Towards Accurate Free Energy Calculations in Ligand Protein-Binding Studies

Thomas Steinbrecher[1] and Andreas Labahn*[,2]

[1]*BioMaps Institute, Rutgers University, 610 Taylor Rd., 08854 Piscataway, NJ*

[2]*Institut für Physikalische Chemie, Universität Freiburg, Albertstr. 23a, 79104 Freiburg, Germany*

Abstract: Cells contain a multitude of different chemical reaction paths running simultaneously and quite independently next to each other. This amazing feat is enabled by molecular recognition, the ability of biomolecules to form stable and specific complexes with each other and with their substrates. A better understanding of this process, i.e. of the kinetics, structures and thermodynamic properties of biomolecule binding, would be invaluable in the study of biological systems. In addition, as the mode of action of many pharmaceuticals is based upon their inhibition or activation of biomolecule targets, predictive models of small molecule receptor binding are very helpful tools in rational drug design. Since the goal here is normally to design a new compound with a high inhibition strength, one of the most important thermodynamic properties is the binding free energy ΔG^0.

The prediction of binding constants has always been one of the major goals in the field of computational chemistry, because the ability to reliably assess a hypothetical compound's binding properties without having to synthesize it first would save a tremendous amount of work. The different approaches to this question range from fast and simple empirical descriptor methods to elaborate simulation protocols aimed at putting the computation of free energies onto a solid foundation of statistical thermodynamics. While the later methods are still not suited for the screenings of thousands of compounds that are routinely performed in computational drug design studies, they are increasingly put to use for the detailed study of protein ligand interactions. This review will focus on molecular mechanics force field based free energy calculations and their application to the study of protein ligand interactions. After a brief overview of other popular methods for the calculation of free energies, we will describe recent advances in methodology and a variety of exemplary studies of molecular dynamics simulation based free energy calculations.

Keywords: Protein-ligand binding, free energy calculations, thermodynamic integration, molecular dynamics, computer aided drug design, virtual screening.

1. INTRODUCTION

A cell's metabolism consists of thousands of chemical reactions running in parallel. The ability to run and regulate these complex biochemical networks to maintain homeostasis and reproduce is not only essential for but can be put forward as the very definition of a living organism [1]. Molecular recognition plays a critical role almost everywhere in this process [2]. Enzymes rely on their ability to tightly and specifically bind their substrates and cofactors to provide efficient catalysis, cellular signalling cascades are initiated by small molecule messenger recognition and the activity and expression of proteins is regulated by such a multitude of biomolecule binding phenomena that even a most basic survey of them is a daunting endeavour [3, 4].

Here, we will focus on protein-ligand binding, where ligand stands for any small organic molecule with 'drug-like' properties [5]. Such interactions are of particular interest because the bulk of drug action mechanisms involves the binding of a pharmaceutical compound to a protein target [6]. Without disregarding other possible action mechanisms such as nucleic acid binding drugs [7, 8], therapeutic antibodies [9] or other pharmacologically active molecules, we will define 'drug-design' in this article as the art of finding small organic molecules with acceptable toxicological and ADME (absorption, distribution, metabolism and excretion) properties that are active, i.e. binding with a high affinity, towards a given drug target.

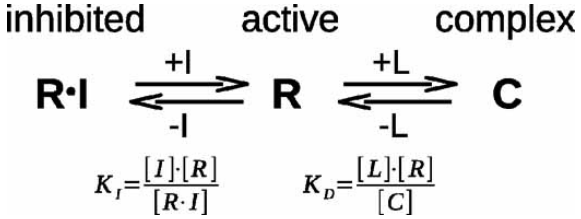


Fig. (1). A minimal model of competitive inhibition where a receptor *R* can either bind a ligand *L* to form the complex *C* or an inhibitor *I*. All binding is reversible to form noncovalent complexes. The binding strength of *L* and *I* is given by their binding constants K_D and K_I, respectively.

Another important property of a promising drug compound, apart from the free enthalpy of binding, is its aqueous solubility. This property is important for e.g. a compound's suitability for oral uptake and its membrane permeability and the development of many early drug candidates is cancelled

*Address correspondence to this author at the Institut für Physikalische Chemie, Universität Freiburg, Albertstr. 23a, 79104 Freiburg, Germany; Tel: +49(761)203-6188; Fax: +49(761)203-6189;
E-mail: andreas.labahn@physchem.uni-freiburg.de

0929-8673/10 $55.00+.00 **© 2010 Bentham Science Publishers Ltd.**

due to insufficient solubility [10]. Theoretical methods to predict solubilities for new compounds are an equally important part of computational drug design (for recent reviews see [11, 12]).

From a thermodynamical perspective, the simplest possible model for ligand binding is the reversible association of a ligand and receptor molecule to form a noncovalent 1:1 complex. The most important thermodynamic quantities describing the binding process are the corresponding standard Gibbs binding free energy ΔG^0_{Bind} and the standard association constant K^0_A[1]:

$$L + R \underset{k_{off}}{\overset{k_{on}}{\rightleftharpoons}} C$$

$$\Delta G_{Bind} = -RT \ln K_A$$

$$K_A = \frac{1}{K_D} = \frac{k_{on}}{k_{off}} = \frac{[C]}{[L]\cdot[R]} \quad (1)$$

in which L, R and C stand for a ligand, free receptor and complex, k_{on} and k_{off} are rate constants and K_A is the complex association constant. Its inverse, the complex dissociation constant K_D is the most widely used measure for a ligand's binding strength. Binding processes can in reality be much more complicated than the simple model given here, due to effects like oligomerisation of receptors, multiple binding sites, allosteric effects, irreversible covalent binding or multistep kinetics to name only a few. Nevertheless, characterizing the activity of a ligand by a single K_D constant is common, even when the binding kinetics are not fully understood.

When the ligand in question acts as a competitive inhibitor I to a second ligand L binding the same receptor, as in Fig. (**1**), its affinity is often given not as an inhibition constant K_I but as an IC_{50} value. This is the inhibitor concentration at which the receptor occupancy is 50% of what it would be in the absence of inhibitor. The inhibition constant K_I and IC_{50} value are related via the Cheng-Prusoff equation [13]:

$$IC_{50} = K_I\left(1 + \frac{[L_0]}{K_D}\right) \quad (2)$$

where $[L_0]$ and K_D are the initial concentration and dissociation constant of the competing ligand L. The derivation of equation 2 assumes that ligand and inhibitor are added in excess of protein and that they bind reversibly and stoichiometrically to the same single binding site on the receptor. If the assay in question was conducted with a ligand concentration far below its K_D, then K_I and IC_{50} are approximately equal. Since the two values are proportional, the same relative (but not absolute) binding free energies for a set of inhibitors will be calculated from either, but only if identical assays were used for all inhibitors. Generally, IC_{50} values are dependent on the conditions under which the inhibition measurement was conducted and experimental data should be interpreted with this consideration in mind.

Drug design has evolved significantly from the trial-and-error approach of earlier times [14]. Modern rational drug design starts from a well characterized target, preferably with structural information from X-ray crystallography. Hundreds of macromolecular drug targets are known today and analysis of the human genome suggests that there may be thousands more to be discovered [15]. Against the selected target a collection of candidate molecules is tested. Such a chemical library can either be one of the standardized sets of thousands of drug-like molecules that pharmaceutical companies maintain [16, 17] or be specifically built by means of combinatorial chemistry for the target in question [18]. High throughput screening (HTS) techniques are then used to find active compounds within the set and once promising lead molecules are identified, their binding affinity and pharmacological properties are improved through chemical optimization until a promising drug candidate is produced [19]. During this optimisation process from early hit to drug candidate molecule, the compound's binding strength will increase by several orders of magnitude, typically from a lower micromolar to below nanomolar K_D.

While the experimental elucidation of all relevant thermodynamic quantities for receptor binding of a ligand is possible [20], it is rarely done in practice. A quantitative in-vitro determination of a lead compound's K_D, using large amounts of time and biochemical sample material, would be prohibitively expensive but of limited usefulness to predict its later in-vivo properties. Therefore, especially in the early HTS stage of drug development, qualitative assays are used to give a fast hit-or-miss assessment of binding affinity. This approach is plagued by generating large numbers of false positive 'hits' [21] and may miss important thermodynamic or kinetic properties [22]. On the experimental side, the emergence of reliable microcalorimetry methods might help to make the measurement of a ligand's full thermodynamic profile more of a standard technique in the future [23, 24]. Alternatively, the tools of computational chemistry can be used as additional filters for chemical libraries and to generate supporting data through every step of the drug design process [25].

Virtual or *'in silico*' screening techniques aim at predicting the binding strengths or complex structures of new ligands from available data, empirically optimized models or physico-chemical first principles [26, 27, 28]. The appeal of studying novel compounds without the need to actually synthesize them first was apparent early on in the development of computational chemistry and has led to a broad and lively field of study [29]. Despite enormous advances in the last decades, the prediction of binding properties is still a precarious task [30, 31], but the number of computationally designed drugs on the market [32, 33], while still small, is rising. The efforts of computational drug designers can be roughly grouped into three different approaches, listed here in order of increasing computational effort, required structural input and predictive power:

- Quantitative Structure Activity Relationship (QSAR) methods try to predict binding affinities from simple

[1]These free energy and equilibrium constants refer to the appropriate standard state of 1M reactants at 1 bar and 298.2 K. All thermodynamic quantities given in this work are standard free energies and equilibrium constants and should strictly speaking be written as ΔG^0 and K^0. For improved legibility, the authors have consistently omitted the corresponding thermodynamic state from here on. Nevertheless, the proper standard state is implied in each case.

structural and physical properties of molecules [34, 35]. Commonly used properties or *descriptors* include the octanol-water partition coefficient, approximate surface area, or polarity of a molecule, but many more are in use [36]. Statistical regression techniques are used to find correlations between the molecular descriptors and affinities for a test set of molecules with known properties. Once a valid prediction function has been established, it can be used to compute properties for new molecules. The results of QSAR methods depend strongly on using significant descriptors and on a training molecule set that covers a sufficient range of chemical space. Both of these requirements are hard to meet. QSAR models are widely applied in drug design because they are fast, allow for the incorporation of diverse experimental data and require no knowledge of receptor structures and complex geometries. More recent QSAR models have begun to utilize 3D-structural information in their predictions [37, 38].

- Ligand Docking calculations aim at computing binding geometries and energies of ligands to a known receptor structure with no or little prior knowledge of the binding modes [39]. Docking programs rely on efficient heuristic ligand placement algorithms and fast empirical scoring functions to minimize the computational effort needed for each ligand. Several well established docking tools exist, such as Dock [40, 41], AutoDOCK [42], FlexX [43, 44], GOLD [45] or Glide [46, 47], that implement different algorithms to solve the docking problem. Traditionally, docking involves the placement of a flexible ligand into a static receptor binding site, but in recent times techniques to include at least some measure of receptor flexibility have been developed [48-52].

- Free energy calculations try to compute free energies for molecular systems based on the principles of statistical thermodynamics. Established molecular mechanics force fields and algorithms are used to describe a system's dynamics and energetics. Thereby, free energy calculations automatically include both ligand and receptor flexibility. No case-by-case parameter fitting is performed, so that the quality of results should be independent of the system studied. Free energy calculations commonly involve conducting extensive computer simulations of the studied biomolecules requiring computational efforts several orders of magnitude higher than the above two techniques. To justify this expense, results must be reliable and close to quantitative. In comparison to QSAR and ligand docking approaches, free energy calculations to compute binding affinities have only recently become employed in drug design applications.

Within the realm of free energy calculation methods, a further subdivision can be made into molecular alchemy methods, potential of mean force calculations and the newer endpoint methods like e.g. MM-PBSA. The differences between these will be discussed in the Theoretical Background section below. The endpoint type of calculations is somewhat less expensive in terms of computational effort than the first two categories and it is uncertain if they can be performed with the same level of rigor as "traditional" free energy calculations. Nevertheless, all these approaches are generally referred to as *free energy* calculations.

This review will focus on the last type of methods, aimed at calculating binding free energies based on the principles of statistical thermodynamics. The QSAR and ligand docking approaches have been extensively reviewed elsewhere, as have earlier applications of the free energy techniques described in the following [53-59]. We will present a brief summary of the various free energy calculation methods, the underlying principles of which have mostly been known for a long time. This will be followed by a description of several recent studies, to show how advances in modern computer simulations have made it possible to apply free energy calculations to study protein ligand binding. Plenty of examples for free energy calculations in other areas of biomolecular studies exist [60, 61], but we limit our selection to examples of protein-small molecule interactions. The authors acknowledge that any selection of examples out of a field as diverse as this is strongly subjective and can make no claim of exhaustiveness.

2. THEORETICAL BACKGROUND

Free energy calculations rely on the fundamental relationship between the Helmholtz free energy and the configuration integral (or partition function). It is furthermore usually assumed that, when condensed phase systems are studied, a state's Standard Gibbs and Helmholtz free energy are approximately equal:

$$G^0 \approx F^0 = -k_B T \ln Z \; ; \; Z = \sum_i e^{-E_i / k_B T} \tag{3}$$

While the partition function Z of even a moderately complex system cannot be calculated explicitly, it can be approximated from an ensemble of Boltzmann-weighted conformational snapshots. These ensembles can be generated via Monte Carlo or Molecular Dynamics simulations. For the extremely complex conformational spaces of biomolecules severe sampling problems can occur, but even in cases where statistical or systematic error generates flawed conformational ensembles, re-weighting schemes can be used to transform them into Boltzmann-weighted distributions [62].

The method used to generate the conformational ensemble is rather unimportant; both techniques will give identical answers in the limit of infinite sampling. Since MD simulations appear to be used more often in molecular modelling applications, they will be briefly summarized in the following. For a more detailed description, see e.g. [63-65].

MOLECULAR MECHANICS FORCE FIELDS AND DYNAMICS SIMULATIONS

The Molecular Mechanics (MM) model makes use of force fields, which are collections of parameters and potential functions to describe a chemical system in terms of predefined atom types and topologies. The total energy is normally expressed as a sum of simple empirical potentials which are additive as described in equation 4.

Electrons are not treated explicitly in MM calculations; instead their effects are accounted for implicitly via harmonic or trigonometric potentials for bonded atoms and via the Coulomb and van-der-Waals equation for non-bonded atoms. As an example, the potential energy equation for the Amber suite [66, 67] of force fields takes the form:

$$E_{FF} = \sum_{bonds} K_r\left(r - r_{eq}\right)^2 + \sum_{angles} K_\theta\left(\theta - \theta_{eq}\right)^2 + \sum_{dihedrals} \frac{V_n}{2}\left[1 + \cos\left(n\phi - \gamma\right)\right] + \sum_{i<j}\left[\frac{1}{k_{ex}}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + \frac{1}{l_{ex}}\frac{q_i q_j}{\varepsilon r_{ij}}\right] \quad (4)$$

in which K_r, K_θ, V_n, r_{eq}, θ_{eq}, n, φ and γ are parameters describing bonded interactions, A_{ij} and B_{ij} are van der Waals parameters and q_i, q_j atomic partial charges. k_{ex} and l_{ex} are constants describing the exclusion of certain atoms from non-bonded interactions: van-der-Waals and Coulombic potentials are normally not evaluated for atoms connected by one or two intervening bonds and they are reduced by a given factor for atoms separated by three bonds (In the recent Amber force field, k_{1-4} is 2.0 and l_{1-4} is 1.2). Several force fields exist that are designed to describe biochemical systems, the most common of which are the Amber [68-71], CHARMM [72], GROMOS [73] and OPLS [74] force fields. In addition to these, there are non-additive or polarizable force fields like Amber ff02 or AMOEBA [75, 76]. While polarizable force field are being used in computational drug design studies [77], most protein-ligand binding calculations do use simple additive force fields for MD simulations and our discussion will focus only on these types.

To accurately describe condensed phase systems under standard conditions, various algorithms are employed to control the system temperature and pressure. To simulate solvated systems, the aqueous environment can be described by an implicit solvent such as Generalised Born [78-85] or Poisson-Boltzmann models [86, 87] or explicitly, in which case a simulation box containing the system including a shell of water molecules is infinitely replicated in all directions via periodic boundary conditions. Non-bonded interactions are then only explicitly calculated up to a given cutoff radius and corrections for long range interactions are accounted for through e.g. Ewald sum calculations [88, 89].

A molecular dynamics simulation uses a force field equation to compute the time evolution of a system according to Newton's equations of motion. The state of the system at any time *t* can be derived from a given starting conformation via a Taylor expansion of the atomic coordinates' time dependence [90]:

$$\vec{r}\left(t + \Delta t\right) = \vec{r}\left(t\right) + \vec{\upsilon}\Delta t + \frac{\vec{f}}{2m}\left(\Delta t\right)^2 + O\left(\left(\Delta t\right)^3\right) \quad (5)$$

The propagation of the system via equation 5 usually can not be done analytically and numerical methods are used in which the system is propagated in increments of a given time step Δt using various integrator algorithms. Since the higher order terms in equation 5 are normally neglected, values for Δt are limited by the desired simulation accuracy. Δt should be smaller than the period of the fastest oscillation in the system and is normally about 1-2 fs. Larger time steps of 2 fs are possible if the fast stretching vibrations of bonds containing hydrogen atoms are constrained by algorithms like SHAKE [91].

The rapid increase in computer power and better use of parallel computing in the last decades [92] allows MD simulations in the multi-nanosecond range on systems containing tens of thousands of atoms to be routinely performed in hours or a few days at most. This enables simulations of most proteins and large nucleic acid oligomers and binding constants could in principle be calculated by simulating enough biomolecule-ligand binding and unbinding processes to determine the k_{on} and k_{off} rates. However, the available timescales are still far below of what would be necessary to observe spontaneous binding in MD simulations (the highest association rates are in the $10^7 M^{-1}s^{-1}$range [93] and half-time rates for protein-ligand complex dissociation can be as high as hours or days [94]). Therefore, more complex techniques must be employed to calculate binding free energies.

COMPUTATIONAL ALCHEMY CALCULATIONS

This work focuses on rigorous free energy calculations based on statistical thermodynamics which generally speaking aim at computing the free energy difference between two states or chemical systems, A and B. Since the start and end states can be in principle arbitrarily different, free energies for chemically impossible reactions can be computed, e.g. transitions from A to B which change the number of atoms or their chemical properties. Therefore these calculations are also referred to as *computational alchemy*. In terms of computational setup, the two states are represented by two potential functions, V_A and V_B, and exchanging the potential function used to describe the system model during simulation is a way of transforming it from one into the other. Below, different analysis approaches via which a free energy change can be computed from such a transformation are described, but first, the concept of a thermodynamic cycle with embedded free energy calculation steps will be discussed.

Alchemical free energy calculations describe processes that are not necessarily experimentally accessible and the computed free energy changes can only be determined with respect to an arbitrary zero of energy. Therefore, for these calculations to have any practical use, they need to be combined in thermodynamic cycles so that experimentally relevant data can be obtained. These cycles involve performing two or more transformations in a way that their free energy difference becomes a meaningful quantity. This means that such calculations can only compute relative free energies and are best suited to determine the change of a physical quantity when a compound's environment changes.

As an example, Fig. (**2**) shows one possible thermodynamic cycle, in which the binding free energy of two ligands (A and B) to a receptor (R) is compared. Here, the free energy calculations are set up so that the transition from start to end state corresponds to changing ligand A into ligand B and the transformations are simulated both for the ligand bound and unbound to its receptor. The free energy difference be-

tween the two results is then equal to the binding free energy difference of A and B:

$$\Delta G_{unbound}(A \to B) - \Delta G_{bound}(A \to B) = \Delta\Delta G_{Bind} = \Delta G_{Bind}(A) - \Delta G_{Bind}(B) \quad (6)$$

With the thermodynamic cycle set up as shown, if the free energy cost of transforming ligand A into B were larger in the unbound than in the bound form, ligand B would bind stronger to the receptor than ligand A. Free energy calculations are therefore able to predict the effect of ligand modifications on their binding free energies. This can be a very useful tool for drug design applications, where it is necessary to improve the binding strength of a lead compound by systematic chemical modification [95]. Apart from the process depicted in Fig. (**2**), other thermodynamic cycles can be designed to compute different physical properties such as pK_a-changes of deprotonatable groups, solvation free energies and the effects of amino acid point mutations on protein stability.

In practice, a single alchemical transformation is often broken down into several steps, e.g. first changing the partial charges on all disappearing atoms to zero and then in a second independent calculation removing the chargeless atoms' van-der-Waals interactions with their surroundings. One reason for this two-step procedure is that removing the van-der-Waals potential of an atom with a nonzero charge can lead to simulation instabilities, the other is that the two free energies computed for the two sub steps can give an indication of the relative importance of electrostatic and van-der-Waals energies in the studied transition. This interpretation should not be overrated however, since the free energy contributions for the substeps depend on the order in which they are executed. They are not state functions and their physical interpretation is therefore ambiguous.

One particular application, the calculation of absolute binding free energies, deserves further mentioning: In principle it is possible to compute a ligand's absolute binding free energy by having it completely disappear during the transformation, i.e. transforming it into 'nothing'. In practice this often involves transforming the ligand into so called 'ghost' or 'dummy' particles that do not interact with their surroundings, which is equivalent to removing the particles themselves since any energy contribution from the ghost particles would exactly cancel in the two calculations. There is one problem associated with this procedure. While completely removing a molecule from the system poses no difficulties in isotropic surroundings like solution, it is problematic for a ligand bound to a receptor binding site. Near the end state of the transformation (which translates into λ-values close to one in the TI formalism below) the almost fully decoupled ligand only weakly interacts with its surroundings anymore and starts to drift away from its binding position. This leads to severe convergence problems and must be avoided by adding a suitable set of positional restraints before disappearing the ligand and accounting for their free energy effect. Several applications of this procedure have been published recently and the technique shows promising results at least for small test ligands [96-102].

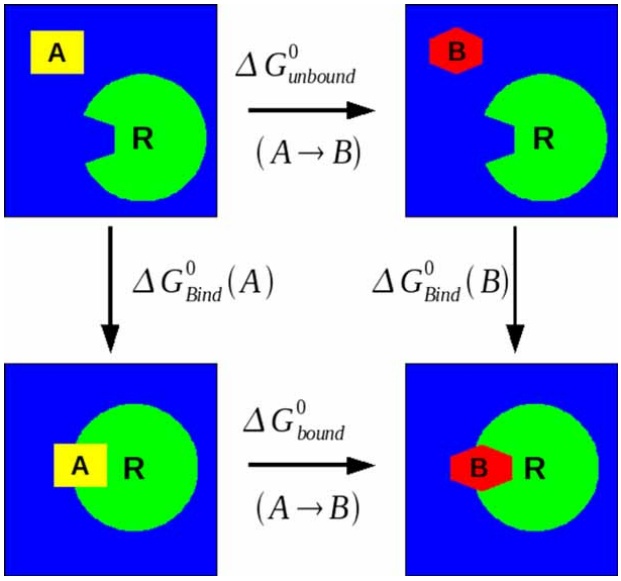


Fig. (2). Thermodynamic cycle to compute changes in binding free energy via free energy calculations. In this figure, horizontal arrows indicate alchemical transformations and vertical arrows indicate ligand binding processes. The difference in the free energy change of the two free energy calculations is equal to the difference in binding free energy for ligands A and B. The unbound transformation does not need to actually contain the receptor and normally just simulates the ligand transformation in a box of water molecules.

DATA ANALYSIS SCHEMES TO COMPUTE FREE ENERGIES

Computing free energy changes associated with a computational alchemy transformation is in principle a straightforward process, but the peculiarities of molecular interactions and the force fields used to describe them introduce quite particular problems of sampling and statistical convergence of the results. The development of optimal analysis schemes for free energy calculations is an ongoing process and there have been a few exiting recent developments. We will begin the description of analysis schemes with the Thermodynamic Integration and related Free Energy Perturbation approaches, which are still employed in the majority of applied computational alchemy studies, followed by a discussion of newer schemes that appear to be of superior efficiency, like the Bennett acceptance ratio method.

The free energy difference between two states can be evaluated with the Zwanzig formula [103] or Free Energy perturbation (FEP) approach. Starting from equation 3 it can be easily shown that:

$$\Delta G^{FEP} = G_B - G_A = -k_B T \ln\left\langle e^{-[V_B - V_A]/k_B T}\right\rangle_A = +k_B T \ln\left\langle e^{-[V_A - V_B]/k_B T}\right\rangle_B \quad (7)$$

where V_A and V_B denote the potential functions of states A and B and the angular brackets indicate that the exponential term should be evaluated over a Boltzmann-weighted ensemble average generated according to the potential function of the indicated state. This gives FEP calculations a sense of direction, in that the potential energy differences can be av-

eraged over an ensemble generated using the start or end state potential function for a forward and backward transition, respectively. This allows estimating the convergence from comparing free energy results of forward and backward transformations.

A calculation according to equation 7 will only converge within a reasonable time if the thermally accessible phase space of state A has considerable overlap with that of state B, i.e. if a conformational ensemble generated with potential function V_A samples all relevant conformations of state B, too. In practice, as this is rarely the case, a FEP calculation is broken down into several small steps by simulating the transition from A to B via non-physical intermediate states and obtaining the total free energy as the sum of all free energy changes for transitions between intermediates. This approach of breaking down a transition into multiple smaller steps, or windows, shares some similarities with an alternative way to analyse free energy calculations, the Thermodynamic Integration (TI) approach [104].

In a TI calculation the two states of interest are considered to be connected via an additional non-physical coordinate, commonly called λ, and a transition from system A to B along this non-physical reaction coordinate is simulated. At any point along λ the system is described by a mixed potential function $V(\lambda)$, chosen so that it corresponds to V_A for $\lambda = 0$ and V_B for $\lambda = 1$. The free energy difference between the states can then be obtained by integrating the Boltzmann-weighted λ-derivative of the mixed potential function over λ:

$$\Delta G^{TI} = \int_0^1 \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_\lambda d\lambda \approx \sum_i \omega_i \left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle_{\lambda_i} \tag{8}$$

The last part of the equation indicates that, since the integration can not be performed in closed form, the integral is approximated numerically by evaluating $\langle \partial V(\lambda)/\partial\lambda \rangle$ at various fixed values λ_i and by constructing the free energy curve via linear or higher order interpolation [105]. A TI calculation therefore consists of multiple independent simulation windows that compute the gradient of the free energy curve at given λ-points. λ will always be considered a static variable here, but dynamic λ-approaches exist as well [106]. Unlike calculations using the FEP formalism, there are no forward and backward transformations in TI calculations.

There are many possible ways in which the mixed potential function $V(\lambda)$ can be constructed, the easiest of which is a linear interpolation between the two end state's potential functions which also reduces the free energy expression to a simple form:

$$V(\lambda) = \lambda V_B + (1-\lambda)V_A \; ; \Delta G = \int_0^1 \langle V_B - V_A \rangle_\lambda d\lambda \tag{9}$$

Even though there are many possible ways to couple the two states via mixing their potential functions [107], the resulting free energy change for the transition $A \rightarrow B$ is a state function and path independent. Choosing the optimal mixing function for a given system is therefore only a question of convergence and simulation stability.

If equation 9 is used to conduct TI calculations, the much discussed 'endpoint catastrophe' problem [108, 109, 110] can occur: When the two states differ in the number of van-der-Waals particles, the computed $\langle \partial V/\partial\lambda \rangle$-values diverge for high or low values of λ (The divergence is mild enough so that the overall integral in equation 8 stays finite). This behavior is caused by the functional form of the van-der Waals potential, the r^{-12} repulsive term of which causes large potential energy fluctuations when scaled down linearly. Several solutions for this problem have been proposed, such as slow-growth calculations [111], bond-shrink protocols [112] or non-linear mixing functions. One form of the later are so called soft core potentials that use a λ-dependent form of the van-der-Waals equation and efficiently solve the endpoint singularity problem [113, 114].

Free energy perturbation can be considered a special case among methods to compute free energies from non-equilibrium work distributions, but the exponential averaging in equation 7 is far from statistically optimal[2]. Switching to the thermodynamic integration formalism of computing free energies to avoid this problem only works if smooth and converged free energy curves constructed from $\left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle$-values are obtained, which is difficult when large chemical changes are simulated. Alternatively, several suggestions have been made for more efficient free energy estimation schemes.

Originally developed in the framework of Monte-Carlo simulations, the Bennett acceptance ratio (BAR) method [115] offers an improved way to estimate free energy differences between two states when both the work for the forward ($A \rightarrow B$) and the backward ($B \rightarrow A$) transformation is computed. From this data, the ratio of the two state's configuration integrals Q_A and Q_B is given by:

$$\frac{Q_A}{Q_B} = \frac{\langle M(V_A - V_B) \rangle_B}{\langle M(V_B - V_A) \rangle_A} \tag{10}$$

where angular brackets as usual denote Boltzmann averaging. V_A and V_B are the two potential functions and M represents the Metropolis function $M(x) = min\ \{1, \exp(-x)\}$ [116] which was used in the first formulation of equation 10. Any other (finite) weighting function can also be used instead of M, and from the expected square error in a computed free energy, Bennett could show that the Fermi function $f(x)=1/(1+\exp(x))$ minimizes the expected square error, especially if one of the potential functions is shifted by a constant value C. C depends on the sample size and the ratio of the configuration integrals and can be determined self-consistently from simulation data. Although available for some time, BAR has only recently been further investigated and applied to free energy calculations.

Its connection to other thermodynamic perturbation theorems [117] has been explored and Shirts *et al.* presented a

[2] Note that the potential energy difference between the two states V_B-V_A is equal to the work needed to instantaneously exchange their potential functions.

Table 1. Three TI Calculations are Performed in Total. All three resulting free energy contributions taken together give the solvation free energy of acetaminophen. $(AAP)^0$ symbolizes a chargeless molecule with all atomic partial charges set to zero. The fourth step of transforming chargeless AAP into nothing in vacuum need not be performed as the specific TI implementation used sets its corresponding free energy to zero. The index EEL indicates that the corresponding TI transformation changes only electrostatic properties, while VDW indicates a transformation that changes van-der-Waals and bonded potentials as well as atom numbers. The resulting free energies are given in kcal/mol.

	Start State (V_0)	End State (V_1)	Free Energy	Result
Step 1	AAP in water	$(AAP)^0$ in water	$\Delta G_{EEL,\,Solution}$	141.8
Step 2	$(AAP)^0$ in water	pure water	$\Delta G_{VDW,\,Solution}$	-0.8
Step 3	$(AAP)^0$ in vacuum	AAP in vacuum	$\Delta G_{EEL,\,Vacuum}$	-128.5
(Step 4)	$(AAP)^0$ in vacuum	vacuum	$\Delta G_{VDW,\,Vacuum}$	$\equiv 0$

new interpretation of BAR in terms of maximum-likelihood estimators [118]. They could show that use of the acceptance ratio method will always yield a lower variance than exponential averaging as in FEP. This is of particular interest to applications, since a set of data collected for a typical FEP calculation (forward and reverse transformations) could directly be used for BAR analysis as well. Additionally, an extension to the technique comparing multiple equilibrium states, called MBAR, was developed [119]. It uses the fact that for most free energy calculations several intermediate states are considered between the initial and final one and MBAR attempts to extract an optimal free energy estimate by using data from a comparison of all these states.

These advanced data analysis techniques are close to being statistically optimal ways of computing free energies from potential energy differences and it has been suggested that they are significantly more efficient than TI calculations at least for the study of larger molecular changes [120]. It is likely that they will find broad application in the near future, maybe replacing Thermodynamic Integration as the most widely employed data analysis scheme in MD-based statistical mechanics free energy calculations.

TI CALCULATION EXAMPLE: PREDICTING THE SOLVATION FREE ENERGY OF ACETAMINOPHEN

In the following we will give a detailed example of how a typical free energy calculation, for a simple test system, would be performed. Evaluating relative binding free energies for different ligands according to the thermodynamic cycle in Fig. (**2**) often involves complex preparations of the system to ensure that a correct receptor structure and initial complex binding mode are used. Additionally, MD simulations of biomolecules are computationally expensive and require the recording of long trajectories to include sufficient sampling. We will therefore give a simpler example for a typical thermodynamic integration calculation that nevertheless yields a thermodynamic quantity of pharmaceutical interest: the absolute solvation free energy for the small molecule acetaminophen (or paracetamol, a widely used analgesic).

To computationally predict this quantity, two TI transformations need to be performed: In the first an acetaminophen molecule (AAP) is transformed into 'nothing' while embedded in a box of water molecules. This transformation yields the free energy contribution of AAP-water interactions plus the internal free energy of acetaminophen. Since only the first of these two terms is needed, a second TI calculation is performed, namely transforming AAP into 'nothing' in vacuum. The difference between the two TI results will be interpreted as our estimate of the molecule's solvation free energy. This approach neglects free energy contributions from polarisation effects and changes in molecular vibration modes upon removal from the solvent.

The first transformation in water was further broken down into two substeps, one in which only the atomic partial charges of the molecule are removed (step 1 in table 1) and a second one in which the chargeless acetaminophen molecule is transformed into nothing (step 2). In vacuum, only the charge removal step needs to be performed (step 3), because all other contributions to the internal free energy of the molecule (such as van-der-Waals, bond, angle and dihedral terms) are automatically excluded from the $\left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle$-term in the particular software implementation used (version 10 of the Amber modelling suite). This sets the free energy change of transforming a chargeless molecule to nothing as zero (step 4). This is valid because these free energy contributions occur in both TI calculations and cancel in the final result. *Chargeless* here means not only a neutral molecule but one with a partial charge of zero on every atom. Table 1 summarizes the different calculations performed, the respective start and end states and the free energy contributions obtained.

The total solvation free energy is then given by:

$$\Delta G_{Solv} = -\left(\Delta G_{EEL,Solution} + \Delta G_{VDW,Solution} + \Delta G_{EEL,Vacuum}\right) \quad (11)$$

The minus sign is necessary because the simulation computes the free energy change for moving an acetaminophen molecule from aqueous solution into vacuum and the solvation free energy describes the reverse process.

The system was prepared using the gaff force field [121], RESP charges [122] for the acetaminophen molecule and a solvation box containing 938 TIP3P water molecules [123]. The system was rapidly brought to simulated conditions of 300 K and 1 bar pressure and further equilibrated over 1 ns of MD simulation time. $\left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle$-values were collected over

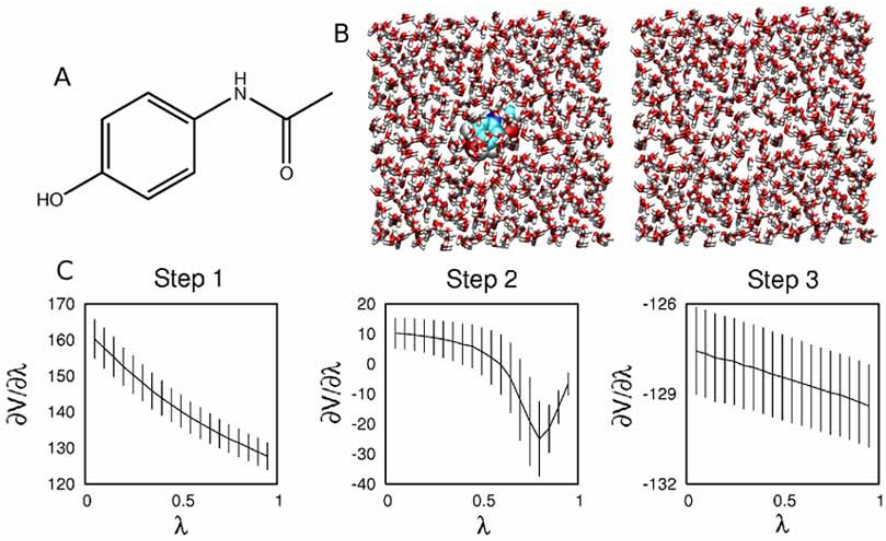


Fig. (3). Solvation free energy of acetaminophen (A). The start and end state for Step 2 of the transformation (see table 1) correspond to a chargeless acetaminophen molecule in a box of 938 water molecules, and a box of the same number of water molecules without the solute (B). The transformation simulates removing the molecule from water and computes the free energy contribution of this process. The three free energy curves for the three substeps are markedly different (C). The curves corresponding to electrostatic changes (Step 1 and 3) are fairly smooth but contain large absolute values. In contrast, the Step 2 transformation yields a highly nonlinear curve. $\left\langle \frac{\partial V(\lambda)}{\partial \lambda} \right\rangle$-values are given in kcal/mol.

a subsequent 1 ns length simulation under NTP conditions. The simulation protocol used was adapted from previous work [114] and uses the modified λ-dependent van-der-Waals equation (*soft core* potentials) mentioned there. 19 λ-values were used for each transformation (spaced equally from 0.05 - 0.95) and the free energy curves were integrated by linear interpolation.

The results add up to a total solvation free energy of -12.6 kcal/mol, in good agreement with acetaminophen's polar nature and moderate solubility in water [124]. As a comparison, computing the solvation free energy via continuum solvent methods yields ΔG_{Solv}-values of -19.8 kcal/mol for Poisson-Boltzmann calculations [86, 87] and -18.9 kcal/mol using a Generalized Born solvation model [125, 126]. Moving from a simplified continuum solvent representation to an explicit description of water-solute interactions results in a markedly different free energy. Absolute solvation free energies are hard to measure and the authors are not aware of such experiments for acetaminophen, but a direct comparison with experimental data could be facilitated by performing a second TI calculation to compute the solvation free energy of acetaminophen in octanol, yielding the octanol-water partition coefficient for the molecule.

Note that the two ΔG_{EEL}-values have large absolute values but mostly cancel each other, indicating that they mainly stem from internal electrostatic interactions. The ΔG_{VDW}-value is negative, mainly because it includes the important entropy effect of removing the solvent exclusion volume of the molecule. Fig. (**3**) contains a snapshot of the start and end states for Step 2 as well as the free energy curves from all transformations. Transformations involving electrostatic changes produce fairly linear curves, but the removal of chargeless AAP from water in Step 2 result in a highly nonlinear curve, which is typical for TI calculations utilizing the *soft core* potentials mentioned above. In subsequent steps, a more accurate estimate of the predicted free energy could be obtained by performing additional or longer simulations at selected λ-values, especially were the free energy curves are particularly steep.

ALTERNATIVE FREE ENERGY CALCULATION METHODS

Potential of Mean Force Calculations

The FEP and TI formalism described above mostly deal with free energies for chemical changes along non-physical or alchemical reaction coordinates. They are complemented by a set of methods aimed at computing the free energy change associated with conformational changes along a physically possible path. Their results are *Potentials of Mean Force* (PMF), or free energy profiles calculated along some designated reaction coordinate. The underlying principle is straightforward and gives the free energy of the system as a function of a designated coordinate x_i:

$$G(x_i) = -k_B T \ln P(x_i);\ P(x_i) = \underbrace{\int e^{-V(\vec{x}^{\ddagger})} d\vec{x}^{\ddagger}}_{=Z(x_i)} / Z \qquad (12)$$

Here, $P(x_i)$ is the probability of finding the system at a given point along the coordinate, which depends on a modified partition function $Z(x_i)$ in which $\vec{x}^{\ddagger}$ indicates that the integration is performed over all degrees of freedom, except x_i. Equation 12 is given in the classical approximation where the sum in equation 3 is replaced with an integral. The probability $P(x_i)$ could in principle be directly extracted from a Boltzmann-weighted conformational ensemble of the system. Unfortunately, this is rarely possible in practice, because even small potential barriers along the reaction coordi-

nate will prevent sufficient sampling to take place in an MD simulation of the system. Therefore, the *umbrella sampling* approach [108, 127] introduces additional biasing potentials to improve sampling and speed up the convergence of PMF calculations.

This is done by dividing the reaction coordinate into distinct windows and running a separate simulation for each window with a harmonic potential added to the system to enforce sampling in a given region. From a simulation using this modified potential function $V^{\dagger}$ a free energy profile $G^{\dagger}(x_i)$ is computed. The unbiased free energy profile can be computed from $G^{\dagger}(x_i)$ because the form of the biasing potential is known:

$$G(x_i) = G^{\dagger}(x_i) - k(x - x_W)^2 \quad (13)$$

Here, k and x_w are the force constant and midpoint of the biasing potential, optimal values of which, together with the optimum number of simulation windows, must be empirically determined for each system. Each simulation window yields a portion of the complete free energy curve from which the complete free energy profile can be constructed if the phase spaces sampled by adjacent windows have sufficient overlap. For each pair of neighboring windows, an optimal constant energy offset K_i must be found for which they overlap best. This is often done via the weighted histogram analysis method (WHAM) [128, 129], which uses error minimization techniques to calculate the best estimate for the total free energy profile.

An alternative way to compute a free energy profile along a physical reaction coordinate is the Jarzynski relationship [130]. This method is distinct from every other approach presented here in that it allows the calculation of equilibrium free energies from non-equilibrium simulations. The equation can be considered a generalization of previous free energy relationships which simplifies to equations 7 and 8 in the limiting cases of instantaneous and reversible transitions. It gives the free energy difference between two states separated by a physical reaction coordinate as:

$$\Delta G = -k_B T \ln\left\langle e^{(-W/k_B T)} \right\rangle \quad (14)$$

where W is the work necessary to transform the system from the starting to the final conformation and the angular brackets indicate an average over all possible transition pathways starting from a Boltzmann-weighted ensemble of starting states. The method is also referred to as the "Fast Growth" algorithm and a recent study showed its accuracy to be comparable to converged TI calculations in biochemically relevant types of transformations [131]. It was shown that equation 14 is a special case of a more general theorem of non-equilibrium perturbation theory [117]. In Crooks generalisation, an expression connecting average non-equilibrium work values for forward and backward processes with free energies was derived that extends Jarzynskis equation in a similar way that the Bennett acceptance ratio method extends the FEP approach.

The Jarzynski relationship has been studied intensely over the last decade and many recent applications exist [132, 133, 134, 135]. The method complements the alternative equilibrium free energy methods nicely in that it replaces the requirement for extensive sampling at any point in a reaction path with the necessity to conduct many non-equilibrium transitions from different starting points, making it easily parallelisable [136]. It should be noted that while equation 14 is most often thought of in terms of "pulling" experiments in which a potential of mean force is calculated along a predefined real-space reaction coordinate, there exists no principle obstacle in applying them to the types of alchemical calculations discussed above.

Endpoint Methods to Calculate Free Energies

Endpoint methods, unlike the previously described approaches, attempt to compute the free energy difference between two states from simulations of these states only, with no consideration of either physical or non-physical intermediates.

One such approach to the calculation of free energies is formulated in the MM-PBSA scheme that combines conformational ensembles from molecular dynamics simulations with a continuum solvent model. The underlying ideas have been presented in [137, 138] and many applications of the method have been published [139-143], but the main ideas can be summarized as following:

The approach aims at calculating binding free energies in solution as the sum of the vacuum binding free energy of a free ligand L, free receptor R and ligand-receptor complex C and a correction for the difference in solvation free energy between the species:

$$\Delta G_{Bind}^{MM-PBSA} = \Delta G_{Bind}^{Vacuum} - \left[\Delta G_C^{Solv} - \left(\Delta G_R^{Solv} + \Delta G_L^{Solv} \right) \right] \quad (15)$$

The free energy contributions are calculated for each member of an MD generated conformational ensemble (commonly using 100-1000 MD snapshots obtained at 1-10 ps time intervals) and averaged. The conformational snapshots for the complex, ligand and receptor are usually extracted from a single MD simulation of the complex. Therefore, MM-PBSA calculations assume that the average structure of the receptor and ligand are identical for the bound and unbound forms and that no major conformational changes occur upon ligand binding. While this could be easily remedied by running three independent MD simulations for the three species, the multi-trajectory approach leads to much slower convergence of the resulting binding free energies and seems not be used extensively in practice.

The vacuum binding free energy in equation 15 is calculated from enthalpic and entropic contributions:

$$\Delta G_{Bind}^{Vacuum} = \left\langle H_C - \left(H_R + H_L \right) - T\left[S_C - \left(S_R + S_L \right) \right] \right\rangle_{MM} \quad (16)$$

where the angular brackets denote Boltzmann-weighted averaging. The enthalpy H is given by the molecular mechanics force field and the gas phase entropy S is calculated from its translational, rotational and vibrational contributions using standard statistical thermodynamics models. The contribution of the vibrational degrees of freedom, approximated to be quasi-harmonic, is obtained from normal mode analysis.

The solvation free energy contributions are computed by a continuum solvent model to avoid having to explicitly average over all solvent degrees of freedom. The original for-

mulation of the MM-PBSA approach used Poisson-Boltzmann calculations for the electrostatic component of ΔG_{Solv} and an empirical term linearly dependent on the solvent accessible surface area (SASA) [144] for the hydrophobic component:

$$\Delta G_{Solv} = \Delta G_{PB}^{Electrostatic} + \Delta G_{SASA}^{Hydrophobic} \quad (17)$$

Alternatively, the solvation free energy could be computed from Generalized Born models (MM-GBSA) or by directly averaging the solute-solvent interactions and assuming a linear solvent response [145].

In order to compute standard free enthalpies of binding, the MM-PBSA method assumes ideal solutes with activity coefficients of unity and negligible pressure work for binding of two molecules in condensed phase systems. In addition, since ΔG in equation 15 is calculated by combining results from gas-phase interaction enthalpies, implicit solvent solvation free energies and gas-phase entropies, it is not immediately clear how the correct standard state of 1M solutes at room temperature and pressure should be accounted for. It is quite complex to rigorously derive an expression for the standard free energy of binding from elementary statistical thermodynamics with full consideration of all approximations made (see Gilson *et al.* for an exhaustive review of the subject [97]). Of particular interest here is a correct representation of the change in entropy upon binding. The loss of translational and rotational entropy when a free receptor and ligand combine into a single complex is a large negative contribution to the binding affinity and it is an open question if these entropy terms should be computed via the Sackur-Tetrode equation as in *e.g.* [146] or via alternative models [147].

One problem of MM-PBSA is that its results are computed as differences of very large energies computed by different models. Especially the electrostatic interaction energy and the change in solvation free energy upon binding tend to cancel each other to a certain degree and both are orders of magnitude larger than typical binding free energies. Additionally, entropy calculations via normal mode analysis are prone to overestimate the entropy loss for binding a ligand. Therefore they are sometimes skipped completely and relative binding free energies for two ligands are calculated by simply assuming that the entropic contribution to the vacuum binding free energy for both ligands is the same.

This makes the reliability of MM-PBSA calculations depend on fortuitous cancellation of errors which might differ from system to system and requires careful checking of results against experimental data. Nevertheless, the MM-PBSA method generated some highly encouraging results in early applications and has been successfully applied to protein ligand binding studies as well as other questions of macromolecular stability. The emerging consensus appears to be that MM-PBSA calculations can be useful as drug-design tools if applied skillfully, but that the method may perform poorly at times. From an extensive compilation of MM-PBSA results [148], Shirts *et al.* give mean square errors in calculated binding free enthalpies of ca. 5 kcal/mol or more, which would put the achievable level of accuracy for MM-PBSA below that of computational alchemy type free energy calculations. Nevertheless, the method is highly appealing for drug design studies, because in principle binding free energies for arbitrarily different ligands can be compared and the calculation of absolute binding free energies, while challenging, is possible.

Another popular free energy method, the linear interaction energy (LIE) approach pioneered by Åquist [149], is like MM-PBSA an endpoint method that calculates binding free energies from MD simulations of bound or free ligands only, without relying on any intermediate states like in the FEP or TI formalism. The LIE approach has been reviewed extensively in the past [150, 151] and will be presented only briefly here. The main ideas are that the binding free energy of a ligand can be computed from the difference in interaction energy with its surroundings when the bound and free state are compared. Furthermore, it is assumed that the binding free energy can be divided into an electrostatic and non-polar component. These two components are then weighted via empirical parameters, optimized with respect to experimental data:

$$\Delta G_{Bind}^{LIE} = \alpha \left(\left\langle V_{bound}^{VDW} \right\rangle - \left\langle V_{free}^{VDW} \right\rangle \right) + \beta \left(\left\langle V_{bound}^{EEL} \right\rangle - \left\langle V_{free}^{EEL} \right\rangle \right) + \gamma \quad (18)$$

where angular brackets denote ensemble averages generated by MD simulation, V represents the interaction energy of a ligand with its surroundings in the free (i.e. solvated) and receptor bound state and α, β and γ are the empirical constants. EEL and VDW stand for electrostatic and Lennard-Jones contributions to the interaction energy, as defined by a MM force field. In the first application of LIE, the three constants were set to 0.16, 0.5 and 0.0, but more recent studies have used a ligand-dependent β which led to a revised parameter set [152, 153].

The LIE approach has been used in several studies and generally leads to accurate calculations of binding free energies when compared to experimental data [154]. It is computationally efficient since no simulations of intermediate states are required and it allows straightforward calculations of not only relative but also absolute ligand binding free energies. The method as outlined in equation 18 does not consider intramolecular energy changes in the ligand and receptor. It approximates them to be linearly dependent on the interaction energies computed and relies on the empirical parameters to correctly account for these contributions. In several studies the LIE constants were reparametrised specifically for the biochemical systems studied [155-157] and if this case-by-case refitting should prove to be necessary in general, it would diminish the predictive power of the LIE method. However, Åquist *et al.* argue [58] that their revised LIE parameter set is robust and predictive for systems not used in the parameterization, which indicates validity of the method over a broad range of applications.

3. APPLICATIONS

In the following we will present several exemplary studies that used high-level 'alchemical' free energy calculations

to study protein-ligand binding. We emphasize recently published accounts, since the methodology in the field is advancing rapidly. The references in this review are confined to works of general interest in the field of free energy calculations, therefore citations particular to the studied systems described below are omitted here and can be found in the respective papers.

Binding of Small Aromatics to an Artificial T4 Lysozyme Binding Site

For free energy calculations to be used in applied protein-ligand binding studies, it is imperative to have a good estimate of their accuracy. Extensive validation studies of multiple ligands binding to a receptor are still largely confined to small test systems due to the high computational demand of the necessary calculations. We will therefore present one study on the binding of non-pharmaceutical test compounds to a model binding site before moving on to more applied examples of computational alchemy.

The L99A mutant of T4 lysozyme which contains an artificial lipophilic binding site is a popular test system for protein ligand binding phenomena both for theoreticians and experimentalists. In a recent study, Mobley *et al.* examined the binding of small organic molecules to this receptor [99], a problem that multiple other computational chemistry groups have tackled as well [158-161]. In selecting a binding site that does not contain titratable groups, ordered water molecules or cofactors and does not undergo large scale conformational changes upon binding, the authors exclude errors stemming from these issues. The accuracy of their results will therefore depend mainly on force field quality, completeness of MD sampling and the use of suitable starting conformations.

The authors first perform a retrospective study in which the known binding free energies for 11 binding and two non-binding molecules are recalculated. All ligands studied were small, rigid, aromatic molecules, e.g. benzene, indole and p-xylene. For complex starting geometries, ligand docking poses generated by the DOCK program were used. The free energy calculations included an additional restraining step for the complex which allowed absolute binding free energies to be computed. Free energy evaluations were performed using the Bennett acceptance ratio method.

The authors were able to quantify the effect of several protocol modifications on the computed ΔG_{Bind}-values. Computation using only the best-ranking docking pose as starting geometry resulted in a total rmsd value for the computed vs. experimental binding strength of 3.5 kcal/mol. This spread could be reduced to 2.6 kcal/mol when multiple starting conformations from different docking poses were used. Additionally accounting for the free energy contribution of a protein sidechain reorientation (for aar Val111, which is known to populate a rotameric state different from that in the apo-protein for several bound ligands) via a *confine-and-release* approach [162] further lowered the rmsd for the whole set of ligands to 2.2 kcal/mol. Finally, the use of an improved charge model for the ligands (AM1-BCC instead of the previously used AM1CM2) resulted in a best-case average error for the computed ΔG_{Bind} of 1.89 kcal/mol with a correlation coefficient R of 0.8. At this point, the agreement with experimental data of the free energy calculation was significantly higher than that of the binding scores from the previous docking step and allowed a clear distinction between binding and non-binding compounds and a fairly accurate ranking of the ligands according to their affinity.

Using the optimal simulation protocol identified above, a subsequent prospective study predicted the binding strengths of five new compounds (again, small rigid aromatics) identified by ligand docking calculations. The prediction of four binding and one non-binding molecule was affirmed by measurements of the changing melting temperature of the protein as well as by directly determining ΔG_{Bind} for the new compounds via isothermal titration calorimetry. Additionally, the binding modes of three of the new ligands were determined by X-ray crystallography and found to be in good agreement with predictions from the MD simulations.

This study shows that the determination of absolute binding free energies to within 2 kcal/mol accuracy is possible even for compounds where no binding mode is previously known. However, the compounds studied were small and rigid molecules binding to a simple model binding site. For the routine application of absolute binding free energy calculations to arbitrary protein ligand binding phenomena, the careful consideration of alternate binding modes or protein conformational changes as performed in this study seems critical for reliable results.

Glycogen Phosphorylase Binding of Glucose Hydantoin Derivatives

The strength of MD based free energy methods lies in their explicit consideration of potential conformational changes important for binding that are hard to account for in simpler methods. Apart from conformational changes in the protein or ligand, these can also be changes in the network of bound water molecules often surrounding a binding site. The recent study of Archontis *et al.* [163] about glucose derivatives binding to glycogen phosphorylase is one such example:

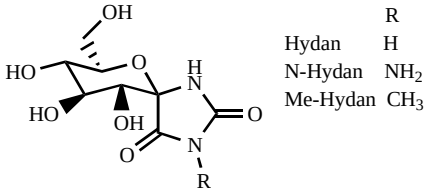


Fig. (4). Three hydantoine ligands to glycogen phosphorylase. The effect of a different N-substituent on the inhibition strength was investigated.

The enzyme glycogen phosphorylase (GP) is involved in the degradation of glycogen to supply energy to skeletal muscles. It is also, via binding glucose, involved in regulation of the glycogen metabolism where GP inhibition leads to a reduction in the blood glucose level. It is therefore considered a drug target with potential therapeutic applications for diabetes mellitus and GP inhibitors with stronger binding affinities than the relatively weakly binding glucose (with a K_I in the low mM range) would be promising lead candidates.

In a computational study combining MD simulations and thermodynamic integration, the relative binding strength of three different glucose hydantoin derivatives, hydan, 1-methyl-hydan and 1-amino-hydan to GP was calculated, see Fig. (**4**). Two ligand transformations, from hydan to 1-methyl-hydan and from hydan to 1-amino-hydan were performed. Since complex structures and thermodynamic data exist for all three ligands, an extensive comparison to experiment was possible in which good agreement between measured and theoretical results was found. The free energy calculations allowed an atomic-level analysis of energetic contributions to the relative ligand binding strengths.

The authors did not use any remedy for the 'endpoint catastrophy' associated with disappearing functional groups and instead used an analytical fitting procedure to account for the diverging $\frac{\partial V}{\partial \lambda}$-values. This yielded converged results, probably because only small ligand modifications (adding a single heavy atom) were studied. The authors used 17 λ-values, with more simulation windows at either end of the transformation. A reduced protein representation of a 20 Å sphere around the ligand was used, together with stochastic boundary conditions.

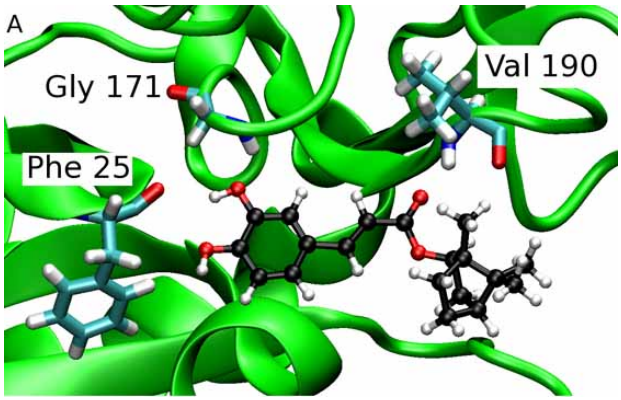
It was found that a critical water molecule changed its position in the binding site during the transformations (water X4 in the original nomenclature). The additional sterical influence from the added functional group of 1-methyl or 1-amino hydan compared to hydan forced the water molecule to move from its initial position in the GP-hydan complex. The free energy effect of moving the water molecule to its new location was estimated to be ca. 3 kcal/mol, and additional umbrella sampling simulations confirmed the size of the effect. The conformational change important in this binding process involves the rearrangement of water molecules instead of a rearrangement of protein parts. This is probably fortunate, since water molecules rapidly equilibrate and readjust their positions compared to macromolecules. Even though protein flexibility is in principle fully accounted for in free energy calculations like the one presented here, full convergence of such simulations remains a daunting challenge.

The TI calculations correctly computed the binding strengths of both 1-methyl-hydan and 1-amino-hydan to be lower than that of hydan (to within <0.5 kcal/mol and 1.3 kcal/mol of the experimental result, respectively). The reason for the loss of binding affinity, besides the necessary displacement of a water molecule was found to be sterical interaction with protein aar Asp283. The correct treatment of bound water molecules and changes in their hydrogen bonding pattern upon ligand binding found here has been highlighted by other recent studies [98] and is more and more being considered an important requirement for reliable binding free energy calculations.

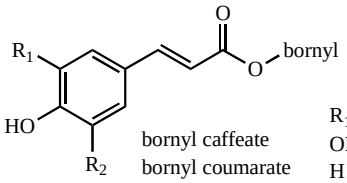
Affinity Prediction, Organic Synthesis and Validation of an Improved Neutrophile Elastase Inhibitor

In a recent study by the authors of this review, the binding of phenolic compounds to the human neutrophile elastase was investigated with ligand docking and MD based free energy methods [164]. The information obtained from studying natural inhibitors was then used to suggest and test a modified ligand, which showed improved binding properties [165]. While the computational effort and number of compounds simulated was modest compared to other studies described here, this example shows how computer simulations, organic synthesis and pharmacological activity testing can be combined to form mini drug design projects that could become routine tools in biomolecular research.

The human neutrophile elastase (HNE) is a serine protease, produced in white blood cells, with specificity for cleavage after small aliphatic aar. Besides its physiological function in the defense against pathogens and the degradation of foreign protein material, it is implicated in several inflammatory diseases. Part of its pathological activity is thought to stem from the ability to hydrolyze the elastin protein in healthy tissue and artificial HNE inhibitors would be of high interest for therapeutical applications. No high-affinity drug-like inhibitors of HNE are known, but a previous study had identified several natural caffeic acid derivatives that inhibit the enzyme with IC_{50} values in the μM range. The structure of the 218 aar functional enzyme and the location of the binding site are known from X-ray crystallography.



B



	R_1	R_2
bornyl caffeate	OH	H
bornyl coumarate	H	H
bornyl ferulate	OCH_3	H
bornyl (3,4,5-trihydroxy)-cinnamate	OH	OH

Fig. (5). The putative binding mode of bornyl caffeate to the human neutrophile elastase (A). The protein receptor is shown in green secondary structure cartoon representation with amino acids important for binding depicted in full. (B) The effect of small variations in the aromatic moiety substitution pattern on the binding affinity was studied and a new ligand with improved binding strength was designed.

The computational study of this system had two goals, first to identify putative binding modes for the new natural compound ligands and second to use calculations of their binding free energies to suggest improved inhibitors. The first was done by a combination of ligand docking and mo-

lecular dynamics calculations. First the docking program FlexX was used to place the strongest binding caffeic acid inhibitor, bornyl caffeate into the enzyme binding site. A clustering algorithm was used to group the top 30 placements into six putative ligand binding modes according to their respective rmsd values. Starting from the docked complex structures, 2 ns length MD simulations using the Amber MD suite were performed and subjected to MM-PBSA analysis. From this post processing, one of the putative binding modes was found to result in the most stable ligand binding geometry during the MD runs, plausible hydrogen bonds to the enzyme active site and the most negative computed ΔG values and this was picked as the binding mode model for all caffeic acid esters in subsequent calculations. Neither the ligand docking, nor the MM-PBSA calculations could reproduce the experimentally known binding strength of the ligands to a satisfactory degree, though. The same procedure was attempted for a different class of ligands, fucinolic acid derivatives, but no putative binding mode could be reliably determined or validated by subsequent TI calculations for these type of compounds.

Based on the suggested binding mode for caffeic acid derivatives, TI calculations were performed to calculate the relative binding free energies for three known caffeic acid derivatives, see Fig. (**5**). As the ligands differed only in a single functional group each, 4 ns of total MD simulation time were sufficient in each case to obtain converged free energy estimates that agreed to within 1 kcal/mol with the experimental differences in binding affinity. This result further indicated the validity of the proposed binding mode and encouraged the calculation of relative binding free energies for several hypothetical new HNE inhibitors. For one of these, bornyl (3,4,5-trihydroxy)-cinnamate, a 3.7 kcal/mol higher binding affinity than for the original ligand was predicted. A 4-step synthesis for enantiomerically pure bornyl (3,4,5-trihydroxy)-cinnamate from commercially available starting compounds was developed and the final product could be directly used in an HNE activity assay.

The new compound was found to have an IC_{50} of 540 *nM*, making it the strongest known cinnamic acid HNE inhibitor know up to date. Its high binding strength is consistent with the known trend of compounds rich in catecholic hydroxyl groups to be strong HNE binders. While qualitatively correct, the TI calculations somewhat overestimated the gain in binding strength. Insufficient sampling and force field errors cannot be excluded as possible causes for this deviation, but it should be noted that the new ligand also exhibited a slightly different binding kinetic than previous tested ligands, making a direct comparison of computed $\Delta\Delta G$ and measured difference in binding constants difficult. Even without fully quantitative agreement between theory and experiment, the TI calculation helped validate the suggested binding mode and proved sufficiently predictive to guide the design of a novel inhibitor.

AMP Analogues as Inhibitors of Fructose 1,6-Bisphosphatase

In their extensive study of nucleotide-like inhibitors, Reddy and Erion calculated the relative binding and solvation free energies of 22 adenosine monophosphate (AMP) derived inhibitors [166]. The inhibitors target the enzyme fructose 1,6-bisphosphatase (FB), which catalyses the hydrolysis of fructose 1,6-bisphosphate to fructose 6-phosphate in the gluconeogenesis pathway. Excessive production of glucose via gluconeogenesis is associated with diabetes mellitus disease and its critical position in this pathway makes FB an important drug target. The enzyme contains an allosteric regulation site that binds AMP, inducing a conformational change that reduces the enzyme activity. The study focused on modifying AMP in search for a more potent inhibitor and systematic chemical modifications of the phosphate moiety, sugar residue and adenine base were investigated, see Fig. (**6**).

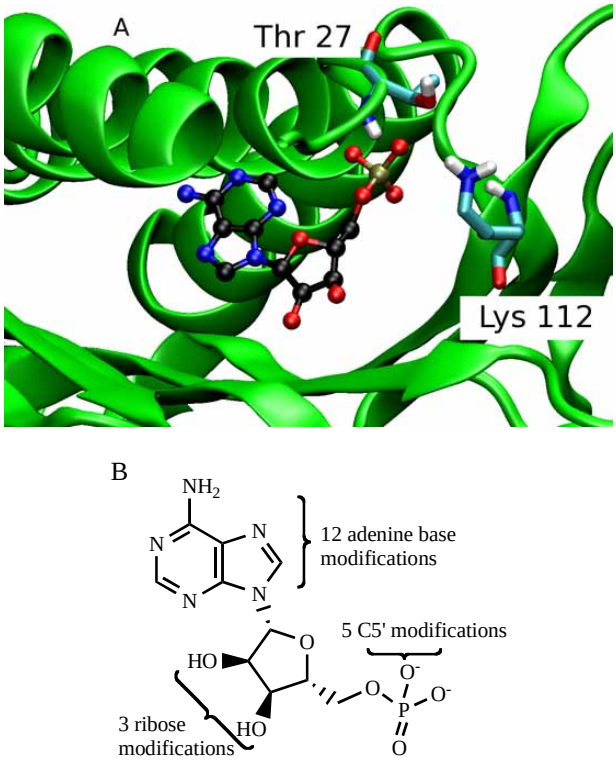


Fig. (6). (**A**) AMP bound to the allosteric binding site of fructose 1,6-bisphosphatase, with color as in Fig. (**5**). (**B**) A variety of AMP derivatives were studied, with different chemical modifications of the ribose moiety, adenine base or phosphate bound to C5'.

The authors simulated ligand transformations not only in water and in a reduced protein ligand complex representation to calculate relative *binding* free energies (see Fig. (**2**)), but also in vacuum, which gives, together with the free energy change in water, the relative *solvation* free energy of the two different ligands. An unusual TI protocol with a total of 51 and therefore quite narrow λ-windows was used, but every window was only run for a short 5 ps data collection time. To improve convergence, the results were calculated by starting runs both from complex structures of the initial and final ligand and averaging the results. Convergence may have benefited from the fact that the chemical modifications studied were mostly small single functional group changes or atom substitutions in the rigid adenine ring. For ligand transformations involving more degrees of freedom, much longer TI simulations can be expected to be necessary.

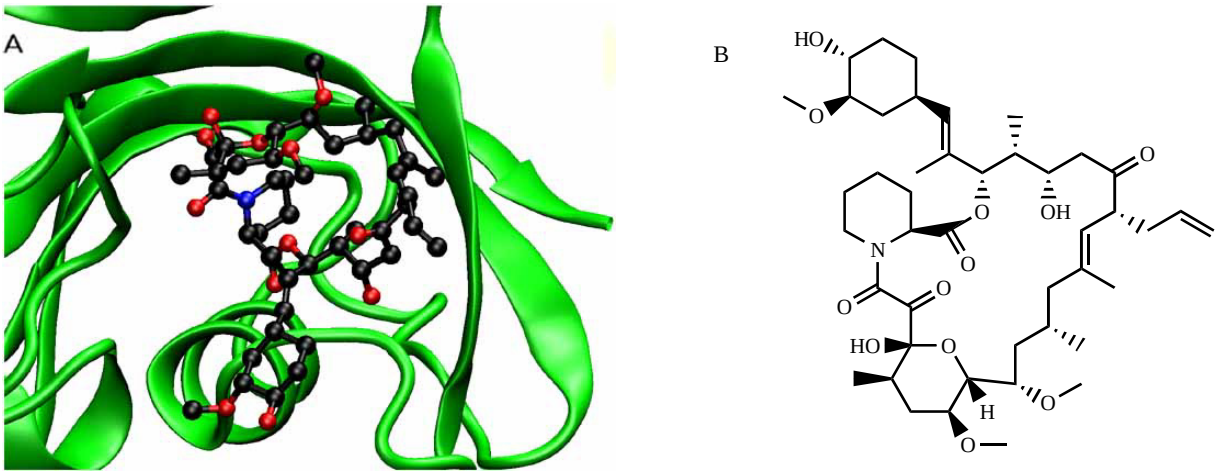


Fig. (7). The macrocyclic ligand FK506 bound to FKBP (**A**) and chemical structure of the ligand (**B**). Free energy calculations resulted in converged and accurate binding free energy estimates, even for this large and flexible molecule.

Experimentally determined binding affinities are available for several studied inhibitors and the TI calculations generally reproduced relative binding free energies computed from these in good agreement (mostly to within <1 kcal/mol). For some new inhibitors, higher binding affinities than for AMP were predicted. It is encouraging that even for charged and polar molecules like AMP accurate relative binding affinities can be calculated, but unconverged results were reported for compounds were the ligand total charge differs from that of AMP. The authors additionally tested a much faster technique to compute relative binding affinities based on molecular mechanics energy minimization in different surroundings. Results from this technique, while more or less correctly reproducing the sign of the free energy change, were significantly less predictive and far from quantitative. This study serves as a good example of the fact that MD based free energy methods with extensive sampling of conformational space are still the only reliable way to compute relative binding free energies.

The calculation of relative solvation free energies for all ligand transformations, while not necessary to estimate binding affinity, can give some additional insight into the reasons a given compound is a better or worse binder. Since ligand binding always involves a loss of beneficial ligand-water interactions in exchange for (more) beneficial ligand-receptor interactions, improving a compounds binding strength can be accomplished in two ways: By optimizing its interactions with the receptor without influencing its solvation free energy or by reducing its solvation free energy without changing its interactions with the receptor. Any real chemical modification of a ligand will involve both elements and while a standard TI binding affinity prediction cannot distinguish between them, the additional computation of the relative solvation free energy allows the separate assessment of both components.

Absolute Binding Free Energies of Diverse Pharmaceuticals to FKBP

FKBPs (short for FK-506 binding proteins) are proteins belonging to the immunophilin family. They act as peptidyl prolyl isomerases and are binding immunosuppressant drugs such as cyclosporin A or rapamycin. Their role in suppressing calcineurin-mediated T-cell activation makes them interesting drug targets. Since many known FKBP inhibitors are complex and flexible molecules (see Fig. (**7**)), this system can be considered a challenging test bed for the ability of free energy calculations in predicting inhibitors strengths.

Shirts *et al.* have studied the binding of ligands to FKBP12 both using the massively distributed computation scheme *folding@home* [167] and using a Fujitsu BioServer parallel computer [168] to compute absolute binding free energies for the same set of eight FKBP inhibitors. The difference between the two studies lay in the simulation protocol: The Fujitani *et al.* work used an improved version of the GAFF force field describing the ligands as well as a much longer equilibration phase of 10-20 ns for the FKBP-inhibitor complexes prior to the free energy calculations. This allowed the calculations of absolute binding free energies for the same set of ligands with even higher agreement to experimental data at only about 10% of the computational cost and we will focus on this study here.

The double annihilation method was used in the simulations, meaning that even at high λ-values the ligands where not restrained to the binding site using artificial additional potentials. While this method, unlike the double decoupling approach which uses restraints [97], has no rigorous connection to the standard state, a later study on the same system [169] showed that by defining a correction term based on a suitable binding site volume, standard binding free energies could be calculated. Additionally, an alternative study on the same system [102] calculated absolute binding free energies using the double-decoupling approach with ligand restraints and obtained results very similar to those of the work described here.

The Bennett acceptance ratio was used to evaluate free energy differences and data collection at 33 λ-values for 1 ns each was sufficient to obtain converged ΔG-estimates. The calculations correctly ranked the inhibitors according to their binding strength and quantitatively agreed with experimental

binding free energies except for a constant ΔG-offset of 3.2 kcal/mol. In comparison, a recent study on inhibitor design for FKBP concluded that ligand docking calculations alone could not correctly rank five inhibitors (one of which, FK506 was also considered in this study) according to their binding affinity [170].

This study shows that, if careful equilibration protocols, modern force fields and efficient free energy calculations schemes are used, the fast and quantitative calculation of absolute binding free energies for a diverse set of complex ligands to a protein receptor can be within the range of today's computational chemistry.

4. CONCLUSION

The goal of accurate, fast and reliable computations of binding free energies has been vigorously pursued since the beginnings of computational chemistry [120]. Despite many methodological advances, the traditional statistical thermodynamics based free energy methods of *computational alchemy* are still the most accurate and robust tools. They are not fast and simple enough yet to become routinely employed by anybody but a molecular modeling expert. However, the development of intermediate-level techniques like MM-PBSA and LIE, continuous improvement of existing program packages and the ever-increasing power of computers all help to advance free energy calculations further into the mainstream of pharmaceutical and medicinal chemistry.

A major point for the applicability of free energy calculations in binding free energy predictions is their achievable accuracy. In a best case scenario of sufficient sampling, a physiologically correct binding model and free energy estimation from rigorous statistical thermodynamics, errors would only stem from inaccuracies in the potential function used (e.g. the neglect of polarization effects in common fixed charge force fields or the use of simple water models). From the studies cited here, these errors appear to be in the range of roughly 1-2 kcal/mol for drug-like molecules binding to proteins. The question is if a method that is capable of predicting the binding strength effect of a chemical modification of a compound to within 2 kcal/mol (still corresponding to a 30-fold change in the inhibition constant) would be of practical use to a medicinal chemist. Given that chemical modifications in drug design very rarely improve ΔG-values by more than 1 kcal/mol [95], free energy calculations in drug design need to predict changes of a scale comparable to their statistical error.

However, in their recent work, Shirts *et al.* point out that free energy predictions even with a considerable level of inaccuracy can still be useful in practical applications [148]. In their thought experiment, chemical changes to a compound were first randomly selected. The chemical space of possible modifications was assumed to have an effect on the compounds binding strength that is Gaussian distributed and centered at zero. Then modifications were picked after being filtered through a free energy screening process which predicts their ΔG-effects with a random noise of 2 kcal/mol. The number of new compounds that have to be synthesized on average before one with a tenfold increase in binding strength is found decreased threefold in the second process. Assuming a more accurate screening step decreased the amount of necessary synthetic work even further. This example may be overly favorable towards free energy calculations since it assumes that inaccuracies in the predictions are solely caused by random instead of systematic errors and that the effort of actually performing the calculations is neglectable when compared to the work of making new compounds. Nevertheless, it appears to be a realistic account of what free energy calculations might accomplish once routinely applied to practical drug design work.

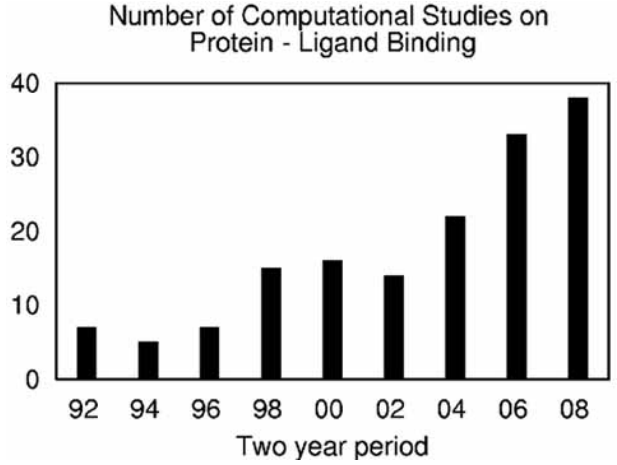


Fig. (8). An ISI Web of Knowledge search for studies on the topics "free energy calculation" and "ligand binding", grouped into two year periods.

As a note of caution, the prognosis that rigorous free energy calculations are about to emerge as mainstream tools for biochemical researchers is about as old as reviews on the subject [171, 172]. However, a survey using Thomson Reuters' ISI Web of Knowledge [173] of studies published on the topics *free energy calculation* and *ligand binding* locates 157 published accounts and sorting them by publication year reveals a recent marked increase of publishing activity, with the number of papers tripling in the last five years (see Fig. (**8**)). The authors will refrain from making any predictions, but from the recent advances presented in this work and the current high activity and interest in the field, an imminent breakthrough of free energy calculations to a wider applicability in biomolecular and drug design studies seems not only possible but more likely than ever.

5. ACKNOWLEDGEMENTS

The authors would like to thanks Prof. D.A. Case for invaluable discussions and Prof. T. Koslowski for a critical read of the manuscript and many helpful suggestions. Financial support from NIH grant RR12255 is gratefully acknowledged. TS acknowledges funding from a DFG postdoctoral fellowship.

ABBREVIATIONS

ADME	=	absorption, distribution, metabolism and excretions
HTS	=	high throughput screening
QSAR	=	quantitative structure activity relationship
MD	=	molecular dynamics

MM	=	molecular mechanics
FEP	=	free energy perturbation
TI	=	thermodynamic integration
AAP	=	acetaminophen
VDW	=	van-der-Waals
BAR	=	Bennett acceptance ratio
PMF	=	potential of mean force
WHAM	=	weighted histogram analysis method
MM-PB(GB)SA	=	molecular mechanics - Poisson Boltzman (Generalized Born) surface area method
LIE	=	linear interaction energy
GP	=	glycogen phosphorylase
HNE	=	human neutrophile elastase
AMP	=	adenosine monophosphate
FB	=	fructose 1,6-bisphosphatase
FKBP	=	FK-506 binding protein
GAFF	=	general Amber force field
ISI	=	Institute for Scientific Information, now Thomson Scientific

REFERENCES

[1] Voet, D.; Voet, J.G. *Biochemistry*. J. Wiley and Sons, New York, **2004**.

[2] Dominy, B.N. Molecular recognition and binding free energy calculations in drug development. *Curr. Pharm. Biotech.*, **2008**, *9*, 87-95.

[3] Ideker, T.; Thorsson, V.; Ranish, J.A.; Christmas, R.; Buhler, J.; Eng, J.K.; Bumgarner, R.; Goodlett, D.R.; Aebersold, R.; Hood, L. Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science*, **2001**, *292*, 929-934.

[4] Strogatz, S.H. Exploring complex networks. *Nature*, **2001**, *410*, 268-276.

[5] Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Del. Rev.*, **1997**, *23*, 3-25.

[6] Hopkins, A.L.; Groom, C.R. The druggable genome. *Nat. Rev. Drug Disc.*, **2002**, *1*, 727-730.

[7] Chaires, J.B. Energetics of drug-DNA interactions. *Biopolymers*, **1997**, *44*, 201-215.

[8] Propst, C.L.; Perun, T.J., Eds. *Nucleic Acid Targeted Drug Design*. Marcel Dekker, Inc., New York, **1992.**

[9] Presta, L.G. Molecular engineering and design of therapeutic antibodies. *Curr. Opin. Immunol.*, **2008**, *20*, 460-470.

[10] Lipinski, C.A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharm. Tox. Met.*, **2000**, *44*, 235-249.

[11] Balakin, K.V.; Savchuk, N.P.; Tetko, I.V. *In silico* approaches to prediction of aqueous and DMSO solubility of drug-like compounds: trends, problems and solutions. *Curr. Med. Chem.*, **2006**, *13*, 223-241.

[12] Bergstrom, C. Computational models to predict aqueous drug solubility, permeability and intestinal absorption. *Expert Opin. Drug Metab. Toxicol.*, **2005**, *1*, 613-27.

[13] Cheng, Y.; Prusoff, W.H. Relationship between inhibition constant (K_i) and the concentration of inhibitor which causes 50 per cent inhibition (I_{50}) of an enzymatic reaction. *Biochem. Pharmacol.*, **1973**, *22*, 3099-3108.

[14] Drews, J.; Ryser, S. The role of innovation in drug development. *Nat. Biotechnol.*, **1997**, *15*, 1318-1319.

[15] Imming, P.; Sinning, C.; Meyer, A. Opinion - Drugs, their targets and the nature and number of drug targets. *Nat. Rev. Drug Disc.*, **2006**, *5*, 821-834.

[16] Spencer, R.W. High-throughput screening of historic collections: Observations on file size, biological targets, and file diversity. *Biotech. Bioeng.*, **1998**, *61*, 61-67.

[17] Jacoby, E.; Schuffenhauer, A.; Popov, M.; Azzaoui, K.; Havill, B.; Schopfer, U.; Engeloch, C.; Stanek, J.; Acklin, P.; Rigollier, P.; Stoll, F.; Koch, G.; Meier, P.; Orain, D.; Giger, R.; Hinrichs, J.; Malagu, K.; Zimmermann, J.; Roth, H.J. Key aspects of the Novartis compound collection enhancement project for the compilation of a comprehensive chemogenomics drug discovery screening collection. *Curr. Top. Med. Chem.*, **2005**, *5*, 397-411.

[18] Diller, D.J. The synergy between combinatorial chemistry and high-throughput screening. *Curr. Opin. Drug Disc. Devel.*, **2008**, *11*, 346-355.

[19] Manly, C.J.; Chandrasekhar, J.; Ochterski, J.W.; Hammer, J.D.; Warfield, B.B. Strategies and tactics for optimizing the Hit-to-Lead process and beyond - A computational chemistry perspective. *Drug Disc. Today*, **2008**, *13*, 99-109.

[20] Chaires, J.B. Calorimetry and thermodynamics in drug design. *Ann. Rev. Biophys.*, **2008**, *37*, 135-151.

[21] Keseru, G.M.; Makara, G.M. Hit discovery and hit-to-lead approaches. *Drug Disc. Today*, **2006**, *11*, 741-748.

[22] Lundquist, T. The devil is still in the details - driving early drug discovery forward with biophysical experimental methods. *Curr. Opin. Drug Disc. Devel.*, **2005**, *8*, 513-519.

[23] A. Ababou, J.E. Ladbury. Survey of the year 2004: literature on applications of isothermal titration calorimetry. *J. Mol. Recog.*, **2006**, *19*, 79-89.

[24] Bruylants, G.; Wouters, J.; Michaux, C. Differential scanning calorimetry in life science: Thermodynamics, stability, molecular recognition and application in drug design. *Curr. Med. Chem.*, **2005**, *12*, 2011-2020.

[25] Guido, R.V.C.; Oliva, G.; Andricopulo, A.D. Virtual screening and its integration with modern drug design technologies. *Curr. Med. Chem.*, **2008**, *15*, 37-46.

[26] Jorgensen, W.L. The many roles of computation in drug discovery. *Science*, **2004**, *303*, 1813-1818.

[27] Gilson, M.K.; Zhou, H.X. Calculation of protein-ligand binding affinities. *Ann. Rev. Biophys. Biomol. Struct.*, **2007**, *36*, 21-42.

[28] de Azevedo, W.F.; Dias, R. Computational methods for calculation of ligand-binding affinity. *Curr. Drug Target*, **2009**, *9*, 1031-1039.

[29] Jorgensen, W.L. Efficient drug lead discovery and optimization. *Acc. Chem. Res.*, **2009**, *42*, 724-733.

[30] Michel, J.; Essex, J.W. Hit identification and binding mode predictions by rigorous free energy simulations. *J. Med. Chem.*, **2008**, *51*, 6654-6664.

[31] Mobley, D.L.; Dill, K.A. Binding of small-molecule ligands to proteins: "What you see" is not always "what you get". *Structure*, **2009**, *17*, 489-498.

[32] Greer, J.; Erickson, J.W.; Baldwin, J.J.; Varney, M.D. Application of the 3-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.*, **1994**, *37*, 1035-1054.

[33] von Itzstein, M.; Wu, W.Y.; Kok, G.B.; Pegg, M.S.; Dyason, J.C.; Jin, B.; Phan, T.V.; Smythe, M.L.; White, H.F.; Oliver, S.W.; Colman, P.M.; Varghese, J.N.; Ryan, D.M.; Woods, J.M.; Bethell, R.C.; Hotham, V.J.; Cameron, J.M.; Penn, C.R. Rational design of potent sialidase-based inhibitors of influenza-virus replication. *Nature*, **1993**, *363*, 418-423.

[34] Hansch, C.; Fujita, T. rho-sigma-pi analysis. A method for the correlation of biological activity and chemical structure. *J. Am. Chem. Soc.*, **1964**, *86*, 1616-1626.

[35] Leo, A.; Hansch, C.; Church, C. Comparison of parameters currently used in study of structure-activity relationships. *J. Med. Chem.*, **1969**, *12*, 766-771.

[36] Labute, P. A widely applicable set of descriptors. *J. Mol. Graphics Model.*, **2000**, *18*, 464-477.

[37] Briggs, J.M.; Marrone, T.J.; McCammon J.A. Computational science new horizons and relevance to pharmaceutical design. *Trends Card. Med.*, **1996**, *6*, 198-204.

[38] Yang, G.F.; Huang, X.Q. Development of quantitative structure-activity relationships and its application in rational drug design. *Curr. Pharm. Des.*, **2006**, *12*, 4601-4611.

[39] Brooijmans, N.; Kuntz, I.D. Molecular recognition and docking algorithms. *Ann. Rev. Biophys. Biomol. Struct.*, **2003**, *32*, 335-373.

[40] Ewing, T.J.A.; Kuntz, I.D. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.*, **1997**, *18*, 1175-1189.

[41] Moustakas, D.T.; Lang, P.T.; Pegg, S.; Pettersen, E.; Kuntz, I.D.; Brooijmans, N.; Rizzo, R.C. Development and validation of a modular, extensible docking program: Dock 5. *J. Comp. Aided Mol. Des.*, **2006**, *20*, 601-619.

[42] Morris, G.M.; Goodsell, D.S.; Halliday, R.S.; Huey, R.; Hart, W.E.; Belew, R.K.; Olson, A.J. Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *J. Comput. Chem.*, **1998**, *19*, 1639-1662.

[43] Kramer, B.; Rarey, M.; Lengauer, T. Evaluation of the FLEXX incremental construction algorithm for protein-ligand docking. *Proteins - Struct. Func. Gen.*, **1999**, *37*, 228-241.

[44] Rarey, M.; Kramer, B.; Lengauer, T. A fast flexible docking method using an incremental construction algorithm. *J. Mol. Biol.*, **1996**, *261*, 470-489.

[45] Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.*, **1997**, *267*, 727-748.

[46] Friesner, R.A.; Banks, J.L.; Murphy, R.B.; Halgren, T.A.; Klicic, J.J.; Mainz, D.T.; Repasky, M.P.; Knoll, E.H.; Shelley, M.; Perry, J.K.; Shaw, D.E.; Francis, P.; Shenkin, P.S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.*, **2004**, *47*, 1739-1749.

[47] Halgren, T.A.; Murphy, R.B.; Friesner, R.A.; Beard, H.S.; Frye, L.L.; Pollard, W.T.; Banks, J.L. Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J. Med. Chem.*, **2004**, *47*, 1750-1759.

[48] Carlson, H.A. Protein flexibility is an important component of structure-based drug discovery. *Curr. Pharm. Des.*, **2002**, *8*, 1571-1578.

[49] Cavasotto, C.N.; Singh, N. Docking and high throughput docking: Successes and the challenge of protein flexibility. *Curr. Comput. Aided Drug Des.*, **2008**, *4*, 221-234.

[50] Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FlexE: Effient molecular docking considering protein structure variations. *J. Mol. Biol.*, **2001**, *308*, 377-395.

[51] Cozzini, P.; Kellogg, G.E.; Spyrakis, F.; Abraham, D.J.; Costantino, G.; Emerson, A.; Fanelli, F.; Gohlke, H.; Kuhn, L.A.; Morris, G.M.; Orozco, M.; Pertinhez, T.A.; Rizzi, M.; Sotriffer, C.A. Target Flexibility: An emerging consideration in drug discovery and design. *J. Med. Chem.*, **2008**, *51*, 6237-6255.

[52] Teodoro, M.L.; Kavraki, L.E. Conformational flexibility models for the receptor in structure based drug design. *Curr. Pharm. Des.* , **2003**, *9*, 1635-1648.

[53] Kollman, P. Free energy calculations: Applications to chemical and biochemical phenomena. *Chem. Rev.*, **1993**, *93*, 2395-2417.

[54] Reddy, M.R.; Erion, M.D.; Agarwal, A. Free energy calculations: Use and limitations in predicting ligand binding affinities. *Rev. Comput. Chem.*, **2000**, *16*, 217-304.

[55] Wang, W.; Donini, O.; Reyes, C.M.; Kollman, P.A. Biomolecular simulations: Recent developments in force fields, simulations of enzyme catalysis, protein-ligand, protein-protein, and protein-nucleic acid noncovalent interactions. *Ann. Rev. Biophy. Biomol. Struct.*, **2001**, *30*, 211-243.

[56] Lazaridis, T. Binding affinity and specificity from computational studies. *Curr. Org. Chem.*, **2002**, *6*, 1319-1332.

[57] Simonson, T.; Archontis, G.; Karplus, M. Free energy simulations come of age: Protein-ligand recognition. *Acc. Chem. Res.*, **2002**, *35*, 430-437.

[58] Brandsdal, B.O.; Osterberg, F.; Almlof, M.; Feierberg, I.; Luzhkov, V.B.; Åqvist, J. Free energy calculations and ligand binding. *Protein Simulat.* **2003**, *66* 123-158.

[59] Foloppe, N.; Hubbard, R. Towards predictive ligand design with free-energy based computational methods? *Curr. Med. Chem.*, **2006**, *13*, 3583-3608.

[60] Dolenc, J.; Oostenbrink, C.; Koller, J.; van Gunsteren, W.F. Molecular dynamics simulations and free energy calculations of netropsin and distamycin binding to an AAAAA DNA binding site. *Nucleic Acids Res.*, **2005**, *33*, 725-733.

[61] Kopitz, H.; Zivkovic, A.; Engels, J.W.; Gohlke, H. Determinants of the unexpected stability of RNA fluorobenzene self pairs. *ChemBioChem*, **2008**, *9*, 2619-2622.

[62] Ytreberg, F.M.; Zuckerman, D.M. A black-box re-weighting analysis can correct flawed simulation data. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 7982-7987.

[63] Pissurlenkar, R.S.; Mushtaque, S.; Iyer, R.P.; Evans C. Molecular mechanics force fields and their applications in drug design. *Anti-Infect. Agents Med. Chem.*, **2009**, *8*, 128-150.

[64] Allen, M.P.; Tildesley, D.J. *Computer Simulations of Liquids*. Clarendon Press, Oxford, **1987**.

[65] Case, D.A.; Darden, T.A.; Cheatham III, T.E.; Simmerling, C.L.; Wang, J.; Duke, R.E.; Luo, R.; Crowley, M.; Walker, R.C.; Zhang, W.; Merz, K.M.; Wang, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Kolossvary, I.; Wong, K.F.; Paesani, F.; Vanicek, J.; Wu, X.; Brozell, S.R.; Steinbrecher, T.; Gohlke, H.; Yang, L.; Tan, C.; Mongan, J.; Hornak, V.; Cui, G.; Mathews, D.H.; Seetin, M.G.; Sagui, C.; Babin, V.; Kollman, P.A. *AMBER10*. University of California, San Francisco, **2008**.

[66] Case, D.A.; Cheatham III, T.E.; Darden, T.A.; Gohlke, H.; Luo, R.; Merz, K.M.; Onufriev, A.; Simmerling, C.L.; Wang, B.; Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.

[67] Pearlman, D.A.; Case, D.A.; Caldwell, J.W.; Ross, W.S.; Cheatham III, T.E.; DeBolt, S.; Ferguson, D.; Seibel, G.; Kollman, P. AMBER, a package of computer programs for applying molecular mechanics, normal mode analysis, molecular dynamics and free energy calculations to simulate the structural and energetic properties of molecules. *Comput. Phys. Commun.*, **1995**, *91*, 1-41.

[68] Cornell, W.D.; Cieplak, P.; Bayly, C.I.; Gould, I.R.; Merz, K.M.; Ferguson, D.M.; Spellmeyer, D.C.; Fox, T.; Caldwell, J.W.; Kollman, P.A. A 2nd generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, **1995**, *117*, 5179-5197.

[69] Duan, Y.; Wu, C.; Chowdhury, S.; Lee, M.C.; Xiong, G.M.; Zhang, W.; Yang, R.; Cieplak, P.; Luo, R.; Lee, T.; Caldwell, J.W.; Wang, J.M.; Kollman, P. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *J. Comput. Chem.*, **2003**, *24*, 1999-2012.

[70] Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins*, **2006**, *65*, 712-725.

[71] Weiner, P.K.; Kollman, P.A. Amber - Assisted model building with energy refinement - A general program for modeling molecules and their interactions. *J. Comput. Chem.*, **1981**, *2*, 287-303.

[72] Brooks, B.R.; Bruccoleri, R.E.; Olafson, B.D.; States, D.J.; Swaminathan, S.; Karplus, M. Charmm: A program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.*, **1983**, *4*, 187-217.

[73] Scott, W.R.P.; Hunenberger, P.H.; Tironi, I.G.; Mark, A.E.; Billeter, S.R.; Fennen, J.; Torda, A.E.; Huber, T.; Kruger, P.; van Gunsteren, W.F. The GROMOS biomolecular simulation program package. *J. Phys. Chem. A*, **1999**, *103*, 3596-3607.

[74] Jorgensen, W.L.; Maxwell, D.S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, **1996**, *118*, 11225-11236.

[75] Ren, P.Y.; Ponder, J.W. Consistent treatment of inter- and intramolecular polarization in molecular mechanics calculations. *J. Comput. Chem.*, **2002**, *23*, 1497-1506.

[76] Ren, P.Y.; Ponder, J.W. Temperature and pressure dependence of the AMOEBA water model. *J. Phys. Chem. B*, **2004**, *108*, 13427-13437.

[77] Khoruzhii, O.; Donchev, A.G.; Galkin, N.; Illarionov, A.; Olevanov, M.; Ozrin, V.; Quenn, C.; Tarasov, V. Application of a polarizable force field to calculations of relative protein-ligand binding affinities. *Proc. Natl. Acad. Sci. USA*, **2008**, *105*, 10378-10383.

[78] Still, W.C.; Tempczyk, A.; Hawley, R.C.; Hendrickson, T. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.*, **1990**, *112*, 6127-6129.

[79] Schaefer, M.; Karplus, M. A comprehensive analytical treatment of continuum electrostatics. *J. Phys. Chem.*, **1996**, *100*, 1578-1599.

[80] Edinger, S.R.; Cortis, C.; Shenkin, P.S.; Friesner, R.A. Solvation free energies of peptides: Comparison of approximate continuum solvation models with accurate solution of the Poisson-Boltzmann equation. *J. Phys. Chem. B*, **1997**, *101*, 1190-1197.

[81] Jayaram, B.; Sprous, D.; Beveridge, D.L. Solvation free energy of biomacromolecules: Parameters for a modified generalized born

model consistent with the AMBER force field. *J. Phys. Chem. B*, **1998**, *102*, 9571-9576.
[82] Cramer, C.J.; Truhlar, D.G. Implicit solvation models: Equilibria, structure, spectra, and dynamics. *Chem. Rev.*, **1999**, *99*, 2161-2200.
[83] Bashford, D.; Case, D.A. Generalized born models of macromolecular solvation effects. *Annu. Rev. Phys. Chem.*, **2000**, *51*, 129-152.
[84] Onufriev, A.; Bashford, D.; Case, D.A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **2000**, *104*, 3712-3720.
[85] Lee, M.S.; Salsbury, F.R.; Brooks, C.L.. Novel generalized Born methods. *J. Chem. Phys.*, **2002**, *116*, 10606-10614.
[86] Davis, M.E.; McCammon, J.A. Electrostatics in biomolecular structure and dynamics. *Chem. Rev.*, **1990**, *90*, 509-521.
[87] Sharp, K.A.; Honig, B. Electrostatic interactions in macromolecules: Theory and applications. *Annu. Rev. Biophys. Biophys. Chem.*, **1990**, *19*, 301-332.
[88] Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: An N x log(N) method for Ewald sums in large systems. *J. Chem. Phys.*, **1993**, *98*, 10089-10092.
[89] Essmann, U.; Perera, L.; Berkowitz, M.L.; Darden, T.; Lee, H.; Pedersen, L.G. A smooth particle mesh Ewald method. *J. Chem. Phys.*, **1995**, *103*, 8577-8593.
[90] Verlet, L. Computer "experiments" on classical fluids. I. thermodynamical properties of Lennard-Jones molecules. *Phys. Rev.*, **1967**, *159*, 98-103.
[91] Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H.J.C. Numerical integration of the cartesian equations of motion of a system with constraints: Molecular dynamics of n-alkanes. *J. Comput. Phys.*, **1977**, *23*, 327-341.
[92] Pitera, J.W. Current developments in and importance of high-performance computing in drug discovery. *Curr. Opin. Drug Discovery Dev.*, **2009**, *12*, 388-396.
[93] Maren, T.H. Direct measurements of the rate constants of sulfonamides with carbonic anhydrase. *Mol. Pharm.*, **1992**, *41*, 419-426.
[94] Schloss, J.V. Significance of slow-binding enzyme inhibition and its relationship to reaction-intermediate analogs. *Acc. Chem. Res.*, **1988**, *21*, 348-353.
[95] Hajduk, P.J.; Sauer, D.R. Statistical analysis of the effects of common chemical substituents on ligand potency. *J. Med. Chem.*, **2008**, *51*, 553-564.
[96] Boresch, S.; Tettinger, F.; Leitgeb, M.; Karplus, M. Absolute binding free energies: A quantitative approach for their calculation. *J. Phys. Chem. B*, **2003**, *107*, 9535-9551.
[97] Gilson, M.K.; Given, J.A.; Bush, B.L.; McCammon, J.A. The statistical thermodynamic basis for computation of binding affinities: A critical review. *Biophys. J.*, **1997**, *72*, 1047-1069.
[98] Hamelberg, D.; McCammon, J.A. Standard free energy of releasing a localized water molecule from the binding pockets of proteins: Double-decoupling method. *J. Am. Chem. Soc.*, **2004**, *126*, 7683-7689.
[99] Mobley, D.L.; Graves, A.P.; Chodera, J.D.; McReynolds, A.C.; Shoichet, B.K.; Dill, K.A. Predicting absolute ligand binding free energies to a simple model site. *J. Mol. Biol.*, **2007**, *371*, 1118-1134.
[100] Woo, H.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. *Proc. Natl. Acad. Sci. USA*, **2005**, *102*, 6825-6830.
[101] Olano, L.R.; Rick, S.W. Hydration free energies and entropies for water in protein interiors. *J. Am. Chem. Soc.*, **2004**, *126*, 7991-8000.
[102] Wang, J.; Deng, Y.; Roux, B. Absolute binding free energy calculations using molecular dynamics simulations with restraining potentials. *Biophys. J.*, **2006**, *91*, 2798-2814.
[103] Zwanzig, R.W. High-temperature equation of state by a perturbation method. I. nonpolar gases. *J. Chem. Phys.*, **1954**, *22*, 1420-1426.
[104] Kirkwood, J.G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, **1935**, *3*, 300-313.
[105] Shyu, C.; Ytreberg, F.M. Reducing the bias and uncertainty of free energy estimates by using regression to fit thermodynamic integration data. *J. Comput. Chem.*, **2009**, *30*, 2297-2304.
[106] Knight, J.L.; Brooks III, C.L. λ-Dynamics free energy simulation methods. *J. Comput. Chem.*, **2009**, *30*, 1692-1700.
[107] Blondel, A. Ensemble variance in free energy calculations by thermodynamic integration: Theory, optimal "Alchemical" path, and practical solutions. *J. Comput. Chem.*, **2004**, *25*, 985-993.
[108] Valleau, J.P.; Torrie, G.M. *Modern Theoretical Chemistry, Vol. 5: Statistical Mechanics, Part A.* Berne, B.J., Ed. Plenum Press, New York, **1977**.
[109] Beveridge, D.L.; DiCapua, F.M. Free energy simulation via molecular simulations: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem.*, **1989**, *18*, 431-492.
[110] Simonson, T. Free-energy of particle insertion - an exact analysis of the origin singularity for simple liquids. *Mol. Phys.*, **1993**, *80*, 441-447.
[111] Pearlman, D.A.; Kollman, P.A. A new method for carrying out free energy perturbation calculations: dynamically modified windows. *J. Chem. Phys.*, **1989**, *90*, 2460-2470.
[112] Straatsma, T.P.; Zacharias, M.; McCammon, J.A. Holonomic constraint contributions to free-energy differences from thermodynamic integration molecular-dynamics simulations. *Chem. Phys. Lett.*, **1992**, *196*, 297-302.
[113] Pitera, J.W.; van Gunsteren, W.F. A comparison of non-bonded scaling approaches for free energy calculations. *Mol. Simulat.*, **2002**, *28*, 45-65.
[114] Steinbrecher, T.; Mobley, D.L.; Case, D.A. Non-linear scaling schemes for Lennard-Jones interactions in free energy calculations. *J. Chem. Phys.*, **2007**, *127*, 214108.
[115] Bennett, C.H. Efficient estimation of free energy differences from Monte Carlo data. *J. Comput. Phys.*, **1976**, *22*, 245-268.
[116] Metropolis, N.; Rosenbluth, A.W.; Rosenbluth, M.N.; Teller, A.H.; Teller, E. Equation of state calculations by fast computing machines. *J. Chem. Phys.*, **1953**, *21*, 1087-1092.
[117] Crooks, G.E. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E*, **2000**, *61*, 2361-2366.
[118] Shirts, M.R.; Bair, E.; Hooker, G.; Pande, V.S. Equilibrium free energies from nonequilibrium measurements using maximum-likelihood methods. *Phys. Rev. Lett.*, **2003**, *91*, 140601.
[119] Shirts, M.R.; Chodera, J.D. Statistically optimal analysis of samples from multiple equilibrium states. *J. Chem. Phys.*, **2008**, *129*, 124105.
[120] Shirts, M.R.; Mobley, D.L.; Chodera, J.D. Alchemical free energy calculations: Ready for prime time? *Ann. Rep. Comput. Chem.*, **2007**, *3*, 41-59.
[121] Wang, J.; Wolf, R.M.; Caldwell, J.W.; Kollman, P.A.; Case, D.A. Development and testing of a general amber force field. *J. Comput. Chem.*, **2004**, *25*, 1157-1174.
[122] Bayly, C.I.; Cieplak, P.; Cornell, W.D.; Kollman, P.A. A well-behaved electrostatic potential based method using charge restraints for deriving atomic charges: The RESP model. *J. Phys. Chem.*, **1993**, *97*, 10269-10280.
[123] Jorgensen, W.L.; Chandrasekhar, J.; Madura, J.P.; Impey, R.W.; Klein, M.L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, **1983**, *79*, 926-935.
[124] Granberg, R.A.; Rasmuson, A.C. Solubility of paracetamol in pure solvents. *J. Chem. Eng. Data*, **1999**, *44*, 1391-1395.
[125] Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Pairwise solute descreening of solute charges from a dielectric medium. *Chem. Phys. Lett.*, **1995**, *246*, 122-129.
[126] Hawkins, G.D.; Cramer, C.J.; Truhlar, D.G. Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.*, **1996**, *100*, 19824-19839.
[127] Roux, B. The calculation of the potential of mean force using computer simulations. *Comput. Phys. Comm.*, **1995**, *91*, 275-282.
[128] Kumar, S.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A.; Rosenberg, J.M. The weighted histogram analysis method for free-energy calculations on biomolecules. 1. The method. *J. Comput. Chem.*, **1992**, *13*, 1011-1021.
[129] Kumar, S.; Rosenberg, J.M.; Bouzida, D.; Swendsen, R.H.; Kollman, P.A. Multidimensional free-energy calculations using the weighted histogram analysis method. *J. Comput. Chem.*, **1995**, *16*, 1339-1350.
[130] Jarzynski, C. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.*, **1997**, *78*, 2690-2693.
[131] Oostenbrink, C.; van Gunsteren, W.F. Calculating zeros: Non-equilibrium free energy calculations. *Chem. Phys.*, **2006**, *323*, 102-108.

[132] Ko, Y.J.; Huh, J.; Jo, W.H. Ion exclusion mechanism in aquaporin at an atomistic level. *Proteins - Struct. Func. Bioinf.*, **2008**, *70*, 1442-1450.
[133] Liu, M.; Sun, T.; Hu, J.; Chen, W.; Wang, C. Study on the mechanism of the BtuF periplasmic-binding protein for vitamin B12. *Biophys. Chem.*, **2008**, *135*, 19-24.
[134] Cossins, B.P.; Foucher, S.; Edge, C.A.; Essex, J.W. Protein-ligand binding affinity by nonequilibrium free energy methods. *J. Phys. Chem. B*, **2008**, *112*, 14985-14992.
[135] Narambuena, C.F.; Beltramo, D.M.; Leiva, E.P.M. Polyelectrolyte adsorption on a charged surface. Free energy calculation from Monte Carlo simulations using Jarzynski equality. *Macromolecules*, **2008**, *41*, 8267-8274.
[136] Cossins, B.P.; Foucher, S.; Edge, C.M.; Essex, J.W. Assessment of nonequilibrium free energy methods. *J. Phys. Chem. B*, 2009, 113, 5508-5519.
[137] Kollman, P.A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.H.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D.A.; Cheatham III, T.E. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*, 889-897.
[138] Srinivasan, J.; Cheatham III, T.E.; Cieplak, P.; Kollman, P.A.; Case, D.A. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate-DNA helices. *J. Am. Chem. Soc.*, **1998**, *120*, 9401-9409.
[139] Wang, W.; Kollman, P.A. Free energy calculations on dimer stability of the HIV protease using molecular dynamics and a continuum solvent model. *J. Mol. Biol.*, **2000**, *303*, 567-582.
[140] Reyes, C.M.; Kollman, P.A. Structure and thermodynamics of RNA-protein binding: Using molecular dynamics and free energy analyses to calculate the free energies of binding and conformational change. *J. Mol. Biol.*, **2000**, *297*, 1145-1158.
[141] Lee, M.R.; Duan, Y.; Kollman, P.A. Use of MM-PB/SA in estimating the free energies of proteins: Application to native, intermediates, and unfolded vilin headpiece. *Proteins*, **2000**, *39*, 309-316.
[142] Wang, J.M.; Morin, P.; Wang, W.; Kollman, P.A. Use of MM-PBSA in reproducing the binding free energies to HIV-1 RT of TIBO derivatives and predicting the binding mode to HIV-1 RT of efavirenz by docking and MM-PBSA. *J. Am. Chem. Soc.*, **2001**, *123*, 5221-5230.
[143] Marinelli, L.; Cosconati, S.; Steinbrecher, T.; Limongelli, V.; Bertamino, A.; Novellino, E.; Case, D.A. Homology modeling of NR2B modulatory domain of NMDA receptor and analysis of ifenprodil binding. *ChemMedChem*, **2007**, *2*, 1498-1510.
[144] Connolly, M.L. Analytical molecular surface calculation. *J. Appl. Cryst.*, **1983**, *16*, 548-558.
[145] Morreale, A.; de la Cruz, X.; Meyer, T.; Gelpi, J.L.; Luque, F.J.; Orozco, M. Linear response theory: An alternative to PB and GB methods for the analysis of molecular dynamics trajectories? *Proteins - Struct. Func. Bioinf.*, **2004**, *57*, 458-467.
[146] Gohlke, H.; Case, D.A. Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf. *J. Comput. Chem.*, **2003**, *25*, 238-250.
[147] Amzel, L.M. Loss of translational entropy in binding, folding, and catalysis. *Proteins: Struct. Func. Gen.*, **1997**, *28*, 144-149.
[148] Shirts, M.R.; Mobley, D.L.; Brown, S.P. *Free energy calculations in structure-based drug design*. in *Structure Based Drug Design*. Cambridge University Press, in press.
[149] Åqvist, J. Comment on transferability of ion models. *J. Phys. Chem.*, **1994**, *98*, 8253-8255.
[150] Åqvist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen.*, **2001**, *4*, 613-626.
[151] Åqvist, J.; Luzhkov, V.B.; Brandsdal, B.O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.*, **2002**, *35*, 358-365.
[152] Åqvist, J.; Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem.*, **1996**, *100*, 9512-9521.
[153] Hansson, T.; Marelius, J.; Åqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comp. Aided Mol. Des.*, **1998**, *12*, 27-35.
[154] Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Åqvist, J. Computation of affinity and selectivity: Binding of 2,4-diaminopteridine and 2,4-diaminoquinazoline inhibitors to dihydrofolate reductases. *J. Comput. Aided Mol. Des.*, **1998**, *12*, 119-131.
[155] Paulsen, M.D.; Ornstein, R.L. Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.*, **1996**, *9*, 567-571.
[156] Lamb, M.L.; Tirado-Rives, J.; Jorgensen, W.L. Estimation of the binding affinities of FKBP12 inhibitors using a linear response method. *Bioorg. Med. Chem.*, **1999**, *7*, 851-860.
[157] Wall, I.D.; Leach, A.R.; Salt, D.W.; Ford, M.G.; Essex, J.W. Binding constants of neuraminidase inhibitors: An investigation of the linear interaction energy method. *J. Med. Chem.*, **1999**, *42*, 5142-5152.
[158] Wei, B.Q.; Baase, W.A.; Weaver, L.H.; Matthews, B.W.; Shoichet, B.K. A model binding site for testing scoring functions in molecular docking. *J. Mol. Biol.*, **2002**, *322*, 339-355.
[159] Hermans, J.; Wang, L. Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. Application to a complex of benzene and mutant T4 lysozyme. *J. Am. Chem. Soc.*, **1997**, *119*, 2707-2714.
[160] Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem. B*, **2009**, *113*, 2234-2246.
[161] Deng, Y.; Roux, B. Calculation of standard binding free energies: Aromatic molecules in the T4 lysozyme L99A mutant. *J. Chem. Theor. Comput.*, **2006**, *2*, 1255-1273.
[162] Mobley, D.L.; Chodera, J.D.; Dill, K.A. Confine-and-release method: Obtaining correct binding free energies in the presence of protein conformational change. *J. Chem. Theor. Comput.*, **2007**, *3*, 1231-1235.
[163] Archontis, G.; Watson, K.A.; Xie, Q.; Andreou, G.; Chrysina, E.D.; Zographos, S.E.; Oikonomakos, N.G.; Karplus, M. Glycogen phosphorylase inhibitors: A free energy perturbation analysis of glucopyranose spirohydantoin analogues. *Proteins - Struct. Func. Bioinf.*, **2005**, *61*, 984-998.
[164] Steinbrecher, T.; Case, D.A.; Labahn, A. A multistep approach to structure-based drug design: Studying ligand binding at the human neutrophil elastase. *J. Med. Chem.*, **2006**, *49*, 1837-1844.
[165] Steinbrecher, T.; Hrenn, A.; Dormann, K.; Merfort, I.; Labahn, A. Bornyl (3,4,5-trihydroxy)-cinnamate - An optimized human neutrophil elastase inhibitor designed by free energy calculations. *Bioorg. Med. Chem.*, **2008**, *16*, 2385-2390.
[166] Reddy, M.R.; Erion, M.D. Computer aided drug design strategies used in the discovery of fructose 1,6-bisphosphatase inhibitors. *Curr. Pharm. Des.*, **2005**, *11*, 283-294.
[167] Shirts, M.R. *Calculating Precise and Accurate Free Energies in Biomolecular Systems.* Ph.D. Thesis, Stanford University, **2005**.
[168] Fujitani, H.; Tanida, Y.; Ito, M.; Jayachandran, G.; Snow, C.D.; Shirts, M.R.; Sorin, E.J.; Pande, V.S. Direct calculation of the binding free energies of FKBP ligands. *J. Chem. Phys.*, **2005**, *123*, 084108.
[169] Jayachandran, G.; Shirts, M.R.; Park, S.; Pande, V.S. Parallelized-over-parts computation of absolute binding free energy with docking and molecular dynamics. *J. Chem. Phys.*, **2006**, *125*, 084901.
[170] Sun, F.; Li, P.Y.; Ding, Y.; Wang, L.W.; Bartlam, M.; Shu, C.L.; Shen, B.F.; Jiang, H.L.; Li, S.; Rao, Z.H. Design and structure-based study of new potential FKBP12 inhibitors. *Biophys. J.*, **2003**, *85*, 3194-3201.
[171] Reynolds, C.A.; King, P.M.; Richards, W.G. Free energy calculations in molecular biophysics. *Mol. Phys.*, **1992**, *76*, 251-275.
[172] Straatsma, T.P.; McCammon, J.A. Computational alchemy. *Ann. Rev. Phys. Chem.*, **1992**, *43*, 407-435.
[173] Sevinc, A. Web of science: A unique method of cited reference searching. *J. Nat. Med. Assoc.*, **2004**, *96*, 980-983.

Received: October 15, 2009 **Revised: January 11, 2010** **Accepted: January 12, 2010**