliminated. Eq 13 is of course more expensive to implement because it requires the simulation of several intermediate states. In a typical binding free energy simulation the intermediate states do not describe a chemically meaningful ligand and are therefore not the subject of much interest. It is thus advisable to find ways to minimize the number of intermediate states required, to minimize computational cost.

A technique proposed to reduce this cost is the double wide sampling approach [44]. In this technique, the free energy differences of two perturbed states λ_{k+1} (the forwards state) and λ_{k-1} (the backwards state) are evaluated with a simulation conducted at λ_k (the reference state). The free energy change to proceed from λ_{k-1} to λ_{k+1} is then obtained from a single simulation. As a result, the number of simulations required is halved compared with the so-called direct sampling method shown in Eq. 13.

Other techniques that rely on different averaging schemes have been proposed. For instance in the overlap sampling (OS) method, [45] the free energy difference is computed according to

$$\Delta G^{\mathrm{OS}}(PL1 \rightarrow PL2) = -kT \ln \left[\frac{\langle \exp(-\beta(U_M - U_{PL1})) \rangle_{PL1}}{\langle \exp(-\beta(U_M - U_{PL2})) \rangle_{PL2}} \right] \quad (14)$$

where M is an intermediate system that links the two systems PL1 and PL2. To solve Eq. 14 it is necessary to run two simulations with reference potentials U_{PL1} and U_{PL2}, and perturb in both cases to a potential U_M. In the so called 'simple overlap sampling' approach, U_M is defined as $0.5*(U_{PL1} + U_{PL2})$. This combination increases the likelihood that the low energy configurations of the perturbed potential are also low energy configurations of the reference potential. This simple modification increases the

rate of convergence of the free energy change compared with a direct application of Eq. 12.

Another interesting technique to compute free energy changes that has attracted interest recently is the Bennet acceptance ratio (BAR) method [46, 47]

$$\Delta G^{\mathrm{BAR}}(PL1 \rightarrow PL2) = -kT \ln \left(\frac{\left\langle \{1+\exp(\beta(U_{PL2}-U_{PL1}-C))\}^{-1} \right\rangle_{PL1}}{\left\langle \{1+\exp(\beta(U_{PL2}-U_{PL1}-C))\}^{-1} \right\rangle_{PL2}} \right) + C \quad (15)$$

where the parameter C is a constant that can be iteratively adjusted to obtain the optimal value which minimizes the variance of the ensemble averages of Eq. 15.

Multi-staged versions of OS and BAR, as in Eq. 13 for EXP, are of course possible and usually required to achieve acceptable precision of the computed free energy changes. A multistate Bennet acceptance ratio version that uses statistics collected across multiple thermodynamic states has been recently proposed [48]. In general, the evaluation of the ensemble averages to estimate free energy changes in Eqs. 12, 14 or 15 only requires recalculating the potential energy of the perturbed ligand and is therefore efficient compared to the evaluation of the potential energy of the protein and solvent. This observation is exploited by the double wide sampling technique to reduce the cost of a free energy calculation, but it is possible to achieve further time-savings by evaluating the energies of multiple different perturbed ligands from a single common reference ligand [49, 50]. In principle this approach allows substantial gains in efficiency as only one free energy simulation is required to provide multiple relative binding free energies. However, as the number of perturbed ligands to consider increases it becomes difficult to find a reference ligand such that all the relative binding free energies converge well, and it becomes useful to use a non-physical reference ligand constructed so as to maximize overlap with the set of perturbed ligands [51]. Given the limitations, the method appears better suited to study substituent effects on sets of congeneric or rigid ligands [26, 52].

Methods based on integration

Another established approach computes with Eq. 16 a free energy change by integrating free energy gradients.

$$\Delta G^{\mathrm{TI}}(PL1 \rightarrow PL2) = \int_{\lambda=0}^{\lambda=1} \left(\frac{\partial G}{\partial \lambda} \right) d\lambda \quad (16)$$

where PL1 is defined at $\lambda = 0$ and PL2 at $\lambda = 1$. This relationship is useful because it can be shown that the free energy gradients are related to an ensemble average [39].

$$\int_{\lambda=0}^{\lambda=1} \left(\frac{\partial G}{\partial \lambda} \right) d\lambda = \int_{\lambda=0}^{\lambda=1} \left\langle \frac{\partial U}{\partial \lambda} \right\rangle d\lambda \quad (17)$$

If the perturbation defined by the reaction coordinate λ only involves changes in force field parameters, it is usually straightforward to work out the derivatives of the potential energy function with respect to λ. However, in some free energy calculations it is common to couple with λ geometrical changes in the ligand as well, in which case expressions for the free energy gradients can be difficult to evaluate. This problem can be avoided through the estimation of the free energy gradients by means of a finite difference approach: [53]

$$\frac{\partial G}{\partial \lambda} = \frac{G(\lambda+\Delta\lambda)-G(\lambda)}{\Delta\lambda} = \frac{\Delta G(\lambda \rightarrow \lambda+\Delta\lambda)}{\Delta\lambda}, \text{ provided } \Delta\lambda \text{ is small} \quad (18)$$

The free energy change in Eq. 18 is then estimated by means of a perturbation technique. Given a method to compute free energy gradients at discrete values of λ, the integral in Eq. 16 is then estimated using numerical integration. In principle TI does not suffer from the issues of asymmetry of the forwards and backwards free energy estimates obtained by FEP techniques.

Non-equilibrium methods

Rather than defining a series of intermediate states and using perturbation or integration methods to compute a free energy change, it is possible to compute the work $W_{PL1 \rightarrow PL2}$ necessary to change PL1 into PL2 in a finite amount of time. According to the second law of thermodynamics the work done to change PL1 into PL2 must be greater or equal to their free energy difference.

$$W_{PL1 \rightarrow PL2} \geq \Delta G_{PL1 \rightarrow PL2} \quad (19)$$

The work can be conveniently computed by increasing the coupling parameter λ at each step of the simulation and by recording the change in potential energy.

$$W_{PL1 \rightarrow PL2} = \sum_{i=0}^{i=n-1} U_{i+1}(r) - U_i(r) \quad (20)$$

While it is straightforward to compute $W_{PL1 \rightarrow PL2}$ in a molecular simulation, the approach has found limited use because it is difficult to estimate how much extra work relative to the free energy difference has been performed, unless the change of PL1 into PL2 is accomplished extremely slowly. An interesting related approach is based on the Jarzynsky relationship [54, 55].

$$\Delta G^{\mathrm{JAR}}(PL1 \rightarrow PL2) = -kT \ln \langle \exp(-\beta W_{PL1 \rightarrow PL2}) \rangle_{PL1} \quad (21)$$

This equation shows that the equilibrium free energy difference can be computed as an ensemble average of many, independent work values to convert PL1 into PL2 (and starting from an equilibrated simulation of PL1). Because Eq. 21 does not limit the rate at which PL1 is converted into PL2, it is conceivable to estimate the free energy change from a large number of short computer simulations where PL1 is rapidly converted into PL2. Thus, the approach offers a much greater degree of parallelizability than other equilibrium approaches and is well suited for distributed computing resources. Research conducted in recent years on Eq. 21 suggests, however, that to obtain precise free energy changes of comparable quality to equilibrium free energy methods, a delicate balance must be struck between the speed at which PL1 is converted into PL2, the number of computed work values and the extent of pre-equilibration of PL1 [56–59].

Dynamic λ techniques

Instead of keeping λ fixed it is possible to sample this value over the interval [0…1] during a simulation, through the use of appropriate Monte Carlo moves or by treating the parameter λ as a fictitious particle during a molecular dynamics simulation. The ratio of the probabilities of observing low and high values of λ can then be used to compute a free energy change.

$$\Delta G^{\mathrm{DYN}}(PL1 \rightarrow PL2) = -kT \ln\left(\frac{P(\lambda = 1)}{P(\lambda = 0)}\right) \tag{22}$$

This approach is in fact commonly used to compute free energy changes along an order parameter in molecular simulations (e.g. radius of gyration of a biomolecule, position of an ion in a ion channel…) and can also be used for "non-geometrical" reaction coordinates. A limitation of this type of technique is that the sampling of λ over the interval [0…1] can be slow due to the presence of high energy barriers. It is possible then to split the problem into series of simulations over smaller intervals and to use umbrella potentials to bias the sampling [60]. The full energy profiles can then be reconstructed using weighted histogram analysis techniques [61]. This requires constructing histograms to maintain statistics about the distribution of λ values. Because the approach can be cumbersome, adaptive methods have been proposed that vary the biasing potentials along the reaction coordinate to guarantee that a simulation will visit the entire range of λ values [62, 63]. Interestingly, in the limit where the range of λ values are sampled equally often, the biasing potential is equal to the negative of the potential of mean force and the relative free energy between the two end states can then be easily determined. Unfortunately, it appears that the convergence of the biasing potential may be slow and that precise predictions of free energy changes may require lengthy simulations [64].

Elaborating on these techniques, a λ-dynamics approach has been proposed by Brooks and coworkers for binding free energy calculations [65]. An advantage of this approach is that multiple ligands can be evaluated within a single simulation. The method is conceptually similar to a competitive binding experiment where a mixture of ligands compete for binding to the receptor. The technique has proven useful to efficiently rank-order the binding affinity of series of ligands, [66] and has recently been reviewed [67]. The metadynamics technique of Laio and Parinello [68] is another related approach that has found applications in ligand binding studies [69].

Grand Canonical ensemble methods

Pathway-based free energy calculations, i.e. those that define a coupling parameter λ, are typically computed from molecular simulations in the NVT or NPT ensembles. An alternative route for the computation of free energies of binding is to perform simulations with a varying number of ligand molecules at constant volume V, temperature T, and chemical potential μ. Grand Canonical ensemble free energy methods can be implemented in a number of ways, but in essence, the approach allows exchanges of ligand molecules between a simulation box that contains the protein and a reservoir of ligand molecules in an ideal gas state (i.e. they do not interact with each other). The procedure requires performing a set of simulations at distinct value of the chemical potential μ. For technical reasons it is simpler to vary a dimensionless parameter B that is related to the chemical potential μ of the system and the ligand concentration in ideal gas reservoir. During each simulation, Monte Carlo moves are periodically attempted to insert or delete one ligand molecule from the simulation box [70]. The probability that an insertion or deletion move will be accepted depends on the interaction energy ΔE of the inserted ligand (ΔE values < 0 kcal mol^{-1} are favoured) and the value of the parameter B. For each distinct B value, a simulation is run until the concentration of ligand molecules in the protein box has equilibrated. At high B values it is relatively easy to insert particles and difficult to remove particles, so the simulation box will fill up with numerous copies of the ligand. As the value of B is lowered, insertions become less likely and deletions more likely, so that weakly bound ligand molecules are progressively removed from the simulation box until only the ligand molecules with the most favourable interactions with the protein are left. Thus systematic annealing of the B-parameter from high to low values provide a titration curve from which binding affinities can be derived.

An attractive feature of Grand Canonical computations over pathway simulations is that, assuming convergence has been achieved, the binding mode of the ligand does not

need to be known in advance: the most stable poses of the ligand are simply those retained at low B-values. In early work, water was the ligand of interest and applications of the methodologies were concerned with the detection of tightly bound water molecules around biomolecules, [71–73] or more recently to assist structure based lead optimization [28]. The approach was recently extended to simulate the titration of small organic molecules binding to a protein surface [74, 75]. Convergence can be enhanced by biasing the insertions of ligands into empty cavities to avoid attempting insertions that would lead to steric clashes with the protein [76]. If the protein binding site is already known, it is also possible to limit insertions/deletions to this region which further enhances convergence [75].

Grand Canonical simulations have a number of drawbacks, however. When the ligand is not water, the simulation of the protein–ligand complexes are typically done in the gas phase and the resulting gas phase binding free energies must be corrected post-simulations to account for solvation energies. Typically this is done with implicit solvent models [75]. Further, it is difficult to allow protein flexibility during the simulation, though it is possible to assess the effect of protein flexibility by repeating the simulations using different protein conformers. Similarly, ligands are usually treated as rigid bodies because allowing ligand flexibility hinders convergence. If the ligand can adopt multiple conformers, they can, however, be simulated independently. The methodology is also inefficient for large ligands because the probability of successfully inserting a molecule in the correct position/orientation in a binding site decreases dramatically as the size of the ligand increases. Recent efforts to address this limitation have been focused on the assembly of larger ligands from multiple fragments simulated independently [77].

Overall, Grand Canonical methods appear well suited to identify small molecule binding sites on a protein surface or to select promising scaffolds from a database of small rigid molecules for further optimization. In that respect, the methodology may prove useful to support fragment-based drug design programs.

Direct evaluation of configuration integrals

An innovative approach to the computation of binding affinities has been proposed by Gilson and co-workers [78]. The "mining minima" algorithm estimates configuration integrals as a sum of contributions from the different low energy conformations of a protein–ligand complex identified by a conformational search. Owing to the large number of minima on a potential energy surface the method does not guarantee converged results and is currently only practical using implicit solvent models and with a rigid model of a protein. Nonetheless the technique has proven instrumental in recent theoretical studies to shed novel insights into protein–ligand interactions [79, 80].

Implementation of free energy calculations

Once a methodology to compute a relative or absolute binding free energy has been selected it is necessary to define how the calculation will be implemented in a computer simulation. This usually requires adopting a coupling scheme to perturb L1 into L2 in water and bound to the protein P (or to vanish L1 in the case of an absolute binding free energy calculation), a sampling technique to generate low energy configurations of the ligand and protein–ligand complex, and the selection of an appropriate potential energy function.

Choice of a coupling technique

The standard terminology is to define a coupling parameter λ over the range [0…1] where PL1 (or L1) and PL2 (or L2) correspond to the extremes (otherwise called "end states"). It is usually necessary to define a mathematical relationship that allows the smooth modification of the potential energy function describing PL1 into one describing PL2. There are two major approaches currently. The first approach is called single topology (Fig. 1) [44]. In its most simple form the force field parameters describing PL1 and PL2 are linearly interpolated ($n = 1$). For example for the atomic partial charges:

$$q_{i,\lambda} = \lambda^n q_{PL1} + (1 - \lambda)^n q_{PL2} \tag{23}$$

In addition, geometrical changes (typically bond lengths, dihedrals) can also be coupled to λ.

The second approach does not rely on interpolation of force field parameters but rather interpolates the potential energy of PL1 and PL2,

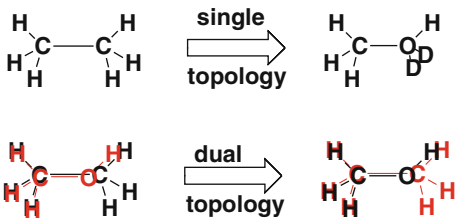

Fig. 1 Computation of the free energy change for the conversion of ethane into methanol. In the single topology implementation two dummy atoms D are added to the molecule of methanol so as to keep the number of atoms constant. Atoms in the two molecular end states are therefore mapped directly onto each other. In the dual topology implementation both molecules are present in the simulation and occupy the same volume. In the initial end-state methanol is not interacting while ethane is fully interacting with the solvent; in the final end-state methanol is fully interacting with the solvent and ethane is not interacting

$$U(\lambda) = U_0 + \lambda^n U(L1) + (1-\lambda)^n U(L2) \quad (24)$$

where U_0 denotes the potential energy arising from the protein and solvent degrees of freedom. U(L1) and U(L2) cover the intermolecular potential energy terms of the ligands. Usually no intermolecular interactions are computed between U(L1) and U(L2). Because this approach requires the simultaneous simulation of L1 and L2 it is often called dual topology (Fig. 1) [41]. In general, it is desirable that the two molecules occupy the same volume in space; this can be achieved for instance by coupling their rigid body motions in a Monte Carlo simulation [17]. Finally it is also possible to make use of hybrid single/dual topology schemes, where part of a molecule is common to both ligands of interest.

When L1 or L2 differ in the number of atoms, or when a dual topology technique is used, the potential energies of a number of ligand atoms have to be turned off/on during the perturbation. Creation and destruction of atoms is accomplished by using dummy atoms (atoms that do not have any force field parameters or whose interaction energy is null). As the Lennard-Jones energy terms of a vanishing ligand atom is progressively weakened, atomic overlaps between non-bonded atoms become allowed and solvent and protein atoms can occupy the space left vacant by the ligand atom. This change in the volume occupied by the ligand gives rise to numerical instabilities in the evaluation of free energy changes that are often referred to as "end-point catastrophes" because they are most apparent near the end-points of the perturbation. Figure 2 illustrates the problem.

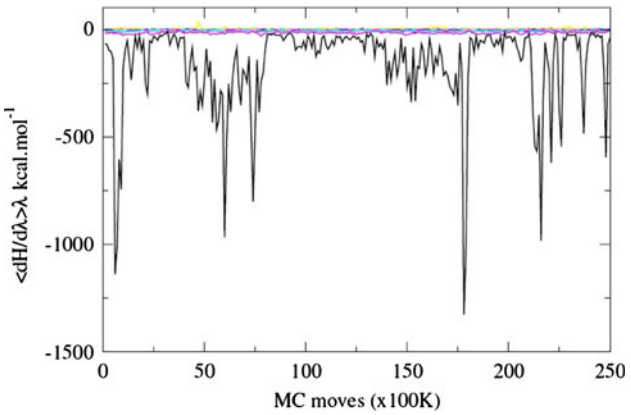

Fig. 2 Fluctuations in the free energy gradients recorded in a dual topology calculation of the free energy change ΔG^{TI} for the conversion of ethane into methanol in a box of TIP4P water.[81] The black line shows the free energy gradients recorded at $\lambda = 1.0$ and the coloured lines the free energy gradients recorded at lower values of λ. The fluctuations in the free energy gradients at $\lambda = 1.0$ are two orders of magnitude greater than at other λ values, making it difficult to obtain a precise estimate of the average gradients. Figure reproduced from ref [82]

To avoid these issues, in single topology calculations it is common to gradually shrink the bonds of atoms that are becoming dummies so that by the time their non-bonded energy is completely turned off they are shielded from molecular overlaps by the Van der Waals radius of the atom to which they are bonded [83]. In general, bonds are not shrunk in dual topology calculations but it is possible to use a non linear interpolation ($n \neq 1$, equation 24) of the ligand potential energy terms to lessen numerical instabilities. Additionally, it can be advantageous to use an irregular spacing of λ values to narrow spacing in the regions of the coupling parameter that are difficult to converge, although it is difficult to estimate the optimum spacing a priori [84]. A more general solution to avoid numerical instabilities is to use a so called soft core potential energy function [85, 86]. For example

$$U_{\text{nonbonded},\lambda} = (1-\lambda)4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}^{12}}{\left(\lambda\delta\sigma_{ij} + r_{ij}^2\right)^6}\right) - \left(\frac{\sigma_{ij}^{6}}{\left(\lambda\delta\sigma_{ij} + r_{ij}^2\right)^3}\right)\right] + \frac{(1-\lambda)^n q_i q_j}{4\pi\varepsilon_0\sqrt{\left(\lambda + r_{ij}^2\right)}} \quad (25)$$

This formula causes the repulsion between two sites to be finite when the ligand L1 or L2 is not fully interacting. Consequently it is possible to bring smoothly to zero the interaction energy of the ligand, and reduce considerably fluctuations in free energy gradients or free energy changes. A problem with Eq. 25 is that the Lennard-Jones interactions can be softened too rapidly which allows the overlap of the softened atom with another atom of opposite charge, causing an infinite Coulombic attraction. To avoid this problem, some workers advocate running two separate simulations, one where all the charges are first turned off, and a second where the Lennard-Jones sites are removed [87]. However, it is possible to also turn off both interactions in a single simulation provided the soft core parameters δ and n in Eq. 25 are appropriately selected to cause the Coulombic interactions to vanish more rapidly than the Lennard-Jones interactions [88]. The latter approach potentially halves the number of simulations required, but the optimal values of δ and n are system dependent and a trial and error phase may be required to adjust these parameters. The former approach is therefore more practical if one wishes to compute free energies for a large number of solutes.

In general it is straightforward to setup a single topology calculation for well localized perturbations, for instance by adding small substituents to a benzyl ring. Single topology has the advantage that the number of degrees of freedom coupled to the perturbation are usually lower than with a dual topology technique. Therefore, single topology calculations are in general more precise and converge faster

than dual topology calculations; they should therefore be preferred for simple calculations [25]. However, it becomes rapidly very complicated to compute free energy changes using single topology if large topological differences exists between ligands L1 and L2 (for instance, a completely different scaffold). The main difficulty for single topology is that the pathway that converts ligand L1 into ligand L2 should in general avoid dramatic distortions of the internal structure of the ligand, because these tend to give rise to abrupt changes in the distribution of the low energy configurations of the intermediate states. This often translates into hysteresis problems or large fluctuations in free energy gradients. In this case dual topology techniques should be preferred because of their versatility. A dual topology approach is also useful to compute the relative free energies of different binding modes of the same ligand.

Additional considerations for absolute binding free energy calculations

Absolute binding free energy simulations are similar to dual topology calculations in that they are usually performed by scaling the interaction energy of a ligand with the surrounding atoms. Because all the atoms of the ligand must become non-interacting, precise free energy changes generally require the use of a soft-core energy function. However, as seen in "Statistical thermodynamics expressions for the computation of binding affinities", absolute binding free energy calculations require additionally the use of spatial constraints on the ligand. The constraints play two roles. First, they allow the unambiguous definition of a volume V that the decoupled but constrained ligand can occupy. These constraints can be defined in several ways, for instance, the centre of the geometry of the ligand could be constrained to stay within a sphere of radius R positioned in the protein binding site. Once V is defined it is possible to compute the free energy change associated with the release of the constraint (Eq. 7) [36]. Further, another benefit of the constraints is that they prevent the ligand from drifting off the binding site once it is no longer interacting with its surroundings, which facilitates convergence of the free energy change associated with the decoupling step (Eq. 8).

Choice of a sampling technique: Monte Carlo and Molecular Dynamics

Most of the techniques to evaluate free energy changes presented in "Computation of free energy changes" require the evaluation of a Boltzmann weighted ensemble averages of different properties. Thus a key aspect of a free energy simulation is the selection of a technique to evaluate these ensemble averages. In principle, one could simply average the value of a target property computed over every distinct configuration the chemical system of interest can adopt. In practice, there are far too many configurations to consider for this approach to be effective with condensed phase chemical systems. The problem is further compounded by the fact that most chemical systems can adopt only a very small number of low energy-high weight configurations amidst a large quantity of high energy-low weight configurations.

To overcome this problem, Markov Chain Monte Carlo methods were developed in 1953 [89]. In this approach, trial configurations are generated by modifying randomly N degrees of freedom of a starting configuration. New configurations are accepted on the basis of a probabilistic test. The functional form of the acceptance test is selected so as to ensure that, in the limit of a large number of samples, the different configurations generated in this fashion are distributed according to a specific distribution (e.g. the Boltzmann distribution). In addition, it is possible to devise intelligent Monte Carlo moves so as to bias the generation of trial configurations towards configurations likely to be of low energy. For example, preferential sampling,[60] configurational bias, [90, 91] and concerted rotations [92, 93]. In general, to propose a new Monte Carlo move it is necessary to verify that the move does lead to a correct limiting distribution. This can be very difficult to prove unless ensuring that the move satisfy the principle of detailed balance or microscopic reversibility, which can be written as in Eq. 26 for the Boltzmann distribution.

$$\frac{P(i \to j)}{P(j \to i)} = \frac{\exp(-\beta U_j)}{\exp(-\beta U_i)} \tag{26}$$

U_i is the potential energy of configuration i and P(i→j) the probability of attempting to move from configuration i to configuration j. Most existing Monte Carlo moves were validated by proving that they respect Eq. 26. However, this requires establishing the ratio of probabilities of attempting a move from configuration i to j and from configuration j to i, which, depending on the nature of the proposed move, can be difficult.

Most Monte Carlo moves are performed in internal coordinates. There is not a unique way to define an internal coordinate representation from Cartesian coordinates, and it can be important to construct an internal coordinate representation that leads to efficient Monte Carlo sampling. For example, improper torsions are often used to allow the concerted rotation of group of atoms with the modification of a single torsion. For large molecules, small random changes of a few internal degrees of freedom at one end of the molecule can translate into large displacements of the other end of the molecule. For this reason, the backbone torsions of a protein are usually not sampled in a Monte Carlo simulation with random moves in dihedral space, but

through concerted rotations or independent translation/rotation of the Cα atom. When constructing an internal coordinate representation for a ligand (often called a *z*-matrix), it is usually best to start defining the ligand degrees of freedom near the centre of geometry of the ligand, to minimize the chance of large displacements induced by a small random change in the value of an internal degree of freedom.

A benefit of the internal coordinate representation is that it is trivial to keep some degrees of freedom frozen during a simulation without evaluating their energy; thus it is possible to simulate the flexible parts of a molecule efficiently while ignoring certain other degrees of freedom for computational efficiency. Because in a Monte Carlo simulation, Monte Carlo moves typically modify a small fraction of the total number of degrees of freedom of a system at a time, it is often necessary to use millions of MC moves to converge the equilibrium properties of the system of interest. Each Monte Carlo move must therefore be very rapid so the overall computational expense is reasonable. To achieve this, it is customary in a Monte Carlo program to evaluate only the terms of the potential energy function that would change after modification of some degrees of freedom, rather than to evaluate the full potential energy function after each move. However, this strategy does not work well if the potential energy function is not separable into pairwise terms. As a consequence Monte Carlo simulations based on quantum mechanical (QM) potentials, certain implicit solvent potentials or with Ewald summations, can be relatively inefficient. It is possible, however, to make use of advanced Monte Carlo moves to eliminate some of this inefficiency [94].

A very common alternative to Monte Carlo simulations is the use of Molecular Dynamics (MD). This technique assumes that in the limit of sufficient time, a single configuration evolved with classical mechanics will visit with the correct frequency all the low energy configurations of a system (condition of ergodicity). Molecular dynamics has the advantage of using information about forces to guide the search for new low energy configurations, and it also handles more easily large scale concerted motions of multiple atoms that can be very slow to sample with simple Monte Carlo moves. Additionally, kinetic information can be more easily extracted from a MD trajectory as it follows a time evolution.

The approach is not however without difficulties; most MD programs perform simulations in Cartesian coordinates which sample more slowly the torsional degrees of freedom of molecules, [95] and the freezing of degrees of freedom requires in MD the use of specialized algorithms. In general, MD simulations are more computationally intensive than MC simulations, although MD programs have often benefited from further software optimization and parallelization. Additionally, many MD integration schemes have been proven to not be ergodic, particularly for long trajectories, and therefore the quality of the ensemble averages is affected [96]. Finally, it is possible to consider molecular dynamics as a Monte Carlo move in the context of hybrid Monte Carlo/molecular dynamics algorithms [97, 98].

In general, none of the current techniques are generally applicable to protein–ligand complexes because they do not yet navigate the potential energy landscape efficiently, and therefore a large range of protein conformational changes are not captured on practical time scales by current sampling techniques. Research in novel sampling algorithms to improve binding affinity predictions is an active topic of molecular simulations.

Among the recent developments, replica exchange methodologies have attracted considerable interest. In this approach, a system of interest is simulated simultaneously at high and low temperatures. Occasionally, configurations generated by the sampling algorithm are exchanged, subject to a particular acceptance test that ensures each simulation converge to its equilibrium distribution. The benefit of this approach is that the high temperature simulations can explore a wider range of low energy states as they will overcome more easily energy barriers separating low energy states. The replica exchange test allows the propagation of newly discovered states to low energy temperature simulations [99]. A disadvantage of this approach is that it increases significantly the computational expense; it is not uncommon to simulate a dozen replicas for one system of interest, and statistics about the ensembles of states generated at a high temperature are often not the subject of much scrutiny, although temperature dependent properties such as entropy can be derived in principle in this fashion [100]. It has been proposed to use replica exchange moves across different values of the coupling parameter λ used in free energy simulations (rather than different temperatures) [101]. Replica Exchange Thermodynamic Integration has been shown to increase the precision of computed hydration and binding free energies [101, 102]. Since a free energy simulation already typically requires performing a dozen independent simulations at different values of this coupling parameter, the approach is no more computationally intensive than a regular free energy simulation, although it requires all the replicas to be run in parallel.

Another strategy to enhance sampling is to use approximate, computationally inexpensive potential energy functions to rapidly generate new low energy configurations. Without any further considerations, this would be expected to generate ensembles of low accuracy. However, if the newly generated configurations are periodically evaluated using an accurate potential energy function, it is possible to correct errors and generate ensembles of high quality with

fewer expensive potential energy evaluations [94, 103]. This strategy has been shown to accelerate considerably conventional force field simulations,[94, 103] implicit solvent simulations, [104] or QM/MM simulations [105]. It is comparable to multiple time step molecular dynamics techniques [106].

Another recent strategy to enhance sampling of protein–ligand configurations has been to parallelize the generation of configurations over different simulations. In this technique each simulation is assigned the exploration of a fraction of the configuration space of a protein–ligand complex and the final thermodynamic properties are estimated by combining the results from the individual simulations [107]. While in principle this approach can allow a large decrease of the time to answer with sufficient distributed computing resources, a problem with this technique is that it is difficult to assign unambiguously simulations to different regions of configuration space, and therefore it is not straightforward to merge the results from each simulation.

Choice of potential energy function

The generation of new configurations by Monte Carlo or molecular dynamics simulations requires repeated evaluations of potential energies. For computational efficiency, the vast majority of binding free energy calculations is performed to date using molecular mechanics (MM) force fields.

Proteins

In general the selection of a MM force field to describe a protein does not cause serious issues [108]. There are a number of widely used force fields that have been employed in numerous free energy calculations, for instance: AMBER, [109] CHARMM, [110] OPLS, [111] and GROMOS [112]. However, parameters can still be an issue for metals, co-factors or less common biomolecules that it could nonetheless be necessary to consider, because they are in the vicinity of the binding site. In addition, the force fields parameters, in particular torsional potentials, evolve over time as systematic errors are identified; [113] thus it is important to check for newer parameter sets before starting a new binding free energy study.

Ligands

Because of the large number of functional groups encountered in small drug-like molecules, the parameterization of ligands with a MM force field is usually more difficult than proteins. When parameters are not available, they are often derived by a time consuming fitting procedure to reproduce QM calculations. Typically this represents too much work for most scientists interested in predicting the binding affinities of a series of ligands to a protein, so alternative options have been sought. One solution is the development of general purpose force fields for drug-like molecules. Efforts along those directions are exemplified by the GAFF force field [114]. Another force field is the OPLSAA-CM1A force field which relies on rapid QM calculations to obtain partial charges [115]. The use of atomic partial charges derived from a QM calculation can be advantageous because it allows, for example, taking into account stereoelectronic effects that can affect, for instance, the strength of a hydrogen bond in a protein–ligand complex without the need to develop a very large parameter set [116]. However, the accuracy of these general purpose force fields cannot be expected to be as high as those that have been finely tuned to work for a smaller class of molecules such as proteins. For instance, the average error in the prediction of hydration free energies for 25 diverse organic molecules is 0.7 kcal mol^{-1} with the OPLS-AA force field but 1.0 kcal mol^{-1} with the OPLS-CM1A force field [115].

The setup of many classes of ligands can be complicated by additional considerations such as tautomerization and protonation states. It can be difficult to know which particular state is the most relevant for a given ligand when bound to a protein and in solution. It could be in fact several, but current MM force fields do not easily allow the simulation of multiple tautomeric/protonation states. These issues can also be important for the protein, but tend to be limited to a smaller range of residues, typically acidic and basic amino acid side chains, and for which some information may sometime be inferred from biophysical experiments.

Solvent

The established approach in molecular simulations of biomolecules is to use atomic models of the solvent, usually water. The TIP4P [81] model is often used in Monte Carlo simulations while the less accurate model TIP3P [81] is more often used in molecular dynamics simulations because the presence of off-centre point charges in TIP4P can cause instabilities in the integration of molecular dynamics trajectories. A problem with explicit solvent simulations is that solvation of a protein usually requires the consideration of several thousand water molecules, and this adds a significant computational cost to the simulations. Additionally, it can be difficult to equilibrate the position of water molecules in binding sites that are partially occluded from bulk solvent and it is easy to setup a free energy simulation with an incorrect distribution of water molecules in the vicinity of the ligand—some water molecules may be kinetically trapped in unfavourable

regions, or other regions are left dry and no water molecule is able to diffuse there on the timescale of the simulation. It has been shown that simulation of an incorrect water distribution in the vicinity of a ligand can dramatically affect the computed free energies of binding [117]. Depending on the nature of the binding site it can then be advised to equilibrate a protein–ligand complex using for example Grand Canonical Monte Carlo moves that allow exchange of water molecules with bulk, [72, 118] or other methods based on the computation of absolute binding free energies of water molecules [119]. Another issue is that it is somewhat simplistic to model the solvent as pure water, when in fact the aqueous environment of a protein in a binding assay is likely to be more complex, with buffer and pH influencing the measured affinity.

Alternatively, it is possible to model the influence of the solvent on the protein–ligand interactions through the use of classical electrostatic theories [120]. Salt effects can also be modelled implicitly. Recent extensive studies on the accuracy of computed hydration free energies suggest that explicit solvent models provide more accurate predictions, but that implicit solvent models fare well considering that they are computationally much cheaper and therefore a good option for high-throughput free energy calculations [121, 122]. There is indeed evidence that implicit solvent models parameterized for molecular simulation methods can perform reasonably well to be useful in binding free energy calculations, [16] although it is important to test the stability of the implicit solvent energies in a molecular simulation, as parameter sets have not always been derived with these considerations in mind [123].

Inevitably, an implicit solvent approach will treat protein–ligand interactions more crudely, and there are many documented cases of important water mediated protein–ligand interactions [124]. If this is thought to be the case for a protein–ligand system of interest it might be more fruitful to model explicitly a small number of water molecules bound in the vicinity of the ligand during the simulations [125].

Treatment of long range interactions

The full evaluation of all the terms of a potential energy function is expensive, and it is customary to use cutoffs to discard distant intermolecular interactions [126]. It is important to be aware that this can introduce significant errors into the computed free energy changes. The neglect of distant electrostatic interactions is particularly problematic for simulations of polar molecules because electrostatic interactions are long ranged and significant beyond most typical cutoffs. A standard approach in molecular simulations to lessen the approximation introduced by spherical cutoffs is to add a reaction field correction term that accounts for the electrostatic energy neglected beyond the cutoff radius [126, 127]. The approach implies a medium of homogeneous dielectric beyond the cutoff radius and is reasonable to correct the electrostatic interactions for simulations of a ligand in solution, but is only approximate for simulations of the ligand–protein complex. To reduce errors introduced by unbalanced long-range corrections, it is common to use similar cutoffs and system sizes for simulations of the ligand in solution and the protein–ligand complex; hence if the protein–ligand complex is solvated by a non periodic sphere of 25 Å centred on the ligand, so too is the simulation of the isolated ligand.

Alternatively, for simulations in periodic boundaries, the Ewald summation has proven a popular alternative to spherical cutoffs and reaction field corrections [128]. The particle mesh Ewald technique is an algorithmic improvement that allows more rapid evaluations of the electrostatic energies over simple Ewald summation, and is often used in molecular dynamics simulations [128]. It is important, however, to be aware that the Ewald summation implies a periodicity of the electrostatic interactions which is an approximation for simulations of biomolecules in solution [129].

While the previous techniques are useful to treat dipolar molecules in solution, an accurate treatment of charged molecules is still an ongoing topic of research in molecular simulations. The computation of thermodynamic properties (e.g. hydration free energies) for ions is very sensitive to the choice of the boundary conditions and the approximations used to handle the long range component of electrostatic interactions. A thorough discussion of these issues is presented by Kastenholz and Hünenberger [130, 131].

As proteins almost always consist of charged residues, various strategies are employed to lessen the errors introduced by an incomplete treatment of electrostatic interactions. A realistic approach to long range electrostatic interaction may be the explicit consideration of counter-ions to ensure overall electro neutrality of the protein–ligand complex. This also provides a way to account for the effect of the ionic strength as many binding assays are done in buffered solutions. This approach is, however, fraught with difficulty because the adequate placement of counter ions is typically uncertain. Further, if the ion concentration is low, over the course of the simulation, a small number of ions may slowly drift from their initial position and yet should be able in principle to sample a large fraction of the simulation box, hindering greatly the convergence of thermodynamic properties [132]. An approach that has been favoured, particularly for simulations without periodic boundaries, is to neutralize selected charged protein residues by modelling them in their non ionic form, this effectively eliminates the long range electrostatic contributions due to monopole interactions, but may also lead to inaccurate description of

the electrostatic interactions in the vicinity of this residue; such residues are therefore typically chosen such that they are the most distant from the ligand [133].

This approximate approach cannot be justified for protein residues in the vicinity of a ligand or the ligand itself bearing a non-zero net charge. Progress is being made in developing correction terms for the computation of the absolute hydration free energy of spherical ions, [130, 134] but a general formalism to handle correctly, charged molecules of arbitrary shape is currently lacking.

Finally, long-range correction terms can also be used to account for missing dispersive interactions beyond the cutoff radius [126]. Because dispersive interactions decay more rapidly than electrostatic interactions, they can be accounted for accurately for simulations of ligands in solution, but only approximately in the case of protein–ligand complexes [135]. The proper accounting of missing dispersive interactions is more important for absolute binding free energy calculations as the correction term tends to cancel out in calculations of relative binding free energy calculations between similar ligands.

Binding free energy studies on charged ligands should be conducted with caution. The absolute binding free energies will likely show a large dependence on the simulation protocol. Relative binding affinities may show less dependence due to a systematic cancellation of errors, but only if the net charge is not varied.

Quality and reliability of results

To understand the strengths and weaknesses of the numerous techniques available to compute free energy changes, it is crucial to understand the factors that affect the quality of the results obtained from a free energy simulation.

Accuracy

Regardless of the technique used to estimate a free energy change, it is necessary to use some form of potential energy function to compute the required ensemble averages. Because all current techniques only represent approximately the nature of intermolecular interactions for the sake of efficiency, it is quite possible that a computed free energy change is exact given the underlying potential, but not comparable to an experimental measurement.

So how well can we expect current biomolecular force fields to fare in protein–ligand binding free energy calculations? It seems reasonable to expect that free energy calculations cannot predict more accurately binding free energies than they can predict solvation free energies. There have been several recent studies where free energy calculations were used to compute absolute and relative hydration free energies of small organic molecules. Because the extent of sampling was high, the hydration free energies were computed to a high precision and were reproducible, suggesting that they are truly converged for these simple systems. In general the best performing protocols show a mean unsigned error of around 1 kcal mol^{-1}, but there can be much larger deviations among certain families of compounds (nitro compounds being a typical example). Larger deviations can be expected for more complex polyfunctional drug-like molecules. This should serve as lower bound estimate of the ideal accuracy of current force fields as applied to ligand binding studies [87, 88, 121, 136].

Currently a number of research groups are working on the development and application of polarizable force fields to free energy calculations [137–140]. Because polarization effects are thought to be important in a number of protein ligand interactions (for instance cation-π interactions [141]) this should in principle lead to more accurate predictions of binding free energies. Some applications of such polarizable force fields to studies of ligand binding have started to appear in the literature [142, 143]. However, as of now, these force fields have not been tested extensively and have not found widespread use. Another trend is towards the use of mixed quantum mechanical/molecular mechanical (QM/MM) representations to compute binding affinities [105]. Although given the computational expense of QM calculations, applications to date tend to crudely consider configurational sampling [144]. Hardware and algorithmic advances in quantum chemistry are closing the gap between the current and necessary computing resources required to compute quantum mechanical energies for a large number of protein–ligand configurations [145].

Precision

It is well appreciated that a significant problem with free energy simulations is their inability to converge the computed free energy changes. This is an important issue because it seems difficult to concentrate efforts on systematically improving the accuracy of the potential energy function if highly reproducible free energy changes for a given potential energy function cannot be obtained first. It is important to recognize that there are two main sources of imprecision in a free energy calculation.

The first depends on the nature of the free energy estimator. An ill conceived free energy perturbation study can be highly imprecise if, for example, there is little or no overlap between the low energy configurations of the reference and perturbed state. Methodologies such as double wide sampling, [44] overlap sampling, [45] or Benett

acceptance ratio, [46, 47] acknowledge this problem in different ways and provide means to reduce the magnitude of this error. In general, regardless of the perturbation methodology, the precision of the free energy changes can be improved more rapidly by reducing the spacing between windows. So if there is poor overlap between a reference and perturbed state and if it is feasible to double the computational expense for the measurement of a free energy change, it seems better to double the number of windows rather than double the length of a simulation at each window. While this is highly dependent on the nature of the perturbation attempted, in general it seems that most protein–ligand binding free energy studies use between 10-30 windows or integration points to compute a free energy change, with over 100 windows in some cases for absolute binding free energy calculations on large ligands [146].

With thermodynamic integration there is no longer any problem of overlap between reference and perturbed states, so at first it may seem that this technique offer a better alternative. However, the use of a discrete number of integration points to estimate the potential of mean force can be a source of large inaccuracy. For instance, if a free energy gradient is rapidly changing over a small interval, numerical integration of this potential of mean force may miss out important regions of curvature. This was observed for instance in the comparison of the free energy change computed by non equilibrium perturbation schemes and replica exchange thermodynamic integration schemes [56]. The TI calculations missed a region of large curvature of the potential of mean force and yielded free energy changes very different from the fast-growth calculations. The relative hydration and binding free energies were, however, very similar because this large curvature of the potential of mean force was due to intramolecular non-bonded interactions present in all individual free energy changes, and therefore the effect cancelled out. However, this is still indicative of the kind of problems that TI can suffer. A practical advantage of TI over FEP methods is that it is straightforward to run additional simulations at different values of the coupling parameter to increase the precision of the pmf, while this is not done so simply with FEP methods where the spacing of the individual windows must usually be defined beforehand.

The second and arguably most important source of error is due to the finite size of the sampled ensembles. Because any computer simulation must necessarily run over a finite amount of time, it is not guaranteed that all low energy configurations of the reference state have been visited with a correct frequency. Consequently, even if a free energy estimator scheme has been selected such that the measurement of the free energy change is highly precise, it may still differ from the converged result because many important low energy configurations have been ignored. For instance, despite the theoretical advantages of overlap sampling over double wide sampling, a recent study that systematically compared free energy changes for pairs of substituted benzenes in solution, found only slight improvements, [147] likely because the finite size sample error dominates the source of statistical errors in these calculations.

Block averaging techniques are also commonly used to estimate the uncertainty on a computed free energy change. Essentially, an entire trajectory is divided into blocks and free energy changes are computed with the data available in each block and the standard deviation of the mean of the free energy changes provides an estimate of the stability of the computed free energy changes [126]. A problem with this technique is that samples from two neighbouring blocks are often highly correlated and the method tends therefore to underestimate the statistical error. A way to avoid this is to use correlation functions to compute statistical inefficiencies that provide an estimate of the minimum length a block should be for two consecutive blocks to be uncorrelated [126].

A more expensive but arguably better way to obtain good estimates of the statistical error is to repeat each free energy simulation independently. The free energy change and associated error can then be estimated from the mean and standard deviation of the independent realizations. Another interesting approach that can prove useful if relative free energies are computed between several ligands, is to perform additional simulations between pairs of ligands [148]. Because free energy is a state function, the sum of free energy changes along a closed thermodynamic cycle should be zero. Deviations from this value are usually indicative of a lack of sampling, although they can also highlight subtle systematic differences in the setup of the different protein–ligand complexes. If several thermodynamic cycles are closed, it is possible to identify in this fashion a problematic measurement and improve the quality of the results by extending the length of the simulation and/or reducing the spacing between windows/integration points.

In general it is fair to say that most current free energy simulation protocols probably under sample the low energy configurations of a protein. For example, it has been shown that the binding affinities of a series of ligands of a mutant of the protein T4 lyzozyme depend markedly (over several kcal mol^{-1}) on the position of a nearby Valine side chain that can adopt two distinct low energy configurations, but is unable to interconvert between these two configurations using standard sampling protocols [23]. Similarly, we have observed that binding free energy estimates differed between two different structures of the same protein solved with a different ligand. The differences were greater than the differences observed between different independent

simulations repeated on the same protein structure, thus indicating systematic imprecision due to a lack of sampling [19]. It is likely that future methodological developments in free energy calculations will aim to reduce the dependency of the computed binding free energies on the initial protein conformation.

Comparison with experimental data

A word of caution is necessary when attempting to compare predicted binding free energies to measured binding affinities. The simulations usually imply idealized conditions such as pure water, which rarely reflect the conditions in which binding affinities are measured. Sometimes the protein used in the assay differs somewhat from the protein structure used for the predictions (e.g. it could be a longer construct). A binding assay is often setup to measure binding affinities within a limited range; the binding affinity of weak or strong binders can be under or overestimated. Also, it is relatively uncommon to measure accurately the absolute binding free energy of a ligand as the biophysical techniques available to do so are often of a too low throughput and/or require too much sample. More often the binding affinities are established through some form of competition assay with a reference ligand whose signal can be measured easily. The resulting dose response curves can be converted into IC50 which can be related to K_d through the Cheng-Prusoff relationship although this is not always valid [149, 150].

There are many examples of variability between different in vitro assays reported in the literature. For example, the IC50 of potent inhibitors of EGFR kinase varies by over one order of magnitude depending on the assay used [151]. In general it seems more productive to only consider binding affinities relative to a reference ligand rather than attempt to put them on an absolute scale, and preferably to only use data coming from the same assay/study. Finally it must be mentioned, obviously, that in vivo activity measurements, for instance in cell based assays, may not correlate strongly with a predicted binding affinity since numerous other factors (cell permeability, off target inhibition, stability) will contribute to the ligand efficacy.

Realistic expectations: which approach will work best?

Given the plethora of options available nowadays to compute a binding free energy, it would not seem surprising that a novice user may be confused as to which technique is likely to be more fruitful for the study he/she wishes to conduct. In this section we provide some guidelines to conduct a binding free energy study. Inevitably these observations reflect our judgments drawn from past experiences and will not exhaustively address possible issues. Additional insights may be gained by consulting recent reviews from other co-workers [12, 34, 152, 153].

- *Is the ligand dataset challenging?* One of the first and most important questions to ask is whether free energy calculations will be worth the trouble. Important insights can be obtained by considering the dataset of ligands that are of interest. An obvious question is the size of the dataset. Large datasets of 100 + compounds are likely to represent a significant endeavour for most organizations. The computational cost may appear significant, but the trends in high performance computing are in favor of large scale binding affinity predictions. Increasingly, the major issue is the correct setup of the protein–ligand simulations, as large datasets are more likely to encounter methodological difficulties discussed earlier in this article (large floppy compounds, different net charges, tautomers etc.…). The structural diversity of the ligand dataset can be indicative of the difficulty of the study. In general binding affinities computed for series of structurally similar ligands benefit from cancellation of systematic errors.
- *Is the protein binding site challenging?* There could be a co-factor in the binding site for which force field parameters must be derived. There could be metal ions whose intermolecular interactions may be difficult to treat with a classical force field. In polar binding sites, water mediated protein–ligand interactions may contribute significantly to the binding affinity of a ligand but experimental data about the location of all relevant water molecules may be incomplete. In addition, the more flexible and solvent exposed a protein binding site is, the more likely it is that complex conformational changes may have to be sampled to obtain reliable binding free energies. In other cases, the binding site may collapse completely in the absence of a ligand, or small modifications to a ligand could cause conformational changes in the protein. If available, visualization of multiple apo/holo crystal structures can give advance warnings that long simulations will be required.
- *Are the ligand binding modes known?* It is also critical to assess whether the binding modes of the ligands are likely to be correct, as most free energy calculations methods must be initiated from a correct binding mode. This can be assessed if there is experimental data for a compound structurally related to the ligands under study, but may prove difficult in other cases. Even seemingly simple structural modifications may be challenging. For instance, an ethoxy group could be added to a scaffold in multiple different orientations that are all seemingly plausible. While the sampling of

the ligand and protein degrees of freedom should correct an inaccurate binding mode over the length of a simulation, this could easily require a simulation far longer than what can be run in practice. So in general it cannot be expected to obtain sensible results if the binding mode of a ligand is incorrect.

- *Is there experimental data to validate the setup?* While free energy calculations are, in theory, free of any empirical parameter adjustment for a specific protein–ligand systems, given the dependence of the computed free energies on many parameters (initial protein structure, long range electrostatic treatments, force field, protonation states…) it is very useful if some experimental data for a few systems is available beforehand to validate a protocol before embarking on time consuming calculations for a large dataset.
- *Are absolute or relative binding free energies needed?* Depending on the composition of the dataset, different approaches to the computation of binding affinities are available. The most general way to predict binding affinities is to compute the absolute binding free energy of each ligand. However, this will usually require a significant amount of computing time, especially if the ligand is highly flexible in solution, in which case there may be a large amount of conformations to explore during the simulation. Additionally, absolute binding free energy methods are not guaranteed to give a correct answer because removing a ligand from a binding site often causes conformational changes in the protein that are not well sampled by current techniques. Furthermore, the binding site of an apo protein is rarely empty and will be usually occupied by solvent. The precise amount and location of water molecules in the vicinity of the ligand can remarkably affects the computed free energies of binding, and accurate results should be expected only if the correct solvent structure in the binding site is obtained when the ligand is fully decoupled from the binding site. Rather than wait for bulk water molecules to diffuse inside a binding site, one may want to establish the distribution of water molecules using algorithms like JAWS or GRID for instance, [119, 154] and grow them in while or after the ligand is decoupled [155]. Alternatively, one can use a Grand Canonical Monte Carlo scheme to dynamically insert and remove water molecules from the binding site [72, 118]. Given these considerations, in the absence of extensive computing resources, current absolute binding free energy calculation schemes are expected to be better suited to score small and fairly rigid small molecules binding to a fairly rigid binding site. Thus it is conceivable that absolute binding free energy calculations will find useful applications in fragment-based drug discovery programs. In the cases where the ligands exhibit substantial structural differences but still occupy the same region of the binding site, a dual topology relative binding free energy technique may prove more practical than an absolute binding free energy calculation as issues such as binding site conformational changes or hydration changes are expected to be lessened if a ligand is always present in the binding site during the simulation [17, 19]. On the other hand, if one wishes to rank-order a series of congeneric ligands according to their potency, it will require far less computing time to do this using a single topology approach than by a dual topology relative or absolute binding free energy approach [17]. However, successful computation of relative free energies of binding using a single topology technique requires more knowledge about the protein–ligand system. In particular it is important that all the analogs in the congeneric series adopt the same binding modes, as large conformational re-arrangements of the ligands are unlikely. This situation may occur for instance when ortho/meta substituents on a substituted phenyl ring are being considered. There can be ambiguity in the selection of the appropriate orientation of the substituents in the binding site and a 180 degrees flip of the phenyl ring may be unlikely to happen over the simulation duration. In this situation is it usually instructive to compute relative binding free energies for substituents added at both equivalent positions to identify those that yield the most favourable free energy changes.

If it appears that the dataset of interest is amenable to free energy simulations, that the computed binding affinities are reproducible and well behaved as judged by the various analysis techniques outlined previously, it becomes possible to interpret the predicted quantities to guide ligand design. Our observation of current recent free energy simulation studies is that the spread of the computed binding affinities tends to be larger than those actually measured in a binding assay, see for example refs [16, 19, 27, 33, 35, 153]. The origin of this discrepancy is not yet well understood, although one could blame the accuracy of the force field or the completeness of sampling; we suspect the latter to dominate as hydration free energies for small organic molecules are now generally well converged and typically show reasonable agreement with experiments.

In general it should not be expected that absolute binding free energies can be predicted with higher accuracy than hydration free energies. Extensive benchmarks of the latter suggest that the mean error reaches ca. 1 kcal mol^{-1} in favourable cases with current force fields, so this figure should serve as a lower bound of the expected mean error for absolute binding free energies. In addition the

imprecision of the computed free energies of binding must also be considered. The standard deviation of the compute binding free energies may commonly vary from ca. 0.1 to 1.0 kcal mol^{-1}, depending on the nature of the ligands and the simulation protocol. Furthermore, large systematic biases due to the selection of a particular protein structure can also be observed in some cases.

Given these considerations, improvements in measured binding affinities are more likely to be observed for predictions of a significant gain in binding affinity (lowering of ΔG_{bind} by > 1 kcal mol^{-1}). Even if an analogue is predicted to be less potent than a reference ligand, these predictions should be critically considered. For instance it is possible that the said analogue could bind more strongly in a different orientation that has not been sampled during the simulation.

Outlook

The accurate prediction of the binding affinity of a small molecule to a protein remains a major challenge for computational chemistry. For a long time of a mainly academic research interest, a large increase in computing power, coupled with the development of improved computer simulation techniques, is greatly facilitating the adoption of statistical thermodynamics approaches to predict binding affinities in a more general context. At this stage it would be unrealistic to expect quantitative predictions of binding affinities: free energy calculations still have to overcome numerous technical challenges before they become generally applicable.

Perhaps a major limitation of this technology is that the setup of a binding free energy calculation currently remains a complex task using existing software. Successful binding free energy studies have been primarily conducted by expert users to date, but widespread adoption will probably require a new generation of easier to use software, built around robust simulation protocols, and well integrated with the many other pieces of software a molecular modeller readily uses. Examples of initiatives along these directions include the development of the software MCPRO+, [156] and the addition of free energy calculation features to the molecular dynamics software Desmond [157]. Another major factor that prevents wider usage of free energy calculations in drug design is the perceived lack of validation of the technology. Because of significant developments in high performance computing and advances in simulation protocols, it is becoming feasible to screen large datasets of ligands. There are now initiatives, such as the SAMPL challenge, organized by Open-Eye, to gather high quality binding and structural datasets for drug-like small molecules [158]. In the spirit of the protein structure prediction contest CASP, these datasets will facilitate the organization of blind prediction challenges and provide a critical assessment of the predictive power of free energy calculations.

In spite of these caveats, with an adequate knowledge of the strength and weaknesses of the current techniques, it is today possible to recognize when free energy simulations can enable semi-quantitative predictions to guide structure-based drug design efforts. Indeed, imperfect as they currently are, free energy calculations have the merit of focussing the attention of molecular modellers on factors contributing to binding affinities that are not readily apparent from visualization of static pictures of protein structures, for instance the contribution of protein and ligand flexibility or changes in hydration. It is hoped that the material presented here will be useful to molecular modellers interested in applying effectively free energy simulations in a drug design context.

Acknowledgments This research was partially supported (J.M.) by a Marie Curie International Outgoing Fellowship within the 7th European Community Framework Programme (FP7-PEOPLE-2008-4-1-IOF, 234796-PPIdesign).

References

1. Jhoti H (2007) Structure-based drug discovery. Springer, Dordrecht
2. Jorgensen WL (2004) Science 303:1813–1818
3. Klebe G (2006) Drug Discov Today 11:580–594
4. McInnes C (2007) Curr Opin Chem Biol 11:494–502
5. Taylor RD, Jewsbury PJ, Essex JW (2002) J Comput Aided Mol Des 16:151–166
6. Warren GL, Andrews CW, Capelli AM, Clarke B, LaLonde J, Lambert MH, Lindvall M, Nevins N, Semus SF, Senger S, Tedesco G, Wall ID, Woolven JM, Peishoff CE, Head MS (2006) J Med Chem 49:5912–5931
7. Cozzini P, Kellogg GE, Spyrakis F, Abraham DJ, Costantino G, Emerson A, Fanelli F, Gohlke H, Kuhn LA, Morris GM, Orozco M, Pertinhez TA, Rizzi M, Sotriffer CA (2008) J Med Chem 51:6237–6255
8. Kokh DB, Wenzel WG (2008) J Med Chem 51:5919–5931
9. Sherman W, Day T, Jacobson MP, Friesner RA, Farid R (2006) J Med Chem 49:534–553
10. Verdonk ML, Mortenson PN, Hall RJ, Hartshorn MJ, Murray CW (2008) J Chem Inf Model 48:2214–2225
11. Gilson MK, Zhou HX (2007) Annu Rev Biophys Biomol Struct 36:21–42
12. Mobley DL, Dill KA (2009) Structure 17:489–498
13. Guimaraes CRW, Boger DL, Jorgensen WL (2005) J Am Chem Soc 127:17377–17384
14. Mobley DL, Graves AP, Chodera JD, McReynolds AC, Shoichet BK, Dill KA (2007) J Mol Biol 371:1118–1134
15. Fujitani H, Tanida Y, Ito M, Jayachandran G, Snow CD, Shirts MR, Sorin EJ, Pande VS (2005) J Chem Phys 123:084108
16. Michel J, Verdonk ML, Essex JW (2006) J Med Chem 49:7427–7439
17. Michel J, Verdonk ML, Essex JW (2007) J Chem. Theory Comput 3:1645–1655

18. Deng YQ, Roux B (2006) J Chem. Theory Comput. 2:1255–1273
19. Michel J, Essex JW (2008) J Med Chem 51:6654–6664
20. Wang JY, Deng YQ, Roux B (2006) Biophys J 91:2798–2814
21. Aqvist J, Luzhkov VB, Brandsdal BO (2002) Acc Chem Res 35:358–365
22. Massova I, Kollman PA (2000) Perspectives in Drug Discovery and Design 18:113–135
23. Mobley DL, Chodera JD, Dill KA (2007) J Chem. Theory Comput. 3:1231–1235
24. Boyce SE, Mobley DL, Rocklin GJ, Graves AP, Dill KA, Shoichet BK (2009) J Mol Biol 394:747–763
25. Tanida Y, Ito MS, Fujitani H (2007) Chem Phys 337:135–143
26. Oostenbrink C, van Gunsteren WF (2004) Proteins Struct Funct Genet 54:237–246
27. Steinbrecher T, Hrenn A, Dormann KL, Merfort I, Labahn A (2008) Bioorg Med. Chem. 16:2385–2390
28. Pan CF, Mezei M, Mujtaba S, Muller M, Zeng L, Li JM, Wang ZY, Zhou MM (2007) J Med Chem 50:2285–2288
29. Jorgensen WL, Ruiz-Caro J, Tirado-Rives J, Basavapathruni A, Anderson KS, Hamilton AD (2006) Bioorg Med. Chem. Lett. 16:663–667
30. Ruiz-Caro J, Basavapathruni A, Kim JT, Bailey CM, Wang LG, Anderson KS, Hamilton AD, Jorgensen WL (2006) Bioorg Med. Chem. Lett. 16:668–671
31. Thakur VV, Kim JT, Hamilton AD, Bailey CM, Domaoal RA, Wang LG, Anderson KS, Jorgensen WL (2006) Bioorg Med. Chem. Lett. 16:5664–5667
32. Kim JT, Hamilton AD, Bailey CM, Domoal RA, Wang LG, Anderson KS, Jorgensen WL (2006) J Am Chem Soc 128:15372–15373
33. Zeevaart JG, Wang LG, Thakur VV, Leung CS, Tirado-Rives J, Bailey CM, Domaoal RA, Anderson KS, Jorgensen WL (2008) J Am Chem Soc 130:9492–9499
34. Jorgensen WL (2009) Acc Chem Res 42:724–733
35. Michel J, Harker EA, Tirado-Rives J, Jorgensen WL, Schepartz A (2009) J Am Chem Soc 131:6356–6357
36. Gilson MK, Given JA, Bush BL, McCammon JA (1997) Biophys J 72:1047–1069
37. Boresch S, Tettinger F, Leitgeb M, Karplus M (2003) J Phys. Chem. B 107:9535–9551
38. Ben Naim A (2006) Molecular Theory of Solutions. Oxford University press, Oxford
39. Leach A (1996) Molecular Modelling: Principles and Applications. Longman, Harlow, UK
40. Zwanzig RW (1954) J Chem Phys 22:1420–1426
41. Kollman P (1993) Chem Rev 93:2395–2417
42. Wu D, Kofke DA (2005) J Chem Phys 123:084109
43. Wu D, Kofke DA (2005) J Chem Phys 123:054103
44. Jorgensen WL, Ravimohan C (1985) J Chem Phys 83:3050–3054
45. Lu ND, Kofke DA, Woolf TB (2004) J Comput Chem 25:28–39
46. Bennett CH (1976) J Comput Phys 22:245–268
47. Shirts MR, Bair E, Hooker G, Pande VS (2003) Phys Rev Lett 91:140601
48. Shirts MR, Chodera JD (2008) J Chem Phys 129:124105
49. Liu HY, Mark AE, vanGunsteren WF (1996) J Phys Chem 100:9485–9494
50. Oostenbrink BC, Pitera JW, van Lipzig MMH, Meerman JHN, van Gunsteren WF (2000) J Med Chem 43:4594–4605
51. Oostenbrink C, van Gunsteren WF (2003) J Comput Chem 24:1730–1739
52. Oostenbrink C, van Gunsteren WF (2005) Proc Natl. Acad. Sci. U. S. A. 102:6750–6754
53. Mezei M (1987) J Chem Phys 86:7084–7088
54. Jarzynski C (1997) Physical Review E 56:5018–5035
55. Jarzynski C (1997) Phys Rev Lett 78:2690–2693
56. Cossins BP, Foucher S, Edge CM, Essex JW (2008) J Phys. Chem. B 112:14985–14992
57. Cossins BP, Foucher S, Edge CM, Essex JW (2009) J Phys. Chem. B 113:5508–5519
58. Oostenbrink C, van Gunsteren WF (2006) Chem Phys 323:102–108
59. Ytreberg FM, Swendsen RH, Zuckerman DM (2006) J Chem Phys 125:184114
60. Owicki JC, Scheraga HA (1977) Chem Phys Lett 47:600–602
61. Kumar S, Bouzida D, Swendsen RH, Kollman PA, Rosenberg JM (1992) J Comput Chem 13:1011–1021
62. Bartels C, Karplus M (1998) J Phys Chem B 102:865–880
63. Woods CJ, Camiolo S, Light ME, Coles SJ, Hursthouse MB, King MA, Gale PA, Essex JW (2002) J Am Chem Soc 124:8644–8652
64. Wang J, Gu Y, Liu HY (2006) J Chem Phys 125:094907
65. Kong XJ, Brooks CL (1996) J Chem Phys 105:2414–2423
66. Guo ZY, Durkin J, Fischmann T, Ingram R, Prongay A, Zhang RM, Madison V (2003) J Med Chem 46:5360–5364
67. Knight JL, Brooks CL (2009) J Comput Chem 30:1692–1700
68. Laio A, Parrinello M (2002) Proc Natl Acad Sci USA 99:12562–12566
69. Gervasio FL, Laio A, Parrinello M (2005) J Am Chem Soc 127:2600–2607
70. Mezei M (1987) Mol Phys 61:565–582
71. Guarnieri F, Mezei M (1996) J Am Chem Soc 118:8493–8494
72. Resat H, Mezei M (1994) J Am Chem Soc 116:7451–7452
73. Resat H, Mezei M (1996) Biophys J 71:1179–1190
74. Clark M, Guarnieri F, Shkurko I, Wiseman J (2006) J Chem Inf Model 46:231–242
75. Clark M, Meshkat S, Wiseman JS (2009) J Chem Inf Model 49:934–943
76. Mezei M (1980) Mol Phys 40:901–906
77. Clark M, Meshkat S, Talbot GT, Carnevali P, Wiseman JS (2009) J Chem Inf Model 49:1901–1913
78. Chang CE, Gilson MK (2004) J Am Chem Soc 126:13156–13164
79. Chang CEA, Chen W, Gilson MK (2007) Proc Natl Acad Sci USA 104:1534–1539
80. Rekharsky MV, Mori T, Yang C, Ko YH, Selvapalam N, Kim H, Sobransingh D, Kaifer AE, Liu SM, Isaacs L, Chen W, Moghaddam S, Gilson MK, Kim KM, Inoue Y (2007) Proc Natl Acad Sci USA 104:20737–20742
81. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) J Chem Phys 79:926–935
82. Michel J (2006) *The Use of Free Energy Simulations as Scoring Functions*, PhD Thesis, University of Southampton. Available at http://www.julienmichel.net/docs/jmichel-thesiscomp.pdf
83. Pitera JW, Van Gunsteren WF (2002) Mol Simul 28:45–65
84. Steinbrecher T, Mobley DL, Case DA (2007) J Chem Phys 127:214108
85. Beutler TC, Mark AE, Vanschaik RC, Gerber PR, van Gunsteren WF (1994) Chem Phys Lett 222:529–539
86. Zacharias M, Straatsma TP, McCammon JA (1994) J Chem Phys 100:9025
87. Shirts MR, Pitera JW, Swope WC, Pande VS (2003) J Chem Phys 119:5740
88. Michel J, Orsi M, Essex JW (2008) J Phys Chem B 112:657–660
89. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E (1953) J Chem Phys 21:1087–1092
90. Martin MG, Siepmann JI (1999) J Phys Chem B 103:4508–4517
91. Siepmann JI, Frenkel D (1992) Mol Phys 75:59–70
92. Dodd LR, Boone TD, Theodorou DN (1993) Mol Phys 78:961–996

93. Ulmschneider JP, Jorgensen WL (2003) J Chem Phys 118:4261–4271
94. Hetenyi B, Bernacki K, Berne BJ (2002) J Chem Phys 117:8203–8207
95. Chen JH, Im W, Brooks CL (2005) J Comput Chem 26:1565–1578
96. Cooke B, Schmidler SC (2008) J Chem Phys 129:164112
97. Izaguirre JA, Hampton SS (2004) J Comput Phys 200:581–604
98. Mehlig B, Heermann DW, Forrest BM (1992) Phys Rev B 45:679–685
99. Mitsutake A, Sugita Y, Okamoto Y (2001) Biopolymers 60: 96–123
100. Garcia AE, Onuchic JN (2003) Proc Natl Acad Sci USA 100:13898–13903
101. Woods CJ, Essex JW, King MA (2003) J Phys Chem B 107:13703–13710
102. Woods CJ, Essex JW, King MA (2003) J Phys Chem B 107:13711–13718
103. Gelb LD (2003) J Chem Phys 118:7747–7750
104. Michel J, Taylor RD, Essex JW (2006) J Chem Theory Comput 2:732–739
105. Woods CJ, Manby FR, Mulholland AJ (2008) J Chem Phys 128:014109
106. Humphreys DD, Friesner RA, Berne BJ (1994) J Phys Chem 98:6885–6892
107. Jayachandran G, Shirts MR, Park S, Pande VS (2006) J Chem Phys 125:084901
108. Ponder JW, Case DA (2003) Adv Prot Chem 66:27–85
109. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz KM, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) J Am Chem Soc 117:5179–5197
110. MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, Fischer S, Gao J, Guo H, Ha S, Joseph-McCarthy D, Kuchnir L, Kuczera K, Lau FTK, Mattos C, Michnick S, Ngo T, Nguyen DT, Prodhom B, Reiher WE, Roux B, Schlenkrich M, Smith JC, Stote R, Straub J, Watanabe M, Wiorkiewicz-Kuczera J, Yin D, Karplus M (1998) J Phys Chem B 102: 3586–3616
111. Jorgensen WL, Maxwell DS, TiradoRives J (1996) J Am Chem Soc 118:11225–11236
112. Schuler LD, Daura X, van Gunsteren WF (2001) J Comput Chem 22:1205–1218
113. Duan Y, Wu C, Chowdhury S, Lee MC, Xiong GM, Zhang W, Yang R, Cieplak P, Luo R, Lee T, Caldwell J, Wang JM, Kollman P (2003) J Comput Chem 24:1999–2012
114. Wang JM, Wolf RM, Caldwell JW, Kollman PA, Case DA (2004) J Comput Chem 25:1157–1174
115. Udier-Blagovic M, De Tirado PM, Pearlman SA, Jorgensen WL (2004) J Comput Chem 25:1322–1332
116. Jorgensen WL, Jensen KP, Alexandrova AN (2007) J Chem Theory Comput 3:1987–1992
117. Michel J, Tirado-Rives J, Jorgensen WL (2009) J Am Chem Soc 131:15403–15411
118. Deng YQ, Roux B (2008) J Chem Phys 128:115103
119. Michel J, Tirado-Rives J, Jorgensen WL (2009) J Phys Chem B 113:13337–13346
120. Dong F, Olsen B, Baker NA (2008) Biophysical tools for biologists: vol. 1 in vitro techniques, pp. 843–870
121. Mobley DL, Bayly CI, Cooper MD, Shirts MR, Dill KA (2009) J Chem Theory Comput 5:350–358
122. Mobley DL, Dill KA, Chodera JD (2008) J Phys Chem B 112:938–946
123. Michel J, Taylor RD, Essex JW (2004) J Comput Chem 25:1760–1770
124. Barillari C, Taylor J, Viner R, Essex JW (2007) J Am Chem Soc 129:2577–2587
125. Lee MS, Salsbury FR, Olson MA (2004) J Comput Chem 25:1967–1978
126. Allen MP, Tildesley DJ (1990) Computer Simulations of Liquids. Oxford University Press, Oxford, UK
127. Aqvist J (1990) J Phys Chem 94:8021–8024
128. Sagui C, Darden TA (1999) Annu Rev Biophys Biomol Struct 28:155–179
129. Heinz TN, Hunenberger PH (2005) J Chem Phys 123:034107
130. Kastenholz MA, Hunenberger PH (2006) J Chem Phys 124:224501
131. Kastenholz MA, Hunenberger PH (2006) J Chem Phys 124:124106
132. Donnini S, Mark AE, Juffer AH, Villa A (2005) J Comput Chem 26:115–122
133. Aqvist J (1996) J Comput Chem 17:1587–1597
134. Jensen KP, Jorgensen WL (2006) J Chem Theory Comput 2:1499–1509
135. Shirts MR, Mobley DL, Chodera JD, Pande VS (2007) J Phys Chem B 111:13052–13063
136. Shivakumar D, Deng Y, Roux B (2009) J Chem Theory Comput 5:919–930
137. Gresh N, Cisneros GA, Darden TA, Piquemal JP (2007) J Chem Theory Comput 3:1960–1986
138. Lamoureux G, Roux B (2003) J Chem Phys 119:3025–3039
139. Patel S, Brooks CL (2004) J Comput Chem 25:1–15
140. Ren PY, Ponder JW (2003) J Phys Chem B 107:5933–5947
141. Gallivan JP, Dougherty DA (1999) Proc Natl Acad Sci USA 96:9459–9464
142. Jiao D, Golubkov PA, Darden TA, Ren P (2008) Proc Natl Acad Sci USA 105:6290–6295
143. Khoruzhii O, Donchev AG, Galkin N, Illarionov A, Olevanov M, Ozrin V, Queen C, Tarasov V (2008) Proc Natl Acad Sci USA 105:10378–10383
144. Khandelwal A, Lukacova V, Comez D, Kroll DM, Raha S, Balaz S (2005) J Med Chem 48:5437–5547
145. Brown P, Woods C, McIntosh-Smith S, Manby FR (2008) J Chem Theory Comput 4:1620–1626
146. Chipot C (2008) J Chem Theory Comput 4:2150–2159
147. Jorgensen WL, Thomas LL (2008) J Chem Theory Comput 4:869–876
148. Villa A, Zangi R, Pieffet G, Mark AE (2003) J Comput Aided Mol Des 17:673–686
149. Cheng Y, Prusoff WH (1973) Biochem Pharmacol 22:3099–3108
150. Krohn KA, Link JM (2003) Nucl Med Biol 30:819–826
151. Wissner A, Berger DM, Boschelli DH, Floyd MB, Greenberger LM, Gruber BC, Johnson BD, Mamuya N, Nilakantan R, Reich MF, Shen R, Tsou HR, Upeslacis E, Wang YF, Wu BQ, Ye F, Zhang N (2000) J Med Chem 43:3244–3256
152. Chipot C, Pohorille A (2007) Free Energy Calculations: Theory and Applications in Chemistry and Biology. Heidelberg, Berlin
153. Deng YQ, Roux B (2009) J Phys Chem B 113:2234–2246
154. Goodford PJ (1985) J Med Chem 28:849–857
155. Helms V, Wade RC (1998) J Am Chem Soc 120:2710–2713
156. MCPRO + version 2.6 (2009) Schrödinger. LLC, New York, NY
157. Bowers KJ, Chow E, Xu E, Drorr RO, Eastwood MP, Gregerson BA, Klepeis JL, Kollossvary I, Moraes MA, Sacerdoti FD, Salmon JK, Shan Y, Shaw DE (2006) Proceedings of the ACM/IEEE Conference on Supercomputing (SC06), Tampa, Floria
158. Newman J, Fazio VJ, Caradoc-Davies TT, Branson K, Peat TS (2009) J Biomol Screen 14:1245–1250