

How to Calculate Binding Constants for Drug Discovery Studies

María J.R. Yunta *

Departamento de Química Orgánica I, Facultad de Química, Universidad Complutense, Madrid, Spain

*Corresponding author: mjryun@quim.ucm.es

Received September 05, 2013; Revised October 07, 2013; Accepted December 15, 2013

Abstract Free energy calculations to predict binding ability of molecules to receptors are a main subject for drug design. A wide range of theoretical methods has been applied to the calculation of binding constants. This mini-review seeks to identify such methods to better achieve correct predictions for drug candidates.

Keywords: *molecular modeling, binding constant, computer aided drug design*

Cite This Article: María J.R. Yunta, "How to Calculate Binding Constants for Drug Discovery Studies." *American Journal of Modeling and Optimization* 1, no. 3 (2013): 61-70. doi: 10.12691/ajmo-1-3-7.

1. Introduction

It is a critical goal of any computational chemistry/biology study to effectively predict free energy of binding for protein–ligand associations, being a key component of present-day rational drug discovery [1]. The potential of molecular simulations to enhance our understanding of drugs behavior relies ultimately on their capability to achieve an accurate ranking of drug binding affinities on clinically relevant time scales. Several computational approaches exist to estimate ligand binding affinities and selectivities, with various levels of accuracy and computational expense [2,3], as free energy perturbation (FEP), thermodynamic integration (TI), linear response (LR), and molecular mechanics Poisson-Boltzmann surface area (MM/PBSA), as it will be seen below.

To optimize this kind of studies, a variety of structure-based computational methods have been developed for determining protein–small molecule binding affinities. These methods can be considered as classified in one of two types: accurate but slow (Type A), and fast but approximate (Type B). In type A methods, protein flexibility is explicitly taken into account, and there is also included an atomic-level description of solvation. In this way, quantitatively reproduction of experimental protein–small molecule absolute binding free energies can be achieved. However, computational requirements for this kind of calculations make the screening of thousands of small molecules against a protein, as required for rational drug design, not possible at the moment. On the other hand, type B methods are sufficiently fast to perform such inhibitor screening, although they used limited descriptions of protein flexibility and solvation problems, which limit their ability to accurately select and rank-order the behavior of small molecules by computed binding affinities, being this the most stringent test of docking, and its most useful possible outcome.

It must be recognized that there is not a *single* global model for each protein–ligand system as there is not a well-determined ionized state for each defined residue at physiological pH and so the appropriate protonation state of the enzyme has to be guessed [4], because protons are not static and the ionization state of residues is a group function. In other words, there may be multiple energetically accessible states for each protein–ligand complex. Moreover, proton transfer between molecules and, in particular, proton migration across hydrogen bonds, has been identified as one of the fundamental mechanisms for biological processes [5] and protonation or deprotonation of titratable groups can cause changes in binding affinities, enzymatic activities, and structural properties of proteins and often represents a key event in enzymatic reactions [6]. Thus, modeling protein–ligand complexes of this type as an ensemble of multiple ionization models would seem to be a more biologically reasonable approach, as probed by Miranda and coworkers [7] for an L-glutamate-receptor complex.

The association of two molecular entities in the biological environment is a process governed by free energy, that is to say that entropy is important. Especially significant is the role of solvent, i.e., water. The displacement of water molecules when two biological molecules associate is clearly a major source of entropy. This is manifested by the hydrophobic effect and desolvation of functional groups buried by the association [8].

Solvent effects also modulate the ionization states of acidic and basic functional groups on the enzyme side chains (or of ligand functional groups) and thereby their biomolecular associations. These properties have been little studied and are poorly understood but must be computationally accounted for in any realistic model. It is well known that "local" pH at specific sites within proteins can be considerably different than the global solution pH, and this, in turn, affects the ionization state of those residues [9-14]. Computational approaches to target

the above factors (entropy, hydrophobicity, solvation/desolvation, and pH) in the biological environment are currently at the cutting edge of simulations of molecular interactions.

2. Computational Strategies

2.1. Use of Molecular Dynamic Simulations

There are different tactics applied to the modeling of complex systems such as those involving protein–ligand interactions in a water environment. One of them is based on extensive molecular dynamics simulations of the complete system in a box including explicit water molecules with boundary conditions. By configuration of these simulations for thermodynamic integration (TI) or free energy perturbation (FEP), the free energy for specific events, e.g., ligand binding, can be estimated [15,16,17,18]. Alternatively, the linear response method analyzes the states of multiple dynamics simulations to obtain a semi-empirical estimation of free energy [19,20], as molecular dynamics simulations, guided by experimental information [21] have successfully reproduced experimental trends in binding affinities [22]. These methods are computationally expensive to perform and are tied to the quality of the molecular mechanics (MM) forcefields used. Thus, only interaction types specifically programmed into the forcefield will be modeled and observed. In particular, MM forcefields do not include specific terms for the hydrophobic effect and hydrophobic–hydrophobic interactions are usually indicated to be energetically unfavorable.

This procedure is costly and prevents the consideration of a large number of conformations, but it is important to point out that insufficient sampling can therefore sometimes lead to incorrect results. To decrease the computational cost, the protein can be truncated beyond a certain cutoff distance and the system minimized using a distance-dependent dielectric, which simulates the deleted surroundings [23]. However, a large variation of the calculated entropy term often results from these ‘free’ minimizations. Inclusion of a fixed buffer region (containing several water molecules) beyond the cut-off can lead to more stable entropy predictions [23]. The internal energy terms (E_{bonded}) of the protein and complex can be on the order of a few thousand kcal/mol, and can introduce large uncertainties in the computed binding free energies. This is prevented in the “single-trajectory” approximation [24,25], which employs simulations of a single state (the complex) to generate conformations for all three states (complex, protein and ligand). In this case, the binding energy of a ligand (L) to a protein (P) to form the complex (PL) is obtained by the equation (1) [26]:

$$\Delta G_{\text{bind}} = G(\text{PL}) - G(\text{P}) - G(\text{L}) \quad (1)$$

and the free energy of each of the three molecular systems, P, L, and PL is given by the expression:

$$G(\text{X}) = E_{\text{MM}}(\text{X}) + G_{\text{solv}}(\text{X}) - TS(\text{X}) \quad (2)$$

where E_{MM} is the total molecular mechanics energy of molecular system X in the gas phase, G_{solv} is a correction term to account for the fact that X is surrounded by solvent that can be estimated as the sum of polar

contributions accounted by the generalized Born, Poisson, or Poisson-Boltzmann model, and non-polar contributions, assumed to be proportional to the solvent-accessible surface area [27,28], and S is the entropy of X.

For each MD conformation sampled, the resulting internal energy terms of the protein and ligand are identical in the bound and the unbound states and cancel exactly in Eq. (1). Hence, effectively only the protein–ligand (non-bonded) interaction energies of the E_{MM} term contribute to ΔG_{bind} . ‘Single-trajectory’ simulations significantly reduce computational effort and are generally sufficiently accurate for most applications.

Using this approximation is important to be aware of the fact that any explicit structural relaxation of the protein and ligand upon binding is ignored. Although charge reorganization can be partly taken into account implicitly by setting the protein/ligand (internal) dielectric constants to values larger than $\epsilon_{\text{in}} = 1-2$ [29,30,31,32], the neglect of explicit structural relaxation may introduce errors depending on the system [33].

To improve the accuracy of the computed binding free energies, the various terms of Eq. (2) are averaged over multiple conformations or MD snapshots (typically a few hundred for the E_{MM} and G_{solv} contributions). Depending on the extent of conformational fluctuations in the system under consideration, the convergence into stable values may require relatively long (multi-ns) simulations. The computation of the entropy term, however, requires the extensive minimization of the trajectory conformations for the protein, ligand, and complex, to local minima on the potential energy surfaces, followed then by normal mode analysis [34].

Separate MD simulations for the complex, unbound receptor, and ligands, may also be performed (the “three-trajectory” approximation) but require greater computational effort, although in theory should yield more accurate results. Indeed, Yang and coworkers [35] have recently shown that including separate simulations for the ligand and accounting for the “ligand reorganization” free energy led to significant improvements in binding affinity predictions for a set of ligands targeting the X-linked inhibitor of apoptosis protein (XIAP). In certain cases, therefore, the added expense of separate simulations may be justified.

The Molecular Mechanics/Generalized Born Surface Area (MM-GBSA) and Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) methods calculate binding free energies using molecular mechanics (forcefields) and continuum (implicit) solvation models [36]. They have been successfully applied across a range of targets and are implemented in software programs such as Amber [37], Delphi [38] and Schrödinger [39].

MM-GB(PB)SA methods are widely recognized as valuable tools in computer aided drug design (CADD) applications. However, as with any method, they have limitations and caveats, which need to be considered. First, while useful for ranking relative ligand binding affinities, these methods lack the required accuracy for absolute binding free energy predictions [40,41]. The inclusion of entropic contributions brings the MM-GB(PB)SA values somewhat closer to experimental absolute affinities [42]. However, such entropic terms are costly and contain large uncertainties. Forcefield inconsistencies may also be an issue: PB and GB results depend strongly on adequate

atomic charges and van der Waals radii, which are often optimized for MD simulations. The MM-GB(PB)SA results may be influenced by system-dependent properties, such as the features of the binding site, the extent of protein and ligand conformational relaxation upon association, and the protein and ligand charge distribution [40,43]. Continuum electrostatics models ignore the molecular structure of the solvent, and in some cases this might affect the results, particularly when key receptor-ligand interactions are bridged by water molecules [44]. Furthermore, the value of the protein/ligand dielectric constant is empirically chosen, and takes into account not only the protein and ligand structural relaxation, but also other error-introducing factors such as the ones mentioned above [30,31]. Hou and coworkers suggested in a recent MM-PBSA study that the use of $\epsilon_{in} = 4$ for a highly charged protein-ligand binding interface, $\epsilon_{in} = 2$ for a moderately charged binding interface and $\epsilon_{in} = 1$ for a hydrophobic binding interface may improve ligand ranking [40], but the lack of a consistent optimum dielectric constant for MM-PBSA calculations has been noted [34]. In this context, recent studies on MM/GBSA binding free energy calculations show that a nanosecond scale MD simulation is sufficient to perform a meaningful MM/GBSA calculation [45,46,47].

MM/GBSA method [48,49] was developed for free energy calculations and has been used to estimate the binding affinity for several protein or DNA systems [50,51,52,53] and to perform ensemble average MM/GBSA binding free energy calculation on the snapshots from the MD simulation to compare binding affinity of a known drug with that of the co-crystallized ligands [54]. In this last case, bondi radii (mbondi2) [27,55,56] in AMBER9 was applied to calculate this part of energy. This model was newer than the original version of the GB model and provided a significant improvement and has been recommended for both proteins and nucleic acids. In this case the interior dielectric constant of the molecule of interest was set as 1.0 and the exterior or solvent dielectric constant was set as 78.5. Being the non-polar contribution to the solvation free energy proportional to the solvent-accessible surface area, this was calculated by the LCPO model [57] and the surface tension used to calculate the non-polar part was taken as $0.0072 \text{ kcal/mol}\text{\AA}^2$. It has been stated that the entropic term is the most time-intensive part of the MM/GBSA calculation but, being found to be indistinguishable among different conformational states, and also contributes less than the other two terms in many applications for estimating relative binding free energies [48,58,59]. The entropy change associated with ligand binding has been estimated by normal mode analysis [60] in AMBER9. There are different opinions over the number of snapshots to be used in this type of studies as, for example, some papers have proposed the MM/GBSA calculation for each system to be carried out on the 200 snapshots extracted from the last 2ns of the MD simulation [54] although others used 900 snapshots along the entire simulation [8].

A recent study by Kuhn and coworkers suggests that the MM-PBSA function could be used as a post-docking filter during the virtual screening of compounds [43], as their use of a single relaxed structure provided better results than usual averaging over MD simulation snapshots.

2.2. Use of FEP-based Methods

Other class of tactic for simulating the biological environment is based on the electrostatic properties of the molecules and their response to the dielectric variations at the molecule/solvent interface. Several implementations of solutions to the Poisson–Boltzmann equation (PBE) have been used in the biological environment [9,14,61]. While PBE approaches are very robust in terms of representing electrostatic effects, particularly at molecular surfaces, they do not have terms to represent other non-covalent forces and so can be of limited utility in cases that are not dominated by electrostatic effects. For this reason, some of the more recent simulations have utilized methods combining PBE (or generalized Born, a faster, less robust electrostatic method) with FEP or TI calculations [62,63,64].

Free energy perturbation (FEP)-based methods are low in throughput but produce calculated results quantitatively consistent with experimental results [65,66,67]. Success has been demonstrated for calculations of relative differences in lipophilicity, ionization, covalent hydration, and solvation [68,69,70]. Moreover, and most important for drug design, the approach is reported to accurately predict relative binding free energies for inhibitors of numerous enzymes [71,72,73], including those considered as potential drug targets [74]. Despite these impressive results, FEP calculations are rarely used in the pharmaceutical industry. Much of the resistance stems from the complexity of the method and its low throughput. FEP calculations are CPU demanding and require the availability of validated molecular mechanics force field parameters to achieve high accuracy. Since most drug candidates contain substructures not fully described by existing parameters, the user must develop and input parameters prior to initiating the calculation. This process is time-consuming and often limited by the absence of relevant experimental data. Accordingly, FEP calculations are inherently difficult to automate and require considerable user expertise and judgment to complete successfully.

2.3. Use of Empirical Free Energy Scoring Algorithms

A third tactic for understanding the biological environment is the application of empirical free energy scoring algorithms. These tools have often been developed for feedback in docking or de novo molecule design programs, and thus have been optimized for speed. In general, simple metrics such as partial charge and solvent accessible surface area, combined with structural features as extracted from experimental measurements, produce a “score” that can often be correlated with free energy [75,76]. For example, the hydrophobic effect is usually represented by apolar surface contact area. While useful within their native context, i.e., docking or de novo building, these algorithms are often invalid outside their training and validation environment. Related to this, the HINT forcefield described by Kellog [8] is empirical, but differs from other scoring functions in that all key parameters are derived from an experiment that uniquely measures intermolecular interactions in the biological environment, as their interaction constants are obtained from measurements of the partition coefficient for 1-octanol/water solubility, usually used as $\log P_{o/w}$.

In another popular approach, the linear interaction energy (LIE) method, the absolute binding free energy of a ligand to a protein is obtained by running two independent simulations; being one the ligand free in solution and the other the solvated protein-ligand complex [77]. An important drawback of the LIE method is that energy descriptors have to be extracted from the simulations and correlated empirically to the activity of a training set of compounds. This limits the application of LIE to systems for which sufficient experimental data is available [78]. Continuum solvent models have been introduced in the LIE methodology to increase efficiency and test their accuracy [79,80,81].

The binding free energies obtained in all these cases were in very good agreement with experimental results and the LIE approach seems to be a good alternative to the more expensive FEP calculations. The two main shortcomings of the method are the need for two different MD simulations, one of the complex structure and another for the free ligand in water, and the use of empirically derived constants which may need to be modified for each particular system. These requirements restrict the broad application of the LIE method in docking/scoring procedures [82].

To better account for receptor flexibility, a new virtual-screening protocol called the relaxed complex scheme (RCS) has been developed [83,84]. Rather than docking many compound models into a single NMR or crystal structure, each potential ligand is docked into multiple protein conformations, typically extracted from a molecular dynamics simulation. Thus, each ligand is associated not with a single docking score but rather with a whole spectrum of scores.

While these successes are promising, the relaxed complex scheme certainly has its weaknesses. Aside from being based on molecular dynamics simulations that are themselves subject to crude force-field approximations and inadequate conformational sampling, the scheme relies on computer-docking scoring functions that of necessity are optimized for speed at the expense of accuracy. In order to facilitate high-throughput virtual screening, these scoring functions often treat subtle influences on binding energy, like conformational entropy and solvation energy, only superficially [85,86], thus sacrificing accuracy for the sake of greater speed.

The most practical and convenient approach to address the docking problem seems to be a two-step protocol. Fast and less accurate algorithms are first used to scan large databases of molecules and reduce their size to a reasonable number of hits. This step is then followed by application of more accurate and time-consuming methods which can refine the conformation of the complexes and produce accurate free energies [87,88].

Molecular dynamics simulations present an attractive alternative for structural refinement of the final docked complexes. They incorporate flexibility of both ligand and receptor, improving interactions and enhancing complementarity between them, and thus accounting for induced fit. Moreover, the evolution of the complexes over the simulation timecourse is an indication of their stability and reliability: incorrectly docked structures are likely to produce unstable trajectories, leading to the disruption of the complex, while realistic complexes will show stable behavior. In addition, the ability to

incorporate explicit solvent molecules and their interactions in the simulations of the docked systems is very important for understanding the role of water and its effect on the stability of the ligand-protein complexes [85].

3. Calculation of Relative Binding Affinities

The “alchemical” computation of *differences* in binding affinities (rather than *absolute* affinities) among a set of related ligands for the same target protein is more accurate and technically simpler. A thermodynamic cycle illustrating the basic principles is shown in Figure 1 [89]. The horizontal legs describe the experimentally accessible actual binding processes, with free energies $\Delta G_{\text{bind}}(\text{L1})$ and $\Delta G_{\text{bind}}(\text{L2})$. Since the free energy is a state function, the relative binding free energy $\Delta\Delta G_{\text{bind}}$ is exactly equal to the difference of the free energies in the horizontal or vertical legs:

$$\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}(\text{L2}) - \Delta G_{\text{bind}}(\text{L1}) \quad (3)$$

$$= \Delta G_{\text{complex}}(\text{L1} \rightarrow \text{L2}) - \Delta G_{\text{free}}(\text{L1} \rightarrow \text{L2}) \quad (4)$$

The simulations follow the vertical steps (Eq. (4)) or unphysical processes, by simulations in water solution that gradually change the energy-function of the system from one “endpoint” to the other through a series of intermediate hybrid states. From Figure 1, this involves the stepwise “alchemical” transformation of ligand L1 to L2 both in its ‘free’ state (unbound) and in the bound complex, through gradual changes in the forcefield parameters describing the ligand interactions. This leads to the free energy changes $\Delta G_{\text{free}}(\text{L1} \rightarrow \text{L2})$ and $\Delta G_{\text{complex}}(\text{L1} \rightarrow \text{L2})$, respectively. Averaging over both transformation directions is often used to improve the free-energy estimates, although this is not always the case [90]. These calculations can be accurate, if conducted with the appropriate care. An overview of current state-of-the-art methods for absolute and relative affinity calculations can be found in Chodera's paper [91].

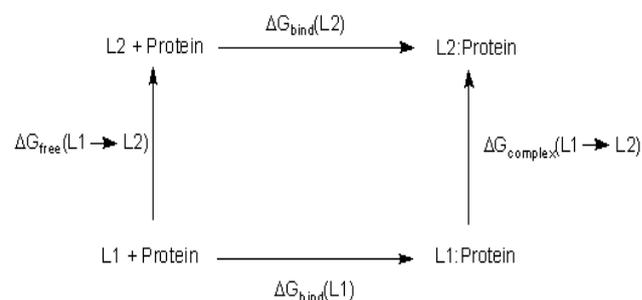


Figure 1. Thermodynamic cycle linking the binding of two ligands L1 and L2 to a protein in solution.

In any case, the results can be compared with the experimentally derived relative free energies of binding, obtained from the IC_{50} values via [92]:

$$\Delta\Delta G_{\text{AB}}^{\text{binding}}(\text{exp}) = -RT \ln \left(\text{IC}_{50}^{\text{A}} / \text{IC}_{50}^{\text{B}} \right) \quad (5)$$

A significant difficulty in the calculation of protein-ligand relative binding free energies arises from the sampling of the many protein and solvent degrees of

freedom, in addition to the ligand degrees of freedom. An implicit solvent framework reduces such complexity, but the degrees of freedom of protein side chains must still be sampled.

Other limitations make free energy calculations less applicable than LIE or MM/PBSA. Usually the study is limited to congeneric inhibitors due to the difficulty of changing one ligand into another unrelated ligand. This is partly an issue of convergence of the free energies, which can be solved by running longer simulations, and complex system setup, which prevents automation. It is, therefore, important to concentrate methodological efforts on the development of rigorous free energy methods that allow fast, reliable calculations of free energy differences between structurally dissimilar ligands.

A trade-off between computational expense and accuracy is therefore required when the goal is to investigate and compare the binding strengths of a structurally diverse and/or larger set of ligands via MD simulations. For this purpose, much less computationally demanding “endpoint” methods are often successfully applied, such as the “linear interaction energy” (LIE) [93] or the molecular mechanics – Poisson Boltzmann (MM-PBSA) [36] and the related molecular mechanics – generalized Born (MM-GBSA) approximation [94]. All these methods compute binding free energies along the horizontal legs of Figure 1, but use only models for the “endpoints” (bound and unbound states).

Like any other spontaneous process, a noncovalent binding event takes place only when it is associated with a negative binding free energy (ΔG), which is the well-known sum of an enthalpic term (ΔH) and an entropic term ($-T\Delta S$). These terms may be of equal or opposite sign and thus lead to various thermodynamic signatures of a binding event, ranging from exothermic to entropy-driven.

A rough correlation between the burial of apolar surface area and free energy could be derived [95], but beyond that, practically useful relationships between structure and the components of free energy have remained elusive. This is not surprising, as both entropy and enthalpy terms obtained from calorimetric experiments contain solute and solvent contributions and thus cannot be interpreted on the basis of structural data alone. The direct experimental estimation of solvent effects has been attempted [96] but always requires additional assumptions. Only theoretical treatments allow a separation of those effects [97]. Thus, computer modeling can support the interpretation of experimental observations.

Since entropic and enthalpic components of binding are highly dependent on many system-specific properties, the practitioner has to conclude that optimizing for free energy is still the only viable approach to structure-based design. Perhaps the greatest advantage in the attempt to interpret components of ΔG is that it forces us to think about two fundamental topics in unprecedented detail: viewing protein-ligand complexes as flexible entities rather than fixed structures and the role of desolvation effects.

A discussion of the thermodynamics of ligand binding is never complete without mentioning the phenomenon of entropy-enthalpy compensation. The validity and generality of this phenomenon have been a contentious topic for many years. There is ample evidence of meaningless and spurious correlations between ΔH and

$T\Delta S$, often due to far larger variations (sometimes through larger experimental errors) in ΔH than in ΔG [98,99]. Also, compensation is no thermodynamic requirement [100,101]. If changes in ΔH were always compensated by opposing changes in $T\Delta S$, optimization of binding affinities would hardly be possible.

Nevertheless, there seems to be a mechanism behind the compensation frequently observed [102] in host-guest chemistry [103] and in protein-ligand interactions [104]. In short, the tighter and more directed an interaction, the less entropically favorable it is. Bonding opposes motion, and motion opposes bonding [105]. However, the detailed nature of these compensatory mechanisms is highly system-dependent and the mechanisms do not obey a single functional form. As a consequence, noncovalent interactions can be positively cooperative; that is, the binding energy associated with their acting together is greater than the sum of the individual binding free energies.

What can we conclude from this statement? First, small changes in ΔG often mask large and mutually compensating changes in ΔH and $T\Delta S$. Focusing on ΔG in designing new molecules is certainly still the safest bet. Second, in medicinal chemistry we often rely on cooperative effects without even noticing, and yet in our minds we attempt to decompose binding free energies into additive elements. There is nothing wrong with this empirical approach as long as we keep in mind that it is primarily useful to teach us about the limits of additivity. The knowledge that specific interactions, in particular strongly directed ones like hydrogen bonds, rigidify a protein-ligand complex may help us to exploit cooperativity in a more rational fashion. The Klebe data [106,107] should also make us think about the degree of mobility required in interactions with different parts of the protein. Flexible domains may require more flexible ligand moieties than highly ordered ones. The thermodynamic signature of a “good” ligand is not necessarily dominated by an enthalpic term. Our traditional focus on visible interactions and on the induced fit model has led to an overly enthalpic view of the world that neglects flexibility and cooperativity. It also neglects local solvation phenomena, whose details strongly affect the thermodynamic profile of a molecular recognition event. We should learn to incorporate these into our thinking, at least in a qualitative fashion.

4. The Water Role

A hydrophobic solute disrupts the structure of bulk water and decreases entropy because of stronger bonding and ordering of water molecules around the solute. Such disruptions can be minimized if nonpolar solute molecules aggregate. In this case, water forms one larger “cage” structure around the combined solutes, whose surface area will be smaller than the combined surface areas of isolated solutes. This maximizes the amount of free water and, thus, the entropy. If this mechanism were the sole driving force for a protein-ligand interaction, all binding events involving hydrophobic partners would be entropy-driven. But spectroscopic evidence indicates that hydrogen bonds at hydrophobic surfaces are weaker [108] and water molecules more flexible [109] than originally assumed. In

addition, already simple water models show that size and surface curvature of solutes have a dramatic impact on their solvation thermodynamics [110]. Complexation thermodynamics driven by enthalpic forces have been regularly observed in host-guest chemistry, being known as the “nonclassical hydrophobic effect” [111].

Any ligand binding event displaces water molecules from the binding site. In structure-based design, most of these are never explicitly considered because they are highly disordered and therefore rarely crystallographically observed [112]. Enzyme binding sites are in fact characterized by easily displaceable water, as shown by Ringe in her seminal work on transferring protein crystals to organic solvents [113]. Those water molecules that are observed need to be carefully analyzed; they might be replaceable or they might be considered as part of the protein structure [114]. Extensive computer simulations are not generally feasible in a fast-paced drug discovery environment, but simple geometric parameters describing the immediate protein environment of a crystallographic water molecule can serve as a useful guide to estimate whether it is displaceable by a ligand moiety [115]. Bissantz published a simple geometric rank function based on the distances and angles of neighboring donor and acceptor atoms in the protein, developed by Kellogg and coworkers, that served as a practically useful metric [116]. Water molecules with high ranks should be regarded as part of the binding site [117,118]. These experience three or more hydrogen bonds with the local environment and are usually located in buried polar cavities.

While the preferred geometries of hydrogen bonds are easily defined, their contribution to affinity is highly context dependent. While the hydrogen bond distances of salt bridges are shorter than those between neutral motifs, indicating stronger interactions, this does not necessarily translate into more favorable binding free energies. The contribution of charge assisted protein-ligand hydrogen bonds to binding depends critically on the protein environment. Theoretical and experimental studies of solvent-exposed salt bridges reveal little free energy gain for the pairing of monovalent ions [119].

But there are also a lot of things to consider: weak hydrogen bonds, halogen bonds, orthogonal multipolar interactions, halogens and aromatic rings and hydrophobic interactions. They contribute in a certain extent to protein-ligand interaction but had not yet become a usual part of crystal structure visualization and molecular design programs. Visualization often does not go beyond the concepts of hydrogen bonds and van der Waals contacts, except in the form of PyMol scripts as part of Proasis2 [120], although this lack had only developed important when detailed studies of interactions are pursued, as many studies have shown that the single best structural parameter correlating with binding affinity is the amount of hydrophobic surface buried upon ligand binding, and this can be just the thing to look for in certain types of relationship studies.

5. Protein Flexibility

The importance of protein flexibility for protein function is well-established [121], with the induced-fit and

population-shift models used to rationalize [122] the binding of small molecules to protein conformations not seen in structural studies of apo proteins, and it is also necessary to use programs allowing full flexibility of enzyme and drug, when there is no Protein Data Bank entries of enzymes complexed with molecules of adequate size to study the binding of new drugs into enzymes, as in this case docking programs actually in use just position the drug molecule in the outside of the enzyme, as there is not enough space in the active center of the rigid enzyme for the drug to enter, giving poor scores.

Explicitly sampling some subset of the protein degrees of freedom during docking is another way of including protein flexibility [123]. One advantage of this approach is that the protein and the small molecule can simultaneously relax to accommodate each other. However, this approach requires judicious choice of a subset of degrees of freedom to sample (e.g. side chains in the binding pocket) in order to remain computationally tractable. A recent combination of simultaneous optimization of protein and small molecule degrees of freedom combined with ensemble docking looks to be promising [124], as does an automated approach in which the small molecule and the protein are iteratively optimized with the other molecule kept rigid [125].

Sanchez-Moreno and coworkers [126] published a way to perform this kind of molecular modeling studies using the AMBER [127] method implemented in Hyperchem 8.0 package [128], modified by the inclusion of appropriate parameters [129]. Starting structures for compounds were built by using Hyperchem capabilities and their geometries were minimized to a maximum energy gradient of 0.1 Kcal/(Å mol) with the AMBER force field, using the Polak-Ribiere (conjugate gradient) minimizer, and ‘simulated annealing’ procedure was used to cover all conformational space. The most stable extended geometry was always used in all calculations of interaction with the enzyme, in order to suit protein-drug interactions. To mimetize the conditions used in the activity measurements, *i. e.*, water as solvent, all calculations were carried out *in vacuo* with distance dependent dielectric constant value. In the absence of explicit solvent molecules, a distance dependent dielectric factor qualitatively simulates the presence of water, as it takes into account the fact that the intermolecular electrostatic interactions should die off more rapidly with distance than in the gas phase [130]. Although the same results can be obtained using a constant dielectric factor greater than 1, they used a distance-dependent dielectric constant ($\epsilon = 4 R_{ij}$) as this was the method used to develop AMBER forcefield [127], assigning charges for all atoms by means of *ab initio* calculations using STO-3G basis set, as it is compatible with the AMBER force field used, prior to energy minimization using AMBER. The enzyme structure was obtained from the Brookhaven protein data bank and its energy minimized in the same way. Interaction studies were done starting from structures having the compound positioned in the border of the enzyme cavity, as this was the best position found with standard docking programs, and entering into the cavity was forced by the use of a restraint to the selected drug-active center distance, starting from 0.15 nm and slowly decreasing this distance, letting the complex to achieve the minimum energy conformation with no restraints for all

the small driving steps, using the same conditions mentioned above.

Closely related, it has been described a ligand-based homology modeling (LBHM) approach that has been used to simulate the conformational changes induced by ligand binding [131]. Interestingly, none of the antagonists used in the study found an energetically stable docking pose in the conventional RBHM-driven A3 model, mainly due to the unfavorable topological complementarity among these antagonists and corresponding RBHM-driven TM binding cavity. In particular, highly destabilizing van der Waals interactions (steric conflicts) seem to be the reason for a lack of topological complementarities. These steric conflicts were drastically reduced or completely eliminated after application of the LBHM approach [131].

Do these results mean free energy calculations should be used more widely? It is believed that the answer is yes if crystallographic evidence or reliable docking predictions are available, the type of ligands studied do not cause significant conformational changes of the binding site, and the goal is to optimize substituent placement on a scaffold. Because of limitations in sampling algorithms, force fields, and protein models, it is unrealistic to expect high accuracy predictions to be obtained all the time, but the data presented here suggest that the calculations work better than the empirical scoring functions typically available to a modeler and at a computational cost that has become affordable.

6. Conclusions

Modern molecular modeling techniques are remarkable tools in the search for potentially novel active agents by helping to understand and predict the behaviour of molecular systems, having assumed an important role in the development and optimization of leading compounds, ranking the possible candidates in accordance with their predicted binding constants.

Although improvements are still needed in the techniques used, they have shown to be invaluable in structure-activity relationship researchs.

Competing Interests

The author declare no competing financial interest.

Acknowledgement

The author thanks the MCINN Project: ConsoliderIngenio CSD2010-00065 for financial support.

References

- [1] Guvench, o., MacKerell Jr. A.D. Computational evaluation of protein - small molecule binding. *Curr Opin Struct Biol.*, **2009**, *19*, 56-61.
- [2] 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. *Annu. Rev. Biophys. Biomol. Struct.* **2001**, *30*, 211-243.
- [3] Wan, S.; Coveney, P. V.; Flower, D. R. Peptide recognition by the T cell receptor: comparison of binding free energies from thermodynamic integration, Poisson-Boltzmann and linear interaction energy approximations. *Philos. Trans. R. Soc. A* **2005**, *363*, 2037-2053.
- [4] Kovalsky, D.; Dubyna, V.; Mark, A. E.; Korenelyuk, A. A molecular dynamics study of the structural stability of HIV-1 protease under physiological conditions: The role of Na⁺ ions in stabilizing the active site. *Proteins: Struct, Funct. Bioinf.* **2005**, *58*, 450-458.
- [5] Bountis, T. (Ed.), Proton Transfer in Hydrogen-Bonded Systems, *NATO ASI Series B: Physics*, vol. 291, Plenum Press, New York, **1992**, pp. 1-355.
- [6] Fornabai, M., Cozzini, P., Mozzarelli, A., Abraham, D. J., Kellogg, G. E. Simple, Intuitive Calculations of Free Energy of Binding for Protein-Ligand Complexes. 2. Computational Titration and pH Effects in Molecular Models of Neuraminidase-Inhibitor Complexes. *J. Med. Chem.* **2003**, *46*, 4487-4500.
- [7] Miranda, C., Escartí, F., Lamarque, L., Yunta, M. J. R., Navarro, P., GarcíaEspaña, E., Jimeno, M. L. New 1H-pyrazole-containing polyamine receptors able to complex L-glutamate in water at physiological pH values. *J. Am. Chem. Soc.* **2004**, *126*, 823-833.
- [8] Kellogg, G. E., Fornabai, M., Spyrikis, F., Lodola, A., Cozzini, P., Mozzarelli, A., Abraham, D. J. Getting it right: modeling of pH, solvent and "nearly" everything else in virtual screening of biological targets. *J. Molec. Graph. Mod.*, **2004**, *22*, 479-486.
- [9] Dong, F., Olsen, B., Baker, N. A. *Computational Methods for Biomolecular Electrostatics. Meth. Cell Biol.* **2008**, *84*, 843-870.
- [10] Miteva, M.A., Tufféry, P., Villoutreix, B.O. PCE: web tools to compute protein continuum electrostatics. *Nuc. Acids. Res.*, **2005**, *33*, w 372-w 375.
- [11] Varekova, R. S., Geidl, S., Ionescu, C-M., Skrehota, O., Bouchal, T., Sehnal, D., Abagyan, R., Koca, J. Predicting pK_a values from EEM atomic charges. *J. Cheminform.* **2013**, *5*:18.
- [12] Witham, S., Talley, K., Wang, L., Zhang, Z., Sarkar, S., Gao, D., Yang, W., Alexov, E. Developing hybrid approaches to predict pK_a values of ionizable groups. *Proteins*, **2011**, *79*, 3389-3399.
- [13] Yikilmaz, E., Rodgers, D. W., Miller, A-F. The Crucial Importance of Chemistry in the Structure-Function Link: Manipulating Hydrogen Bonding in Iron-Containing Superoxide Dismutase. *Biochemistry* **2006**, *45*, 1151-1161.
- [14] Nielsen, J. E., McCammon, J. A. On the evaluation and optimization of protein X-ray structures for pK_a calculations, *Protein Sci.* **2003**, *12*, 313-326.
- [15] Chipot, C., Porhorielle, A. (Eds.) Free energy calculations. Theory and applications in chemistry and biology series. **2007**.
- [16] Reddy, M. R., Erion, M.D., Agarwal, A. Free energy calculations: use and limitations in predicting binding affinities, *Rev. Comput. Chem.* **2000**, *16*, 217-304.
- [17] Straatsma, T. P. Free energy by molecular simulation, 2007, in *Reviews in Computational chemistry*. Vol 9 (Eds. K. B. Lipkowitz and D. B. Boyd), John Wiley & Sons, Inc., Hoboken, NJ, USA, pp 81-127.
- [18] Wu, K. W., Chen, P.C., Wang, J., Sun, Y. C. Computation of relative binding free energy for an inhibitor and its analogs binding with Erk kinase using thermodynamic integration MD simulation. *J. Comput. Aided Mol. Des.* **2012**, *26*, 1159-69.
- [19] Åqvist, J., Luzhkov, V. B., Brandsal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.* **2002**, *35*, 358-365.
- [20] Jorgensen, W. L., Tirado-Rives, J. Potential Energy Functions for Atomic-Level Simulations of Water and Organic and Biomolecular Systems. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 6665-6670.
- [21] Ode, H.; Matsuyama, S.; Hata, M.; Hoshino, T.; Kakizawa, J.; Sugiura, W. *J. Med. Chem.* **2007**, *50*, 1768-1777.
- [22] Stoica, I., Sadiq, S. K., Coveney, P. V. Rapid and Accurate Prediction of Binding Free Energies for Saquinavir-Bound HIV-1 Proteases. *J. Am. Chem. Soc.* **2008**, *130*, 2639-2648.
- [23] Kongsted, J., Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comput. Aid. Molec. Design*, **2009**, *23*, 63-71.
- [24] Gohlke, H., Case, D. Converging Free Energy Estimates: MM-PB (GB) SA Studies on the Protein-Protein Complex Ras-Raf. *J. Comput. Chem.* **2004**, *25*, 238-250.
- [25] Page, C.S., Bates, P.A. Can MM-PBSA Calculations Predict the Specificities of Protein Kinase Inhibitors? *J. Comput. Chem.*, **2006**, *27*, 1990-2007.
- [26] Pearlman, D.A. Evaluating the molecular mechanics Poisson-Boltzmann surface area free energy method using a congeneric

- series of ligands to p38 MAP kinase. *J. Med. Chem.* **2005**, *48*, 7796-7807.
- [27] Onufriev, A., Bashford, D., Case, D.A. Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B*, **2000**, *104*, 3712-3720.
- [28] Rizzo, R. C., Aynechi, T., Case, D.A., Kuntz, I. D. Estimation of absolute free energies of hydration using continuum methods: accuracy of partial charge models and optimization of non-polar contributions. *J. Chem. Theory Comput.* **2006**, *2*, 128-139.
- [29] Archontis, G.; Simonson, T., Karplus, M. Binding free energies and free energy components from molecular dynamics and Poisson-Boltzmann calculations. Application to amino acid recognition by aspartyl-tRNA synthetase. *J. Mol. Biol.* **2001**, *306*, 307-327.
- [30] Archontis, G., Simonson, T. Proton Binding to Proteins: A Free-Energy Component Analysis Using a Dielectric Continuum Model. *Biophys. J.*, **2005**, *88*, 3888-3904.
- [31] Schutz, C. N., Warshel, A. What are the dielectric constants of proteins and how to validate electrostatic models? *Proteins-Structure, Function and Genetics*, **2001**, *44*, 400-417.
- [32] Simonson, T. Electrostatics and dynamics of proteins. *Reports of Progress in Physics*, **2003**, *66*, 737-787.
- [33] Tamamis, P., Morikis, D., Floudas, C. A., Archontis, G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Proteins: Structure, Function & Bioinformatics*, **2010**, *78*, 2655-2667.
- [34] Hayes, J. M., Archontis, G. MM-GB (PB) SA Calculations of Protein-Ligand Binding Free Energies, Molecular Dynamics - Studies of Synthetic and Biological Macromolecules, Wang, L. (Ed.), **2012**, In Tech, Ch. 9; pp 171-190.
- [35] Yang, C-Y., Sun, H., Chen, J., Nikolovska-Coleska, Z., Wang, S. Importance of ligand reorganization free energy in protein-ligand binding affinity prediction. *J. Am. Chem. Soc.* **2009**, *131*, 13709-13721.
- [36] Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D. A., Cheatham, T.E. III, Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*, 889-897.
- [37] Case, D.A., Cheatham, T. E., III, Darden, T., Gohlke, H., Luo, R., Merz, K. M., Jr., Onufriev, A., Simmerling, C., Wang, B., Woods, R. The Amber biomolecular simulation programs. *J. Comput. Chem.*, **2005**, *26*, 1668-1688.
- [38] Rocchia, W., Alexov, E., Honig, B. Extending the applicability of the nonlinear Poisson-Boltzmann equation: Multiple dielectric constants and multivalent ions. *J. Phys. Chem. B*, **2001**, *105*, 6507-6514.
- [39] Du, J., Sun, H., Xi, L., Li, J., Yang, Y., Liu, H., Yao, X. Molecular modeling study of Checkpoint Kinase 1 inhibitors by multiple docking strategies and Prime/MMGBSA. *J. Comput. Chem.*, **2011**, *32*, 2800-2808.
- [40] Hou, T., Wang, J., Li, Y., Wang, W. Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. *J. Chem. Inform. Model.* **2011**, *51*, 69-82.
- [41] Singh, N., Warshel, A. Absolute binding free energy calculations: On the accuracy of computational scoring of protein-ligand interactions. *Proteins: Structure, Function & Bioinformatics*, **2010**, *78*, 1705-1723.
- [42] Gilson, M. K., Zhou, H-X. Calculation of protein-ligand binding affinities. *Ann. Rev. Biophys. Biomol. Struct.*, **2007**, *36*, 21-42.
- [43] Kuhn, B., Gerber, P., Schulz-Gasch, T., Stahl, M. Validation and use of the MMPBSA approach for drug discovery. *J. Med. Chem.*, **2005**, *48*, 4040-4048.
- [44] Hayes, J.M., Skamnaki, V.T., Archontis, G., Lamprakis, C., Sarrour, J., Bischler, N., Skaltsounis, A. L., Zographos, S.E., Oikonomakos, N. G. Kinetics, in silico docking, molecular dynamics, and MM-GBSA binding studies on prototype indirubins, KT 5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: The role of water molecules examined. *Proteins: Structure Function & Bioinformatics*, **2011**, *79*, 703-719.
- [45] Rastelli, G., Del Rio, A., Degliesposti, G., Sgobba, M. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **2009**, *31*, 797-810.
- [46] Hu, R., Barbault, F., Maurel, F., Delamar, M., Zhang, R. Molecular dynamics simulations of 2-amino-6-arylsulphonylbenzotriazoles as HIV inhibitors: interaction modes and binding free energies. *Chem. Biol. Drug. Des.*, **2010**, *76*, 518-526.
- [47] Balus, T. E., Rizzo, R.C. Quantitative prediction of fold resistance for inhibitors of EGFR. *Biochemistry*, **2009**, *48*, 8435-8448.
- [48] 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, T. E. III. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.*, **2000**, *33*, 889-897.
- [49] Lee, M.S., Olson, M.A. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.* **2006**, *90*, 864-877.
- [50] Baris, I., Tuncel, A., Ozber, N., Keskin, O., Kavakli, I. H. Investigation of the Interaction between the Large and Small Subunits of Potato ADP-Glucose Pyrophosphorylase. *PLoS Comput. Biol.*, **2009**, *5*, e1000546.
- [51] Chen, Q., Cui, W., Ji, M. Studies of chirality effect of 4-(phenylamino)-pyrrolo[2,1-f][1,2,4]triazine on p38 α by molecular dynamics simulations and free energy calculations. *J. Comput. Aided. Mol. Des.* **2009**, *23*, 737-745.
- [52] Li, W. H., Tang, Y., Hoshino, T., Neya, S. Molecular modeling of human cytochrome P450 2W1 and its interactions with substrates. *J. Mol. Graph. Model.* **2009**, *28*, 170-176.
- [53] Yang, C. Y., Sun, H. Y., Chen, J. Y., Nikolovska-Coleska, Z., Wang, S. M. Importance of Ligand Reorganization Free Energy in Protein-Ligand Binding-Affinity Prediction. *J. Am. Chem. Soc.* **2009**, *131*, 13709-13721.
- [54] Xie, L., Evangelidis, T., Xie, L., Bourne, P.E. Drug Discovery Using Chemical Systems Biology: Weak Inhibition of Multiple Kinases May Contribute to the Anti-Cancer Effect of Nelfinavir. *PLoS Comput. Biol.*, **2011**, *7*, e1002037.
- [55] Feig, M., Onufriev, A., Lee, M.S., Im, W., Case, D.A., Brooks, C. L. III. Performance comparison of generalized born and Poisson methods in the calculation of electrostatic solvation energies for protein structures. *J. Comput. Chem.*, **2004**, *25*, 265-284.
- [56] Onufriev, A., Bashford, D., Case, D.A. Exploring protein native states and large-scale conformational changes with a modified generalized born model. *Proteins*, **2004**, *55*, 383-394.
- [57] Weiser, J., Shenkin, P.S., Still, W. C. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J. Comput. Chem.* **1999**, *20*, 217-230.
- [58] Gouda, H., Kuntz, I.D., Case, D.A., Kollman, P.A. Free energy calculations for theophylline binding to an RNA aptamer: Comparison of MM-PBSA and thermodynamic integration methods. *Biopolymers*, **2003**, *68*, 16-34.
- [59] 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 villin headpiece. *Proteins* **2000**, *39*, 309-316.
- [60] Ahmed, M., Sadek, M.M., Serrya, R.A., Kafafy, A-H.N., Abouzeid, K. A., Wang, F. Assessment of new anti-HER2 ligands using combined docking, QM/MM scoring and MD simulation. *J. Mol. Graph. Mod.* **2013**, *40*, 91-98.
- [61] Grant, J. A., Pickup, B. T., Nicholls, A. A smooth permittivity function for Poisson-Boltzmann solvation methods. *J. Comput. Chem.* **2001**, *22*, 608-640.
- [62] Kollman, P.A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., Lee, M., Lee, T., Duan, Y., Wang, W., Donini, O., Cieplak, P., Srinivasan, J., Case, D.A., Cheatham, T.E. III. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* **2000**, *33*, 889-897.
- [63] Fogolari, F., Brigo, A., Molinari, H. Protocol for MM/PBSA molecular dynamics simulations of proteins. *Biophys. J.* **2003**, *85*, 159-166.
- [64] Gohlke, H., Kiel, C., Duan, Y., Case, D.A. Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* **2003**, *330*, 891-913.
- [65] Simonson, T., Georgios, A., Karplus, M. Free Energy Simulations Come of Age: Protein-Ligand Recognition. *Acc. Chem. Res.* **2002**, *35*, 430-437.
- [66] 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.

- [67] Lu, N., Kofke, D. A., Woolf, T. B. Improving the Efficiency and Reliability of Free Energy Perturbation Calculations Using Overlap Sampling Methods. *J. Comput. Chem.* **2004**, *25*, 28-39.
- [68] Jorgensen, W. L., Thomas, L. T. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput.* **2008**, *4*, 869-876.
- [69] Zhou, Huan-Xiang; Gilson, Michael K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev.* **2009** *109*, 4092-4107.
- [70] Swanson, Jessica M J., Henchman, Richard H., Andrew McCammon, J. Revisiting free energy calculations: a theoretical connection to mm/pbsa and direct calculation of the association free energy. *Biophys J.* **2004**, *86*, 67-74.
- [71] Khavrutskii, I.V., Wallqvist, A. Computing relative free energies of solvation using single reference thermodynamic integration augmented with hamiltonian replica exchange. *J. Chem. Theory Comput.* **2010**, *6*, 3427-3441.
- [72] Brown, S. P., Muchmore, S.W. Rapid estimation of relative protein-ligand binding affinities using a high-throughput version of mm-pbsa. *J. Chem. Inf. Model.* **2007**, *47*, 1493-1503.
- [73] Michel, J., Verdonk, M.L., Essex, J. W. Protein-Ligand Binding Affinity Predictions by Implicit Solvent Simulations: A Tool for Lead Optimization? *J. Med. Chem.* **2006**, *49*, 7427-7439.
- [74] For reviews, see: *Free Energy Calculations in Rational Drug Design*; Reddy, M. R., Erion, M.D., Eds.; Kluwer/Plenum Press: New York, **2001**.
- [75] Cross, J. B., Thompson, D.C. Rai, B. K., Baber, J.C., Fan, K. Y., Hu, Y., Humblet, C.C. Comparison of Several Molecular Docking Programs: Pose Prediction and Virtual Screening Accuracy. *J. Chem. Inf. Model.* **2009**, *49*, 1455-1474.
- [76] Friesner, R.A., Murphy, R. B., Repasky, M. P., Frye, L.L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., Mainz, D.T. Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein-ligand complexes. *J. Med. Chem.* **2006**, *49*, 6177-6196.
- [77] Amorim, H.L., Caceres, R.A., Netz, P.A. Linear interaction energy (LIE) method in lead discovery and optimization. *Curr. Drug Targets*, **2008**, *9*, 1100-1105.
- [78] Gutierrez-de-Terán, H., Åqvist, J. Linear interaction energy: method and applications in drug design. *Methods Mol. Biol.* **2012**, *819*, 305-323.
- [79] Zhou, R.H., Friesner, R.A., Ghosh, A., Rizzo, R.C., Jorgensen, W.L., Levy, R.M. New linear interaction method for binding affinity calculations using a continuum solvent model. *J. Phys. Chem. B*, **2001**, *105*, 10338-10397.
- [80] Huang, D., Catflisch, A. Efficient evaluation of binding free energy using continuum electrostatics solvation. *J. Med. Chem.*, **2004**, *47*, 5791-5797.
- [81] Carlsson, J., Ander, M., Nervall, M., Åqvist, J. Continuum solvation models in the linear interaction energy method. *J. Phys. Chem. B*, **2006**, *110*, 12034-12041.
- [82] Alonso, H., Blinznyuk, A. A., Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev.* **2006**, *26*, 531-568.
- [83] Amaro, R.E., Baron, R., McCammon, J.A. An improved relaxed complex scheme for receptor flexibility in computer-aided drug design. *J. Comput. Aided Mol. Des.* **2008**, *22*, 693-705.
- [84] Lin, J. H., Perryman, A. L., Schames, J. R., McCammon, J. A. The relaxed complex method: accommodating receptor flexibility for drug design with an improved scoring scheme. *Biopolymers*, **2003**, *68*, 47-62.
- [85] Chodera, J.D., Mobley, D.L., Shirts, M. R., Dixon, R.W., Branson K., Pande V. S. Alchemical free energy methods for drug discovery: progress and challenges. *Curr. Opin. Struct. Biol.*, **2011**, *21*, 150-160.
- [86] Kitchen, D. B., Decornez, H., Furr, J. R., Bajurath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov.* **2004**, *3*, 935-949.
- [87] Taylor, R.D., Jewsbury, P. J., Essex, J. W. A review of protein-small molecule docking methods. *J. Comput. Aided Mol. Des.* **2002**, *16*, 151-166.
- [88] Huang, S-Y., Zou, X. Advances and challenges in protein-ligand docking. *Int. J. Mol. Sci.* **2010**, *11*, 3016-3034.
- [89] Lee, W., Park, H., Lee, S. Molecular dynamics free energy simulation study to rationalize the relative activities of PPAR δ agonists. *Bull. Kor. Chem. Soc.*, **2008**, *29*, 363-371.
- [90] Lu, N., Woolf, T. B. Chapter 6: Understanding and improving free energy calculations in molecular simulations: Error analysis and reduction methods. In: *Free Energy Calculations: Theory & Applications in Chemistry & Biology (Springer Series in Chemical Physics Vol.86)*, Chipot, C., Pohorille, A. (Eds), Springer-Verlag, **2007**, Berlin-Heidelberg.
- [91] Schwab, F., van Gunsteren, W.F., Zagrovi, B. Mechanism for the transport of ammonia within carbamoyl phosphate synthetase determined by molecular dynamics simulations. *Biochemistry* **2008**, *47*, 2935-2944.
- [92] Åqvist, J., Luzhkov, V.B., Brandsdal, B.O. Ligand binding affinities from MD simulations. *Acc. Chem. Res.*, **2002**, *35*, 358-365.
- [93] Gohlke, H., Kiel, C., Case, D.A. Insight into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *Journal of Molecular Biology*, **2003**, *330*, 891-913.
- [94] Olsson, T.S.G., Williams, M.A., Pitt, W.R., Ladbury, J. E. Thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J. Mol. Biol.*, **2008**, *384*, 1002-1017.
- [95] Ma, X.H., Wang, C.X., Li, C.H., Chen, W.Z. A fast empirical approach to binding free energy calculations based on protein interface information. *Prot. Eng.* **2002**, *15*, 677-681.
- [96] Gilson, M.K., Zhou, H.-X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21-42.
- [97] Exner, O. Entropy-enthalpy compensation and anticompensation: solvation and ligand binding. *Chem. Commun.* **2000**, 1655-1656.
- [98] Sharp, K. Entropy-enthalpy compensation: fact or artifact? *Protein Sci.* **2001**, *10*, 661-667.
- [99] Gallicchio E, Levy, R. M. Recent Theoretical and Computational Advances for Modeling Protein-Ligand Binding Affinities. *Adv Protein Chem Struct Biol.* **2011**, *85*, 27-80.
- [100] Ford, D.M. Enthalpy-entropy compensation is not a general feature of weak association. *J. Am. Chem. Soc.* **2005**, *127*, 16167-16170.
- [101] Sharp, K. Entropy-enthalpy compensation: fact or artifact? *Prot. Sci.* **2001**, *10*, 661-667.
- [102] Houk, K. N., Leach, A.G., Kim, S. P., Zhang, X. Binding affinities of host-guest, protein-ligand and protein-transition state complexes. *Angew. Chem., Int. Ed.* **2003**, *42*, 4872-4897.
- [103] Whitesides, G. M., Krishnamurthy, V. M. Designing ligands to bind proteins. *Q. Rev. Biophys.* **2005**, *38*, 385-395.
- [104] Williams, D. H., Stephens, E., O'Brien, D. P., Zhou, M. Understanding noncovalent interactions: ligand binding energy and catalytic efficiency from ligand-induced reductions in motion within receptors and enzymes. *Angew. Chem., Int. Ed.* **2004**, *43*, 6596-6616.
- [105] Muley, L., Baum, B., Smolinski, M., Freindorf, M., Heine, A., Klebe, G., Hangauer, D. Enhancement of hydrophobic interactions and hydrogen bond strength by cooperativity: synthesis, modeling and MD-simulations of a congeneric series of thrombin inhibitors. *J. Med. Chem.* **2010**, *53*, 2126-2135.
- [106] Baum, B., Muley, L., Smolinski, M., Heine, A., Hangauer, D., Klebe G. Non-additivity of functional group contributions in protein-ligand binding: a comprehensive study by crystallography and isothermal titration calorimetry. *J. Mol. Biol.* **2010**, *397*, 1042-1054.
- [107] Scatena, L. F., Brown, M. G., Richmond, G. L. Water at hydrophobic surfaces: weak hydrogen bonding and strong orientation effects. *Science* **2001**, *292*, 908-912.
- [108] Scatena, L. F., Brown, M. G., Richmond, G. L. Water at hydrophobic surfaces: weak hydrogen bonding and strong orientation effects. *Science*, **2001**, *292*, 908-912.
- [109] Dill, K. A., Truskett, T. M., Vlachy, V., Hribar-Lee, B. Modeling water, the hydrophobic effect, and ion solvation. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 173-199.
- [110] Meyer, E. A., Castellano, R. K., Diederich, F. Interactions with aromatic rings in chemical and biological recognition. *Angew. Chem., Int. Ed.* **2003**, *42*, 1210-1250.
- [111] Snyder, P. W., Mecinovic, J., Moustakes, D.T., Thomas, S.W. III, Harder, M., Meck, E. T., Lockett, M. R., Héroux, A., Sherman, W., Whitesides, G. M. Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase. *PNAS*, **2011**, *108*, 17889-17894.
- [112] García-Sosa, A. T. Hydration properties of ligands and drugs in protein binding sites: tightly-bound, bridging water molecules and their effects and consequences on molecular design strategies. *J. Chem. Inf. Mod.* **2013**, *53*, 1388-1405.

- [113] Michel, J., Tirado-Rives, J., Jorgensen, W. L. Prediction of the Water Content in Protein Binding Sites. *J. Phys. Chem. B*, **2009**, *113*, 13337-13346.
- [114] Garcia-Sosa, A. T., Mancera, R. L., Dean, P.M. WaterScore: a novel method for distinguishing between bound and displaceable water molecules in the crystal structure of the binding site of protein-ligand complexes. *J. Mol. Model.* **2003**, *9*, 172-182.
- [115] Amadasi, A., Surface, J. A., Spyraakis, F., Cozzini, P., Mozzarelli, A., Kellogg, G. E. Robust classification of "relevant" water molecules in putative protein binding sites. *J. Med. Chem.* **2008**, *51*, 1063-1067.
- [116] Boehm, H.-J., Boehringer, M., Bur, D., Gmuender, H., Huber, W., Klaus, W., Kostrewa, D., Kuehne, H., Luebbbers, T., Meunier-Keller, N., Mueller, F. Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. *J. Med. Chem.* **2000**, *43*, 2664-2674.
- [117] Garcia-Sosa, A. T., Firth-Clark, S., Mancera, R. L. Including tightly-bound water molecules in de novo drug design. Exemplification through the in silico generation of poly (ADP-ribose)-polymerase ligands. *J. Chem. Inf. Model.* **2005**, *45*, 624-633.
- [118] Vijayan, R., Sahai, M.A., Czajkowski, T., Biggin, P. C. A comparative analysis of the role of water in the binding pockets of ionotropic glutamate receptors. *Phys. Chem. Chem. Phys.* **2010**, *12*, 14057-14066.
- [119] <http://www.desertsci.com/> (accessed September 3, 2013).
- [120] Dodson, G. G., Lane, D. P., Verma, C. S. Molecular simulations of protein dynamics: New windows on mechanisms in biology. *EMBO Reports* **2008**, *9*, 144-150.
- [121] Okazaki, K. I., Takada, S. Dynamic energy landscape view of coupled binding and protein conformational change: Induced-fit versus population-shift mechanisms. *Proc. Natl. Acad. Sci. USA*. **2008**, *105*, 11182-11187.
- [122] Wong, C.F. Flexible ligand-flexible protein docking in protein kinase systems. *Biochim. Biophys. Acta-Proteins and Proteomics* **2008**, *1784*, 244-251.
- [123] Nabuurs, S. B., Wagener, M., De Vlieg, J. A flexible approach to induced fit docking. *J. Med. Chem.* **2007**, *50*, 6507-6518.
- [124] Koska, J., Spassov, V. Z., Maynard, A. J., Yan, L., Austin, N., Flook, P. K., Venkatachalam, C. M. Fully automated molecular mechanics based induced fit protein-ligand docking method. *J. Chem. Inf. Mod.* **2008**, *48*, 1965-1973.
- [125] Sánchez-Moreno, M., Gómez-Contreras, F., Navarro, P., Marín, C., Olmo, F., Yunta, M.J.R., Sanz, A.M., Rosales, M. J., Cano, C., Campayo, L. Phthalazine Derivatives Containing Imidazole Rings Behave as Fe-SOD Inhibitors and Show Remarkable Anti-*T. cruzi* Activity in Immunodeficient-Mouse Mode of Infection. *J. Med. Chem.*, **2012**, *55*, 9900-9913.
- [126] Cornell, W. D., Cieplak, P., Bayly, C.I., Gould, I. R., Merz, K. M. Jr., Ferguson, D. M., Spellmeyer, D.C., Fox, T., Caldwell, J. W., Kollman, P.A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179-5197.
- [127] HyperChem(TM) Professional 8.0, Hypercube, Inc., 1115 NW 4th Street, Gainesville, Florida 32601, USA.
- [128] Miranda, C., Escartí, F., Lamarque, L., Yunta, M. J. R., Navarro, P., García-España, E., Jimeno, M. L. New 1*H*-pyrazole-containing polyamine receptors able to complex L-glutamate in water at physiological pH values. *J. Am. Chem. Soc.* **2004**, *126*, 823-833 and references cited therein.
- [129] Reviriego, F., Navarro, P., García-España, E., Albelda, M. T., Frias, J.C., Domènech, A., Yunta, M. J. R., Costa, R., Ortí, E. Diazatetraester 1*H*-pyrazole crowns as fluorescent chemosensors for AMPH, METH, MDMA (ecstasy), and dopamine. *Org. Lett.* **2008**, *10*, 5099-5102.
- [130] Moro, S., Deflorian, F., Bacilieri, M., Spalluto, G. Ligand-based homology modelling as attractive tool to inspect GPCR structural plasticity. *Curr. Pharm. Des.* **2006**, *12*, 2175-2185.
- [131] Colotta, V., Catarzi, D., Varano, F., Capelli, F., Lenzi, O., Filacchioni, G., Martini, C., Trincavelli, L., Ciampi, O., Pugliese, A.M., Pedata, F., Schiesaro, A., Morizzo, E., Moro, S. New 2-Arylpyrazolo 3, 4-quinoline Derivatives as Potent and Selective Human A3 Adenosine Receptor Antagonists. Synthesis, Pharmacological Evaluation, and Ligand-Receptor Modeling Studies. *J. Med. Chem.* **2007**, *50*, 4061-4074.