American Journal of Modeling and Optimization, 2016, Vol. 4, No. 3, 74-114
Available online at http://pubs.sciepub.com/ajmo/4/3/2
©Science and Education Publishing
DOI:10.12691/ajmo-4-3-2

SciEP Science & Education Publishing

Docking and Ligand Binding Affinity: Uses and Pitfalls

María J. R. Yunta*

Departamento de Química Orgánica I, Facultad de Química, Universidad Complutense, 28040 Madrid, Spain
*Corresponding author: mjryun@ucm.es

Abstract In this review article, we will explore the foundations of different classes of docking and scoring functions, their possible limitations, and their suitable application domains. We also provide assessments of several scoring functions on weakly-interacting protein-ligand complexes, which will be useful information in computational fragment-based drug design or virtual screening.

Keywords: *molecular modeling, docking, binding affinity, drug scoring, computer aided drug design*

Cite This Article: María J. R. Yunta, "Docking and Ligand Binding Affinity: Uses and Pitfalls." *American Journal of Modeling and Optimization*, vol. 4, no. 3 (2016): 74-114. doi: 10.12691/ajmo-4-3-2.

1. Introduction

Molecular recognition phenomena involving the association, usually by noncovalent interactions, of ligands to macromolecules with high affinity and specificity play a key role in biology [1,2,3]. Noncovalent interactions are a number of relatively weak chemical interactions that stabilize the conformations and the interactions between molecules. Noncovalent interactions are abundant in nature [4,5] and are very important in many research areas such as chemistry, [5,6,7,8,9] biology, [10] biochemistry, [5,11] molecular recognition, [12] drug design, [13,14,15] materials science, [16,17,18] and beyond.

Although in medicinal chemistry the fundamental microscopic interactions giving rise to bimolecular association are relatively well understood, designing computational schemes to accurately calculate absolute binding free energies remains very challenging, being one of the main objectives of structure-based drug design to reliably and accurately predict the binding affinity of compounds that bind to a target protein.

Computational approaches currently used for screening large databases of compounds to identify potential lead drug molecules must rely on very simplified approximations to achieve the needed computational efficiency [19]. Nonetheless, the calculated free energies ought to be very accurate to have any predictive value. Furthermore, the importance of solvation in scoring ligands in molecular docking has been stressed previously [20], as water plays an important role in the formation of protein-ligand complexes by forcing hydrophobic groups together abolishing disruptive effects on the hydrogen bonded network in water, which is also known as the hydrophobic effect. Before a protein-ligand complex is formed, the individual partners that are not a part of hydrophobic surface are involved in hydrogen bonds with the surrounding water. Once the complex is formed, these hydrogen bonds are replaced with hydrogen bonds between the ligand and the protein. The contribution of hydrophobic interactions to protein-ligand binding is normally regarded to be proportional to the size of the hydrophobic surface buried during complex formation [21,22]. Hydrophobic interactions are also regarded to be the main driving force of conformational change of the receptor upon ligand binding [23].

It has also been suggested that the electrostatic interactions mainly determines molecular recognition and noncovalent binding [23], but this is by no means a general rule, as there are equal evidence to the importance of shape complementarity [24]. Molecular recognition can be therefore attributed to contributions of electrostatic and van der Waals interactions, solvation/desolvation and flexibility of ligand and protein. In this context, there can be hydrogen bonds resulting from electrostatic attraction between an electronegative atom and hydrogen that is connected to an electronegative atom, which is usually oxygen or nitrogen and less frequently fluorine, or some π-π-interactions, or stacking. This also implies that it is very important for theoretical calculations to have the protonation states of arginine, lysine, aspartic and glutamic acids, as well as histidine correctly determined for an accurate description of electrostatic interactions. Distances of hydrogen bonds are normally 2.5-3.2 Å and angles of 130º-180º are typically found [25]. The strength of a hydrogen bond depends on its directionality and its surroundings. The hydrogen bonds in the interior of proteins are stronger than the ones in the solvent-exposed regions [26]. In addition ionic bonds are very important to ligand-protein binding, but their strength is considerably reduced in water due to shielding.

The van der Waals interactions, or London dispersion forces are used to model the attractive and repulsive forces between molecules. If two atoms are too close to each other they will repel each other, which make it possible to define a fixed radius for the "size" of each atom (van der Waals radius). The contact distance between two atoms is then the sum of their van der Waals radii. Van der Waals interactions can be very important when two surfaces of molecules fit well together.

At the same time, the change in the degrees of freedom of the ligand and protein upon binding results in a change of the entropy [27]. It can be viewed as the ligand and protein both losing three degrees of translational and rotational freedom, while six new vibrational degrees of freedom are created for the complex [28].

One of the especially important directions is the chemistry of large systems with noncovalently interacting constituents such as biomolecular complexes, [29] host−guest complexes, [30,31,32] molecule−surface interactions, [33,34,35] or behavior of large molecular ensembles. [36,37] These applications motivate the development of scalable and robust computational methods [38,39,40,41,42] that can simulate complex processes for large numbers of atoms. These methods typically rely on approximations that are parametrized either by using experimental data, or, more recently, by reference quality calculations.

Due to their technological and fundamental importance, noncovalent interactions are studied very extensively using a broad range of experimental, theoretical, and computational approaches. [16,43,44] In particular, theory and computations are indispensable not only for understanding, interpretation, and validation of experimental measurements but also for obtaining information that is complementary to what is accessible in experiments. [45] Noncovalent interaction studies therefore often combine experiments with theory [46,47] in order to gain more comprehensive insights and deeper fundamental understanding.

There are many distinct types of noncovalent interactions that differ in their origin or nature of the interacting species. Hydrogen bonding and stacking (or π−π interaction) are perhaps the most studied as they play important roles in the structural biology of nucleic acids and proteins. In addition, many more interaction patterns were identified over the years, such as halogen bond, sigma hole interaction, blue-shifting hydrogen bond, dihydrogen bond, or anion/cation-π interaction to name just a few. For a more complete classification of noncovalent interactions see, for example, ref [45]. From the theoretical viewpoint, noncovalent interactions are best understood in terms of electrostatic, induction (or polarization), dispersion, and exchange-repulsion components, whose balance determines the total intermolecular interaction potential. [48] Electrostatic interactions originate from the classical Coulomb interaction of the monomer electron distributions (unperturbed by the interaction). Induction is the change in the electrostatic interaction due to polarization of the monomer charge density by the interacting molecules. Dispersion arises from the interaction of the instantaneous fluctuations of the electronic density and the multipoles induced by this fluctuation. At short distance, the attractive forces are opposed by the repulsive exchange repulsion due to the Pauli principle. Consequently, the dispersion is essentially a correlation effect, and it belongs to the most difficult cases for accurate description by the basis set quantum chemical approaches. The mentioned four components are well-defined within the framework of symmetry-adapted perturbation theory (SAPT, [49]) and provide a solid and insightful background for analysis of intermolecular interactions. [50] Note, however, that most benchmark quantum chemistry methods provide only the total interaction energies, and it is difficult to decompose them to the mentioned basic components.

2. Goals of Ligand Docking and Binding Studies

Molecular docking is a key tool in structural molecular biology and computer-assisted drug design as it tries to predict the structure of the intermolecular complex formed between two or more constituent molecules, trying to predict the position and orientation of a ligand (a small molecule) when it is bound to a protein in order to know the predominant binding mode(s) of a ligand with a protein of known three-dimensional structure. Successful docking methods search high-dimensional spaces effectively and use a scoring function that correctly ranks candidate dockings. Docking can be used to perform virtual screening on large libraries of compounds, rank the results, and propose structural hypotheses of how the ligands inhibit the target, which is invaluable in lead optimization. The setting up of the input structures for the docking is just as important as the docking itself, and analyzing the results of stochastic search methods can sometimes be unclear.

Many approaches to studying protein-ligand interactions by computational docking are currently available. As stated upwards, given the structures of a protein and a ligand, the ultimate goal of all docking methods is to predict the structure of the resulting complex and to predict the biological activity of a given ligand. This requires a suitable representation of molecular structures and properties, search algorithms to efficiently scan the configuration space for favorable interaction geometries, and accurate scoring functions to evaluate and rank the generated orientations. For many of the available methods, tests on experimentally known complexes have appeared in the literature and some of them have been used in predictive studies on antibody-ligand interactions to provide structural insights where adequate experimental information is missing. For example, computational investigations have applied molecular docking and molecular dynamics (MD) simulations to comprehend the binding properties of gp120 and its interaction with the CD4 receptor and co-receptor in the immune cell [51] and a mathematical model that describes the binding of HIV-1 virus to T cells has also been developed to determine the analytical thresholds for the dosage and dosing interval of HIV fusion inhibitor enfuvirtide [52].

There are two different problems in this task: the posing and the scoring. In the first case, the process of determining whether a given conformation and orientation of a ligand fits the active site is usually a fuzzy procedure that returns many alternative results. And in the second case, the pose score is a measure of the fit of a ligand into de active site. Scoring during posing phase usually involves simple energy calculations (electrostatic, van der Waals, ligand strain). Further re-scoring might attempt to estimate more accurately the free energy of binding (ΔG, and therefore K_A) perhaps including properties such as entropy and solvation.

The predicted binding and docked energies are the sum of the intermolecular energy and the torsional free-energy penalty, and the docking ligand's internal energy, respectively, and the inhibition constant (K_i) is usually calculated as follows:

$$K_i = \exp(\Delta G\ x1000) / (R_{cal}\ x\ TK) \quad (1)$$

where ΔG is the docking energy, R_{cal} is 1.98719, and TK is 298. [53,55,56,57]

For that reason, docking calculations currently have limited success beyond the lead identification stage, where more accurate lower-throughput computational methods are needed. In this regard, the Molecular Mechanics/Generalized Born Surface Area (MMGBSA) and Molecular Mechanics/Poisson-Boltzmann Surface Area (MM-PBSA) methods calculate binding free energies using molecular mechanics (force fields) and continuum (implicit) solvation models [58]. They have been successfully applied across a range of targets and are implemented in software programs such as Amber [59], Delphi [60] and Schrödinger [61] among others.

In any case, computational modeling has become a powerful tool in understanding detailed protein-ligand interactions at molecular level and in rational drug design. To study the binding of a protein with multiple molecular species of a ligand, one must accurately determine both the relative free energies of all of the molecular species in solution and the corresponding microscopic binding free energies for all of the molecular species binding with the protein. In this paper, we aim to provide a brief overview of the recent development in computational modeling of the solvent effects on the detailed protein-ligand interactions involving multiple molecular species of a ligand related to rational drug design. In particular, we first briefly discuss the main challenges in computational modeling of the detailed protein-ligand interactions involving the multiple molecular species and then focus on the FPCM model and its applications. The FPCM method allows accurate determination of the solvent effects in the first-principles quantum mechanism (QM) calculations on molecules in solution. The combined use of the FPCM-based QM calculations and other computational modeling and simulations enables us to accurately account for a protein binding with multiple molecular species of a ligand in solution. Based on the computational modeling of the detailed protein-ligand interactions, possible new drugs may be designed rationally as either small-molecule ligands of the protein or engineered proteins that bind/metabolize the ligand. The computational drug design has successfully led to discovery and development of promising drugs. [62]

Nowadays, molecular docking plays an important role in drug design and discovery with the universal application of docking programs, such as Glide, [63] Autodock, [64] FlexX, [65] and GOLD. [66] When these programs are utilized prior to experimental screening, they are usually considered as powerful computational filters to reduce labor and cost. All of these docking programs explore various docked conformations and determine the tightness of interactions between the protein and the ligand, but the performance on predicting the experimentally observed binding poses is not always satisfying. As is widely accepted, the real bottleneck on obtaining the reliable docking result lies in the scoring functions. [67-72] As a matter of fact, considerable efforts have been devoted to the development of approximate computational methods for describing protein−ligand interactions more accurately, but it still lacks a universal scoring function which works reliably for all or most of protein−ligand systems. [73,74] For some particular protein−ligand systems, most of the widely used docking programs are incapable of predicting the correct binding modes, imposing great challenge on the effectiveness of computer-aided drug design. Therefore, improved methods for predicting protein−ligand binding affinities are desperately needed.

Among the approximate methods, the molecular mechanics/Poisson−Boltzmann surface area (MM/PBSA) approach is attractive because it does not contain any parameters that vary for different protein−ligand systems and it involves a set of physically well-defined energy terms. [58,75-80] The validity of such an approach has been explored in previous studies. [75,81,82] In particular, Kuhn et al. validated the MM/PBSA approach on different biological systems by putting forward the idea of using single-minimized structure instead of MD trajectories. [83] Moreover, Hou et al. systematically evaluated the performance of MM/PBSA on predicting the absolute binding affinity for protein−ligand complexes and the accuracy of identifying the correct binding poses generated from molecular docking programs. [84,85] The accuracy of MM/PBSA approach for predicting protein−ligand binding affinity relies on the accuracy of force field, in addition to other factors. It is known that current nonpolarizable force fields, for example, CHARMM and AMBER, often fail to give accurate representation of the electrostatics of the specific protein environment, which is highly inhomogeneous and protein-specific. Recently, polarized protein-specific charges (PPC) based on a fragmentation scheme [86,87,88] for electronic structure calculation of biomolecules and the continuum dielectric model for the solvent in a selfconsistent fashion was developed. [89] Since PPC correctly describes the polarized electrostatic state of a protein at a given structure, it is able to give a more accurate description of the mutual electrostatic polarization effect for protein−ligand binding, resulting in better description of electrostatic interactions between protein and ligand. It has been demonstrated in a number of applications that PPC gives significantly better agreement with experimental data than standard nonpolarizable force fields in protein−ligand binding affinity calculations using MM/PBSA. [90] The effect of bridging water molecules between the protein and ligand attracts more and more attention recently. These water molecules are considered to play an important role in mediating the interaction between protein and ligand. [67,91-98] While only a few scoring functions explicitly take the water mediated protein−ligand interactions into consideration, [99,100,101,102] explicitly including the bridging water molecules in molecular docking and scoring function may be crucial for correctly predicting the binding poses.

Although the influence of water on the stabilization of drug-receptor complexes is well known (the hydrophobic effect) [103], the effect of water on binding kinetics has only recently been recognized [104-111]. At small length scales, on the order of several angstroms, the motion of a few water molecules can be enough to influence binding kinetics. Using a combination of experiment and computer simulations, Schmidtke and co-workers showed that,

when a ligand and a receptor interact via hydrogen bonds shielded from water by surrounding hydrophobic regions, the resulting complex tends to be more kinetically stable than if the hydrogen bonds were less shielded [111] (Figure 1). The difficulty with which water diffuses into and away from these largely hydrophobic sites appears to create a kinetic barrier to ligand binding and unbinding. At larger length scales involving nanometer-scale volumes of water, collective water motion out of a hydrophobic region, or 'dewetting', can present a barrier to drug entry [107,110,112].

Recently, Setny and co-workers explicitly demonstrated the existence of a dewetting barrier to ligand binding in computer simulations of a model system [107]. As the ligand approached the receptor a barrier arose between a wet and dry binding pocket. Surmounting this dewetting barrier presented the major bottleneck to ligand entry. In simulations of beta blockers binding to b-adrenergic receptors we observed a qualitatively similar phenomenon where entry of the hydrophobic ligand into a hydrophobic extracellular vestibule was correlated with the collective evacuation of water from that site and from around the ligand [110]. This dehydration step corresponded to the largest energetic barrier along the drug binding pathway.

By their very nature, computational methods, in particular molecular dynamics simulations, provide detailed structural information on metastable intermediate states and transition states, at atomic spatial and femtosecond temporal resolution [113]. Owing to increases in computational power, it has recently become possible to simulate the full process of spontaneous ligand–receptor association — which typically occurs on the microsecond timescale — in atomic detail, providing direct access to detailed information on binding mechanisms that have been difficult to access experimentally [106,110,114,115]. In recent work from our group, molecular dynamics simulations of the spontaneous binding of several drug molecules to kinases and GPCRs achieved bound poses virtually identical to the crystallographically determined bound structures. Estimates of on-rates from simulation were also in approximate agreement with experimental measurements [110,115]. Although the physicochemical models underlying molecular dynamics simulations remain imperfect, these and other studies demonstrate the beneficial use of such simulations in probing drug binding pathways.

Various other computational methods, ranging from coarse grained molecular dynamics simulations [116,117,118] to biased enhanced-sampling simulations [119-124], have also been used to characterize binding pathways. Because ligand dissociation is slower — often taking seconds to hours — it can usually be observed computationally only by use of these latter techniques. It is important to note, however, that in the absence of external driving forces the unbinding process is the reverse of the binding process, following the same pathway and traversing the same barriers in the opposite order.

The residence time of a ligand−protein complex is a crucial aspect in determining biological effect in vivo. Despite its importance, the prediction of ligand k_{off} still remains challenging for modern computational chemistry. We have developed aMetaD, a fast and generally applicable computational protocol to predict ligand−protein unbinding events using a molecular dynamics (MD) method based on adiabatic-bias MD and metadynamics. This physics-based, fully flexible, and pose-dependent ligand scoring function evaluates the maximum energy (RTscore) required to move the ligand from the bound-state energy basin to the next. Unbinding trajectories are automatically analyzed and translated into atomic solvation factor (SF) values representing the water dynamics during the unbinding event. This novel computational protocol was initially tested on two M3 muscarinic receptor and two adenosine A2A receptor antagonists and then evaluated on a test set of 12 CRF1R ligands. The resulting RTscores were used successfully to classify ligands with different residence times. Additionally, the SF analysis was used to detect key differences in the degree of accessibility to water molecules during the predicted ligand unbinding events. The protocol provides actionable working hypotheses that are applicable in a drug discovery program for the rational optimization of ligand binding kinetics. [125]

Residence time can also be modulated by leveraging water dynamics. Increasing the number of shielded hydrogen bonds, or accentuating the hydrophobic shielding of existing hydrogen bonds by designing a broader ligand, could tend to increase residence time (Figure 1). At larger length scales where the driving force of drug binding is controlled by dewetting, adding hydrophobic groups to the ligand might lower the dewetting barrier [107,112].

With experimental methods for determining drug binding kinetics becoming faster and less expensive, the availability of such data will surely become more widespread and the drive to incorporate it into drug discovery programs will increase. A greater understanding of the molecular determinants of binding kinetics will be crucial for maximizing the impact kinetics data has on drug discovery.

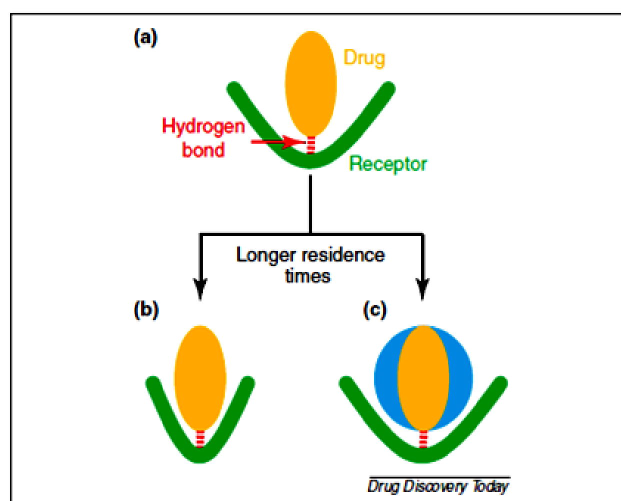

Figure 1. Shielded hydrogen bonds confer longer residence time. (a) Schematic of a drug (yellow) bound to a receptor (green), forming, among other interactions, a hydrogen bond (red dashed line). (b) Compared with (a) the greater curvature of the binding site shields the hydrogen bond from water access, creating a larger barrier to drug dissociation. (c) For a less-curved binding site, as in (a), increasing the ligand size (blue) also shields the hydrogen bond. The difficulty with which water diffuses into and away from a shielded hydrogen bond directly impacts drug residence time, by creating a kinetic barrier to ligand binding and unbinding [111]

A more nuanced route to modulating residence times could involve changing interactions between the drug and those fluctuating parts of the receptor that often appear to be the bottlenecks to drug binding. In these cases, small, rational changes to the drug — for instance adding a group to make a specific interaction, as in the extra methyl group of the InhA inhibitor PT70 [126], or removing a group to break a specific interaction — could result in large changes in residence time. Notably, the atomic groups modified in this way need not be those that confer (the bulk of) binding affinity; in the same way that solubilizing groups can often be added, in an almost orthogonal manner, to a drug binding-core template. Such a design strategy is also promising from the viewpoint of subtype selectivity, especially for receptors in which the binding site is well conserved among subtypes.

3. Ligand Docking.

For a docking process to be successful, it is necessary that both the right conformation of the ligand–receptor complex is predicted, and that the ranking of final structures is correct. The procedure needs to be able to differentiate among similar conformations of the same system, as well as to predict the relative stability of different complexes. There are several different scoring functions for this purpose (for recent comparisons of scoring functions see [67,72,127]). As most contain empirically fitted parameters, their performance on any particular problem will depend on the set of structures used for the calibration. So far, no scoring function has proven to be reliable for every docking case tested. The main constraint on their improvement rests with the need for speed; when ranking hundreds, if not thousands, of complexes a compromise in accuracy must be made. Knowledge-based functions used in the ranking of molecular interactions may not be general and accurate enough, because of the limited number of interactions that can be inferred from crystal structures and the inadequate description of repulsive forces. MM based functions, on the other hand, inherit all common problems of molecular mechanics parameters, and recent calculations have shown that they may result in large electrostatic errors. [128,129,130] Several pilot studies on the use of semi-empirical quantum mechanical methods for a more accurate description of the interactions of proteins with small ligands have been recently published. [131,132,133] Taking account of these factors, the type of scoring functions currently implemented in docking programs cannot be expected to distinguish energetically between close conformations of the same molecule, or even to rank properly a group of ligands of similar activity. Although the combination of several scoring functions into a consensus score has been shown to provide better results, [73,134,135,136] this merely produces a ranking of complexes without offering final energies.

When receptor flexibility is included during the docking process, the risks associated with inadequate conformation of the protein target are reduced. [137,138,139] Although originally restricted to the docking of rigid ligands into rigid receptors, recent advances in docking algorithms have allowed incorporation of ligand flexibility and, to less extent, protein mobility, during the docking procedure. Most modern algorithms account for ligand flexibility; this can be addressed by systematic methods (i.e., incremental search), stochastic methods (i.e., Monte Carlo simulation), and deterministic search (i.e., MD simulation). [140] Programs that incorporate protein receptor flexibility, at least partially, began to appear more recently. [66,141,142,143] The size and complexity of proteins makes it difficult to fully account for their mobility during a docking process and, therefore, its treatment is usually restricted to selected residues.

Docking the ligand against each protein structure in the ensemble constitutes the most comprehensive, although expensive, approach. While this strategy is not a realistic option for the virtual screening of a large library, it is a valid approach for difficult docking problems where even minor conformational changes of the receptor are expected to have a major influence on the binding process. Carlson et al. [144] developed ‘‘dynamic’’ pharmacophore models of HIV-1 integrase using several snapshots from an MD simulation.

Although computationally expensive, docking against individual protein structures has proven to be effective not only in finding the correct docking pose* within a flexible receptor (both in evaluative and predictive contexts), but has been found useful also for discovering alternative binding modes otherwise not apparent from the rigid picture of proteins extracted from crystal structures. This method can have important applications in lead optimization and refinement, despite not being useful for the virtual screening of large libraries. Inclusion of protein flexibility does not necessarily lead to improvements in the final docking results. Increased capacity of the receptor to accommodate several ligand conformations may lead to the generation of very similar complexes not distinguishable by modern scoring functions. Therefore, the validity of the final predictions should be assessed experimentally.

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 time course 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. [145]

While knowledge of the relative stability of different complexes may be an adequate result for an initial screening protocol, estimates of the absolute binding free energy may be necessary in later stages of docking or during lead refinement, when only few selected ligands remain. If stringent rankings or accurate energies are needed, different MD-based calculations can be carried out on the final complexes to estimate the binding free energy. [58,77,146-155] Thermodynamic integration (TI) and free energy perturbation (FEP) are among the most

rigorous methods currently available for the calculation of free energies. Despite providing very accurate free energies, they are not widely applied as they are computationally expensive. [152,154,155] The main limitation of these approaches is the exhaustive conformational sampling required to obtained a proper averaged ensemble, and their slow convergence. Inefficiencies in configurational sampling because of the appearance/disappearance of atoms (explained in more detail below) restrict their use to small transformations, and limit analysis to a few closely related compounds. One of the most important limitations in free energy calculations is the sampling of the conformational space. [156] Exploration of the appropriate conformations is not guaranteed simply by longer simulations. To avoid convergence problems and inadequate sampling during the simulations, only transformations between similar molecules are feasible, constraining the type of ligands that can be compared. This, together with the computational cost of such approaches, has prevented the wide application of FEP for determining binding free energies, despite its accuracy.

Recently developed approaches that provide relatively good energy values at a moderate cost include MD-based methods such as the linear interaction energy (LIE) method, [147,157,158,159,160] and the so called MM-PBSA method. [161]. Aqvist et al. [147] introduced the LIE semi-empirical MD approach for the estimation of binding free energies. [153,162] This method assumes that the binding free energy can be extracted from simulations of the free and bound state of the ligand. The energy is divided into electrostatic and van der Waals components, and the final binding energy is calculated as:

$$\Delta G_{bind} = \alpha V_{bound}^{elec} - V_{free}^{elec} + + \beta V_{bound}^{vdW} - V_{free}^{vdW} + \gamma \quad (2)$$

where $\langle V_{bound}^{elec} - V_{free}^{elec} \rangle$ represents the averaged change in electrostatic energy and $\langle V_{bound}^{vdW} - V_{free}^{vdW} \rangle$ the averaged change in van der Waals energy in going from an aqueous solution to a protein environment. α, β, and γ are empirically determined constants. Two different MD simulations, one for the ligand bound to the protein and another for the free ligand in water, are used to calculate the energies. During the early applications of the LIE approach, only two coefficients, α and β, were considered. Although α, the electrostatic coefficient, appeared to have a constant value of 0.5 for several protein systems, as predicted by the linear response approximation, [147] the van der Waals coefficient, β, seemed to adopt various values depending on the characteristics of the protein receptor. [157,158,163,164] Kollman and co-workers [160] suggested that the value of β depended on the hydrophobicity of the binding site, and that it could be predicted by calculating the weighted desolvation non-polar ratio (WDNR) of the system. Jorgensen's group extended the method to calculate both the hydration and binding free energy, adding a new term to account for the solvent accessible surface and scaling it by a new empirical coefficient. [150,165,166] It was later found, however, that the non-polar component γ, although considered zero in many cases, [147,148,159] could adopt different values [167] and account for the variability earlier assigned to β. In a recent study, Aqvist and coworkers [168] performed a systematic analysis of several ligands in complex with P450cam. Using fixed values for α and β, while optimizing γ, not only provided the best absolute binding free energies for the ligands but also showed that the coefficients of the LIE method are independent of the force field used and that only γ might need to be optimized to account for the hydrophobicity of the active site.

Do MD simulations after docking add any value to results obtained? In terms of structure optimization, MD simulations allow flexibility for both the ligand and protein receptor, facilitating the relaxation of the complete system and accounting for induced-fit effects. The effect of solvent molecules can also be treated explicitly; with the incorporation of water molecules in the simulated system, important stabilizing/destabilizing effects and water-mediated interactions can be observed. Furthermore, the time-dependent evolution of the system during the simulation provides a dynamic picture of the complex and helps to discriminate the correctly docked conformations from the unstable ones.

With respect to free energy calculations, it has pointed out that scoring functions implemented within docking programs are not sufficiently accurate to identify, in every case, the most stable conformation of a given ligand or drug with the highest binding affinity among a set of compounds. Although library-screening processes require fast and inexpensive scoring functions, more accurate and expensive calculations can be employed in the last stages of a docking process, when only a few possible candidates are left, or during lead optimization. MD-based methods are among the most accurate current techniques available for the calculation of free energies. FEP and the more recent LIE and MM-PBSA approaches have been used successfully to predict both relative and absolute binding free energies of many different complexes with errors of chemical accuracy, that is, 1–2 kcal/mol. [145]

Although the physics predicts that a properly set up MD simulation of a solvated protein and its unbound ligand would eventually lead to the formation of the most stable protein–ligand complex, no current simulation protocol can deal with the long time span required for a binding process to occur for such a large and complex system. In addition to time restrictions, the inherent tendency of an MD simulated system to get trapped in local minima (sampling problem) makes the use of ordinary MD simulations as docking techniques infeasible.

To improve the exploration of the free energy landscape and reproduce a possible binding event within feasible computation times, it is necessary to increase the sampling power of conventional MD simulations. [169,170] The two basic approaches developed comprise flattening of the energy surface, which allows the system to overcome large energy barriers, and the simulation of several copies of the system. Multiple-copy simultaneous search [171] and locally enhanced sampling [137] are examples of the latter approach. Alternatively, in replica exchange MD (REMD), [138] several noninteracting replicas of the same system are simulated at different temperatures. At specified intervals, replicas can exchange temperature, thereby overcoming energy barriers when simulated at higher temperatures. Thermal heating of selected components of the system can also be used to selectively enhance the sampling of certain regions, while the barriers

separating local minima can be lowered using the local elevation approach [172] or the conformational flooding technique. [173] These approaches, although faster than conventional MD simulations, are still much slower than typical docking techniques and could not be applied to more than a few examples.

While atomistic MD simulations are still computationally expensive for docking large libraries of compounds, their application for hit discovery and optimization is increasing steadily. Recently, MD simulations have been employed for determining the binding modes of small aliphatic and aromatic molecules into the oncoprotein BCL-6 [174] and isopropyl alcohol into five different proteins [175]. These studies were published 18 years after the minimization-based MCSS protocol of Miranker and Karplus, which in principle allowed also for binding site flexibility by a combination of MD and energy minimization. The MD protocol developed recently in MacKerell's group is called SILCS (Site-Identification by Ligand Competitive Saturation), and as in MCSS the attractive interactions between fragments are switched off. This simulation stratagem makes possible the use of a very high concentration even for hydrophobic fragments, which would otherwise aggregate in the simulation box [176]. We note en passant that this is an interesting example in which a simulation protocol allows one to study a molecular system under conditions that are not accessible by experiments.

Parrinello and co-workers [177] employed a new MD method, metadynamics [178], to find the correct conformation of ligands inside flexible receptors in aqueous solution. A metadynamics run is a standard MD simulation that implements harmonic restraints on certain collective variables (e.g., the distance from the ligand to the binding site), which are explored along a time scale. A potential term, constructed using a sum of Gaussians, prevents the system from re-visiting configurations, so that the system is forced to move around the conformational space. One of the novelties of this approach is that the free energy surface explored during the simulations can be reconstructed from the added Gaussians, and the docking energy can be determined. Four different systems were analyzed, and although the correct geometry was found and the experimental binding energy was predicted within 1 kcal/mol in all cases, most of the calculations started from the crystal configuration, with the ligand already bound within the active site. In only one case, b-trypsin with a small and almost rigid ligand (benzamidine), was simulation of the ligand entering the enzyme presented. Therefore, despite its demonstrated ability to reproduce binding energies and provide a free energy surface as a function of the collective variables, the utility of the method as a predictive docking tool to find the correct binding mode of a free mobile ligand entering its protein receptor remains to be properly tested.

4. Ligand Binding Affinity

Binding reactions are ubiquitous in biology. For example, any substrate needs to bind to its enzyme to be converted to the product, and it can be argued that the activation energy is the difference in binding energy of the substrate and the rate-limiting transition state. However, the arguably most important type of binding reaction is the association of a drug candidate to its target receptor. It is the prime aim of drug development to find a small molecule that binds strongly to a certain biomacromolecule. Moreover, it is also important that the drug candidate does not bind to other, often similar, macromolecules, so that it does not interfere with other key functions in the body, and that it has proper transport, metabolism, and excretion properties, which often are governed by the binding to other biomacromolecules, e.g., transporters and metabolic enzymes. Therefore, the study of binding affinities is of immense interest in pharmaceutical chemistry, and the development of a new drug typically involves the synthesis and test of the binding of thousands of drug candidates. Naturally, it would be of great gain if binding affinities could be estimated fast and accurately by computational methods.

Protein−ligand binding is essential to almost all biological processes. The underlying physical and chemical interactions determine the specific biological recognition at the molecular level. The essential element in drug discovery is to find a molecular ligand that either inhibits or activates a specific target protein through ligand binding. However, finding a ligand that binds a targeted protein with high affinity is a major challenge in early stage drug discovery. Modern technological advances in analytical methods and the availability of experimental tools such as X-ray crystallography and nuclear magnetic resonance (NMR) [179,180] have enabled researchers to obtain atomic resolution structures of protein−ligand complexes. The high-resolution structures of protein and protein−ligand complex provide a chemical basis for understanding protein−ligand interactions at atomic level, [23,181,182,183,184] and they can be effectively used as the basis for the design of small-molecule drugs for the treatment of diseases.

However, given the structure of a protein−ligand complex (such as from experiment or virtual molecular docking), it is not an easy task to calculate its binding affinity reliably, an extremely important but difficult undertaking in computational biology. The strength of binding of a ligand to a protein molecule is governed by the free energy change in the binding process. Besides the accuracy of force field and sufficient sampling of the phase space during molecular simulation, reliable calculation of entropy change is critical to the accuracy of the computed binding free energy. Currently, the most rigorous approaches for accurate calculation of protein−ligand binding free energy are free energy perturbation (FEP) [186-191] and thermodynamic integration [154,192] methods. However, free energy calculations for protein−ligand binding using either FEP (free energy perturbation) or TI (thermodynamic integration) methods are extremely difficult; both can be prohibitively expensive and very difficult to converge numerically as one has to simulate many nonphysical intermediate states of the system. The linear interaction energy (LIE) approach is another class of methods in which the interaction energies are used with adjustable parameters to estimate protein−ligand binding free energies. [162,164] This class of methods often do well for systems with similar interaction characteristics. In contrast, the MM/PBSA approach, [58,60,83,194,195,196,197] which

uses an implicit solvent model to compute solvation energy coupled with MD simulation in explicit water to obtain gas-phase component of the binding free energy, is more general for practical applications in computing binding free energies. However, a major problem in MM/PBSA method is the calculation of entropy change in protein−ligand binding. The current MM/PBSA approach calculates entropy change for protein−ligand binding by using the standard normal mode method, which is approximate in nature, extremely expensive in computation, and often unreliable for protein−ligand binding. As a result, many applications using MM/PBSA approach simply neglect the calculation of entropy change for protein−ligand binding and thus render the computed free energy even more uncertain.

In this report, we present a novel and conceptually more intuitive theoretical paradigm called “interaction entropy” or IE. This new paradigm introduces a novel but more intuitive conceptual understanding of the entropic effect in protein−ligand binding and other general interaction systems as well as a practical method for highly efficient calculation of its effect. This interaction entropy is theoretically rigorous and can be directly obtained from MD simulation of protein−ligand system without any extra computational cost. Thus, the new method is numerically superefficient compared to the normal mode calculation of entropy for protein−ligand binding. For free energy calculation of protein−ligand binding, we can simply employ the standard MM/PBSA method to calculate the solvation free energy component and then combine them with the calculated interaction entropy to obtain the binding free energy. Thus, the interaction entropy method is straightforward to implement and highly efficient to apply for practical computation of protein−ligand binding free energies. To fully demonstrate the efficiency and reliability of the present approach, we carried out computational studies for 15 randomly selected protein−ligand complexes with experimental binding affinities using both the interaction entropy method as well as the standard normal mode method for entropy calculations.

Consequently, numerous methods have been developed with this aim. [23] Most computational methods are based on some sort of energy function. It can be developed either by a statistical analysis of experimentally characterized ligand−receptor complexes or from a physical description of the interactions. Statistical energy functions can come from an analysis of atom−atom distances, converted to an empirical potential of mean force (knowledge-based scoring functions), or from a regression analysis of binding affinities and a collection of terms that are believed to be important for the binding affinity, e.g., hydrogen bonds, ionic interactions, metal bonding, desolvation, hydrophobic effects, stacking, etc. (empirical scoring functions). Physics-based energy functions are typically in the form of a molecular-mechanics (MM) force field that contains terms for the stretching of bonds, bending of angles, rotation of torsion angles, Coulombic interaction between atomic partial charges, and van der Waals attraction (dispersion) and exchange repulsion.

Likewise, many approaches have been used to predict the structure of the ligand−receptor complex and estimate the binding affinity using these energy functions. The most commonly used one is to change the structure until a minimum energy is obtained, i.e., a geometry optimization. This is a formidable task for a biomacromolecule, because the potential-energy surface is extremely complicated with essentially an infinite number of local minima. This is often solved by keeping the macromolecule fixed, excluding the solvent molecules, running many calculations from different starting points, or employing special algorithms (e.g., genetic algorithms) that try to find the global minimum. In their simplest form, such docking calculations can estimate the binding affinity within seconds, often using knowledge-based or empirical scoring functions. They can often predict structures close to the experimentally determined geometry of the complex, but they have problems distinguishing them from other poses and predicting accurate binding affinities for a diverse set of targets. [198,199]

Usually the binding of a small molecule to a biological macromolecule, e.g., a protein or a nucleic acid, is governed by noncovalent interactions, i.e., the reaction is:

$$\mathrm{R} + \mathrm{L} \rightarrow \mathrm{RL}(1) \quad (3)$$

where R is the macromolecule (the receptor), L is the small molecule (the ligand), and RL is their complex. The free energy of this reaction, ΔG_{bind}, is the binding affinity, and it is related to the binding constant K_{bind} by.

$$K_{bind} = e^{-\Delta G_{bind}/RT} \quad (4)$$

where R is the gas constant and T is the absolute temperature. (Strictly speaking, the binding free energy should have a standard-state symbol. In practice, few papers discuss or specify the standard state, although binding affinities calculated with 1 bar or 1 M standard states differ by 8 kJ/mol at ambient temperature, arising from the volume term in the translational entropy; to avoid possible confusion, we have dropped the standard-state symbol throughout this paper.)

From the physical point of view, the ligand binding affinity is defined by the binding energy of ligand to receptor. The binding energy can be estimated either by the docking method [63,65,96,200-216] or molecular dynamics (MD) simulations. The former method is very fast and can be used for screening potential leads from a large number of ligands available in various data bases [217]. However, its predictive power is low due to ignorance of receptor dynamics and a limited number of position trials of ligand. For interaction energy changes, the equation is:

$$\Delta E_{\text{interaction}} = \Delta E_{\text{H-bonding}} + \Delta E_{\text{VdW}} + \Delta E_{\text{electrostatic}}. \quad (5)$$

Hydrogen bonding, Van der Waals contacts, and electrostatic effects were the only intermolecular forces taken into consideration in these calculations. [218]

The receptor theory of drug action posits that a drug works only when bound to its target receptor [219]. Direct measurement of the extent to which a drug is bound to its receptor at equilibrium — the binding affinity — was, however, not possible until long after the theory was first postulated. Accordingly, drug discovery programs historically sought to optimize drug efficacy, not affinity, usually in the context of whole cells, tissues or animals.

Only with the advent of identifiable, and ultimately purifiable, molecular receptors that enabled the direct measurement of binding affinity did optimization of binding affinity guide most early-stage discovery efforts.

This emphasis on binding affinity — quantified either as K_d, the equilibrium dissociation constant, or its proxies, IC_{50} or EC_{50}, the drug concentrations giving half-maximal inhibition or effect — is predicated on the assumption that affinity is an appropriate surrogate for in vivo efficacy. Although many highly efficacious drugs have been discovered on that basis, recent studies have shown that the kinetics of drug–receptor binding could be as important as, and in some cases more important than, affinity in determining drug efficacy [220,221,222]. In an open, in vivo system the concentration of the drug varies over time — potentially on timescales faster than binding and unbinding to its receptor — such that binding equilibrium might not be reached or maintained; for some drugs, attainment of equilibrium might not even be desirable. In these cases, equilibrium binding affinity is no longer an appropriate surrogate for efficacy — instead, the rates of drug–receptor association and dissociation, as reflected by the rate constants k_{on} and k_{off}, are more appropriate

The ligand-protein binding process is too complex to be described by a single representation of the ligand-protein complex produced as a result of the rigid receptor docking [223]. The various levels of approximations necessary to make docking rapid make scientists sceptical to really believe its results [224]. Proteins are not static idles, they are a very complex, moving and viable machines [225,226]. Introducing protein flexibility is increasingly important to describe the ligand-protein binding especially for targets known to be highly flexible such as kinases [227,228,229].

The treatment of very large conformational changes in the receptor induced by ligand or protein-binding remains one of the biggest challenges in calculations of binding free energies. Finding the relevant rotational, translational and conformational degrees of freedom or CVs for a binary complex is far from being trivial, but this would be achievable for less than 20 candidates with the increase of computer power using a multiscale approach that moves through different levels of complexity and precision. In the first step, approximate docking pathways could be sampled with rapid methods such as elastic network models, path-planning approaches and short replica exchange MD simulations based on CG representations of the systems. Next, these pathways could be refined and optimized with metadynamics or other rigorous techniques by retaining a full atomistic description of the system only in regions of interest while describing the rest of the system with elastic network models. [230]

The concepts underlying rational optimization of binding affinity are relatively well understood, but the same is not true for binding kinetics. Much less is known about the molecular determinants of binding kinetics than about those of binding affinity. A major challenge with optimization of kinetics is the fundamental difficulty in characterizing transient states. Binding affinity depends on the free energy difference between the bound and unbound states, both of which are stable and generally easily observable. On- and off-rates depend instead on the height of the (highest) free energy barrier separating those states, yet the atomic arrangement of the drug and the receptor at this point of highest free energy — the transition state — has only a fleeting existence. Understanding the molecular interactions between drug and receptor at this difficult-to-observe transition state is thus central to the rational control of drug binding kinetics.

Despite these challenges, the intentional and rational optimization of k_{on} or k_{off} opens up a new, temporal dimension for controlling drug behavior that has important therapeutic implications for drug efficacy and drug safety. The residence time of a drug–receptor complex, $t_R \equiv 1/k_{off}$, is often a better predictor of efficacy than binding affinity is [220,222,231]. Similarly, when achieving target selectivity is important, a drug with a longer residence time on one receptor can select kinetically for that receptor over another, even when the affinity for both receptors is comparable [220]. Conversely, drugs with faster dissociation rates can increase the therapeutic index (the key measure of drug safety, defined as the ratio of a drug's toxic dose to its efficacious dose) when extended, non-physiological drug occupancy of the target receptor causes toxicity [232,233,234]. Finally, a faster-binding drug might target a short-lived receptor more effectively [235].

Receptor flexibility often plays an important part in modulating the binding kinetics of buried or occluded binding sites. Early studies on carbon monoxide (CO) unbinding from myoglobin revealed the importance of protein breathing motions in enabling CO escape [236]. Indeed, rigidification of myoglobin with an engineered disulfide bond slows CO dissociation [237]. Moreover, receptor flexibility can take the form of intricate loop motions in the binding of more-drug-like molecules [120,226,238,239].

Electrostatic interactions between a charged drug and a charged receptor impact association and dissociation rates, similarly to the effects electrostatics has upon protein–protein binding. Altering the solution ionic strength can greatly affect association rates: increasing ionic strength decreases on-rates but hardly affects off-rates (cf. Debye–Hückel theory [240,241]). Other than as a test for the importance of electrostatics in modulating binding kinetics, however, the pharmacological relevance of this common laboratory manipulation is unclear, because physiological ionic strength is relatively constant. On-rates can also be very sensitive to long-range electrostatic attraction (or repulsion). Off-rates can be modulated by electrostatics, but they tend to be influenced more by short-range drug–receptor interactions such as hydrogen bonds, salt bridges and van der Waals (especially hydrophobic) contacts [242,243]. Binding of a charged acetylcholinesterase inhibitor, for instance, was ~50-fold faster, and unbinding ~10-fold slower, than that of a nearly identical neutral analog in which the inhibitor's trimethylammonium group was changed, by one atom (N^+ to C), to the t-butyl isostere [244].

Experimental studies have shown that it can be difficult to distinguish the electrostatic effects on binding kinetics from the effects of other molecular determinants; deviations from the simple picture described above are not uncommon. The effective charge of a drug or receptor does not necessarily equal its formal charge, and,

paradoxically, 'charge matching' (i.e. negative paired with positive) is not necessarily required for rapid binding [245]. The former issue is demonstrated, for instance, by the insensitivity of the binding rates of nucleotide di- or tri-phosphates to the Na^+/K^+-ATPase, despite the ligand charge varying from -0.8 to -3.8 [243]. Binding of phosphate to (negatively charged) periplasmic phosphate binding protein illustrates the latter issue: surprisingly, despite 'mismatched' charges, the association rate is nearly diffusion controlled [246]. In a similar vein, increasing charge complementarity has been observed to decrease association rates in certain cases [245], and large alterations in (receptor) charge lead to only minor changes in the on-rates of carbonic anhydrase inhibitors [247].

An important component in ligand binding is the strain energy, i.e., the difference in the internal energy of the ligand in the binding site and in solution. [248] One common way to estimate it is by comparison of the energy of the ligand at a certain level of theory calculated in a crystal structure and in continuum solvation after some conformational search. [249,250] However, this will include possible errors in the crystal structure, as well as the disagreement between the energy method used to measure the strain and that used to obtain the crystal structure. A more satisfying approach is to re-refine the crystal structure with a QM/MM approach [251] and use the same method also in solvent. Merz and co-workers have shown that such an approach decreases the estimated strain energy by 80% for a charged ligand. [252]

In the end-point approaches, a conformational sampling is performed, but only of the actual states of the receptor−ligand complex, and possibly also of the free receptor and the free ligand. This is expected to improve the ΔG_{bind} estimates, but it still does not provide true free energies. These calculations often follow the MM/PBSA or LIE approaches, replacing the MM energies by QM, and they will be discussed in separate subsections.

The simplest, but also most approximate, approach to include QM calculations in binding-affinity calculations is the employment of single structures, e.g., obtained from a crystal structure, from docking, or by an energy minimization. The advantage of such an approach is of course the speed; no expensive conformational sampling is performed. On the other hand, such an approach will be strongly affected by the local-minimum problem: An energy minimization will end up in one of an almost infinite number of possible local minima of the receptor−ligand complex, and it is far from certain that this structure is the most important for the binding. As we will discuss below, it is often observed that individual binding affinities estimated by snapshots from MD simulations differ by ~80 kJ/mol even after minimization. [75,252,253] On the other hand, MM studies by Gilson and co-workers on host−guest systems have shown that only a few low-lying conformers contribute to the binding free energy, [254,255] but it is not clear whether this applies to the much more complicated biomacromolecules. Moreover, energies estimated from minimized structures are enthalpies, not the free energies that govern the binding. This is a more serious problem for binding-affinity calculations than, e.g., for enzyme reactions, because the former always involve major entropy terms, owing to the loss of translational and rotational freedom of the ligand, which at least in the gas phase amount to ~60 kJ/mol at ambient temperature. [256] In fact, it is often observed that similar molecules, e.g., enantiomers, have significantly different binding entropies, and there is often a strong inverse correlation between binding enthalpies and entropies of analogous ligands, the much discussed enthalpy−entropy compensation [257,258,259,260].

A natural approach would be to perform of single-point QM energy calculations directly on crystallographic structures, but it is well-known that systematic errors in both the crystallography and the QM calculations would make such energies almost useless, with errors of hundreds of kilojoules per mole. [261] The use of docked structures is usually better, because the docking involves a conformational search for the ligand inside the protein. However, it is still restricted to a single or a few structures, and no valid Boltzmann averaging of the structures is made.

The second approach to study proteins is with QM/MM calculations. Such methods are available in several software products, and they have been much used to study ligand binding. However, most of the studies have employed only QM/MM total energies and mainly discuss structural aspects.

5. Ligands Scoring

The evaluation and ranking of predicted ligand conformations plays central roles in computational drug design, virtual screening of chemical libraries for new lead identification, and prediction of possible binding targets of small chemical molecules, being a crucial aspect of structure-based virtual screening. Even when binding conformations are correctly predicted, the calculations ultimately do not succeed if they do not differentiate correct poses from incorrect ones, and if 'true' ligands cannot be identified. Accurate ligand-protein binding affinity prediction, for a set of similar binders, is a major challenge in the lead optimization stage in drug development. So, the design of reliable scoring functions and schemes is of fundamental importance. Free-energy simulation techniques have been developed for quantitative modelling of protein–ligand interactions and the prediction of binding affinity. However, these expensive calculations remain impractical for the evaluation of large numbers of protein–ligand complexes and are not always accurate.

Accurate prediction of ligand-protein binding affinities plays a crucial role in computer-aided drug design, in particular at the lead optimization stage. The most commonly used structure-based method is still docking and scoring, due to its speed and ease of use [262,263]. Docking fulfills three roles: binding mode prediction; distinguishing binders from nonbinders in a large data set (i.e., virtual screening); and binding affinity prediction of a smaller set of binders. Scoring functions are generally reasonably good at predicting correct binding modes, as has been shown in numerous studies of redocking ligands to cocrystallized complex structures. However, scoring functions are not always able to distinguish the crystallographically correct binding mode, even if it is present in the suggested docking solutions, from other

suggested poses [198,199,264]. In addition, scoring functions have been shown to be successful in enriching binders from a large data set of binders and nonbinders, and therefore are useful for virtual screening [198,199]. However, using docking and scoring at a more fine-tuned level, for accurately predicting binding affinities of a set of binders, or rank compounds accordingly, has proven to be a much more challenging task [127,198,264,265]. The low success rate is mainly because the protein is mostly kept rigid during the docking procedure, allowing only the ligand to be fully flexible. Currently, a number of commonly used docking programs allow for some protein flexibility, either by softening the interactions in the active site, which introduces side-chain flexibility, or by docking to an ensemble of protein structures [266,267,268,269]. However, for many target proteins the allowed protein flexibility is still too small to accurately model ligand-induced changes of the protein conformation or the existence of several protein conformations differently favored by different ligands. This is the classical induced-fit problem. In addition, the scoring functions do not consider the possibility that multiple binding poses contribute to the overall affinity of the ligand. Another challenge for the docking programs is how to treat solvation in the active site. Most programs now offer the possibility to include static or partly rotatable water molecules during the docking procedure, and some programs even offer the possibility of predicting whether certain water molecules should be taken into account for each ligand [99,100,270]. However, the number of water molecules that can be treated this way is generally very low (up to three waters), which can cause problems with larger binding sites, and the overall increase in accuracy by including water molecules is still doubtful. However, there are studies indicating a general improvement in docking results, typically in binding mode prediction [92,271]. The consensus seems to have shifted in favor of including static, or partly rotatable, water molecules in docking calculations, but studies of the actual benefit and molecular accuracy of including them seem to indicate that the improvement is minimal, if any [272,273,274].

In general, docking and scoring functions perform unsatisfactorily. Docking calculations, followed by molecular dynamics simulations and free energy calculations can be applied to improve the predictions. However, for targets with large, flexible binding sites, with no experimentally determined binding modes for a set of ligands, insufficient sampling can decrease the accuracy of the free energy calculations.

An ideal scoring function for protein-ligand interactions is expected to be able to recognize the native binding pose of a ligand on the protein surface among decoy poses, and to accurately predict the binding affinity (or binding free energy) so that the active molecules can be discriminated from the non-active ones. Due to the empirical nature of most, if not all, scoring functions for protein-ligand interactions, the general applicability of empirical scoring functions, especially to domains far outside training sets, is a major concern.

Scoring functions implemented in docking programs make various assumptions and simplifications in the evaluation of modelled complexes and do not fully account for a number of physical phenomena that determine molecular recognition — for example, entropic effects. Essentially, three types or classes of scoring functions are currently applied: force-field-based, empirical and knowledge-based scoring functions.

The comparably fast and inexpensive docking protocols can be combined with accurate but more expensive molecular dynamics (MD) simulation techniques to predict more reliable protein–ligand complex structures [169,275]. On one hand docking techniques are used to search massive conformational space in a short time, allowing the analysis of a large library of drug compounds at a sensible cost [276]. On another hand, MD simulation accounts for both ligand and protein in a flexible way, allowing for an induced fit into the receptor-binding site around the newly introduced ligand [277]. MD simulation can be used: during the preparation of protein receptor before docking, to optimize its structure and account for protein flexibility [278]; for the refinement of the docked complex, to include solvent effects and account for induced fit [148]. This also calculates binding-free energies [153], as well as providing an accurate ranking of the potential ligands [164].

There is no unique solution to a drug design problem. The appropriate experimental techniques or computational methods to use will depend on the characteristics of the system itself and the information available. A variety of computational approaches can be applied at different stages of the drug-design process: in an early stage, these focus on reducing the number of possible ligands, while at the end, during lead-optimization stages, the emphasis is on decreasing experimental costs and reducing times. Although this is simple to articulate, it has been tried many times with only a few fruitful examples. [279-286] The lack of success has led to a re-examination of the underlying principles. For example, recent publications have shown that some of the hypotheses used during the enrichment steps may need to be refined. [287,288] While some drug developers opted for alternative experimental solutions, [289,290] others focused their attention on the improvement of computational protocols. These enhancements include, among others: incorporation of protein flexibility in the docking process, extensive exploration of the ligand conformation within the binding site, refinement and stability evaluation of the final complexes, and estimation of the binding free energies. Not surprisingly, molecular dynamics (MD) simulations have played a dominant role in these attempts to improve docking procedures.

The limited availability of experimentally determined protein structures is one of the bottlenecks of structure-based drug design. In the cases where no protein structures are available, an alternative can be to build a homology model, but they need to be sufficiently accurate to be of use for drug discovery. Validation of these homology models is therefore a crucial aspect in drug development. One important question we aim to address is how errors and inaccuracies of the homology models affect the subsequent molecular modeling of protein-ligand interaction.

Errors introduced into the protein structure through misplacement of side-chains during rotamer modeling led to a correlation coefficient between ΔG_{calc} and ΔG_{exp} of 0.75 compared with 0.90 for the correctly placed side chains. This is in contrast to homology models for members of the retroviral protease family with template

structures ranging in sequence identity between 32% and 51%. For these protein models, the correlation coefficients vary between 0.84 and 0.87, which is considerably closer to the original protein (0.90). It is concluded that HIV-I low sequence identity with the template structure still allows creating sufficiently reliable homology models to be used for ligand-binding studies, although placement of the rotamers is a critical step during the modeling. [291]

In order to further explore binding characteristics of various substrates, Szklarz et al performed binding free energy calculations and compared binding free energies calculated from molecular dynamics simulations to those from experiment. Binding free energies were calculated according to the equation:

$$\Delta G_{bind} = 0.5\Delta V_{el} + \alpha\Delta V_{vdW} \quad (6)$$

where ΔV_{el} is the difference in the electrostatic interaction energy between the ligand and its surroundings in the protein and in aqueous solution, and ΔV_{vdW} represents the difference in van der Waals interaction energies. This linear response approximation is based on the work of Aqvist *et al*. [147]. The value of parameter α used was 1.043, since this value was found to be optimal in earlier studies on binding free energy calculations for P450cam-substrate complexes with the same consistent valence force field [292]. Kollman and coworkers found a similar value for α for P450cam substrates using the Amber force field. Their work suggests that the value of α depends primarily on the class of substrates being examined [160], since both set of substrates were composed of uncharged primarily nonpolar molecules.

Much less computationally demanding is the method described by Lewis, who uses a QSAR approach to estimate binding free energy on the basis of substrate properties such as surface area, number of hydrogen bond donors, log P, pKa and log $D_{7.4}$ [293,294]. In the case of P450 2C9, the calculated values correlated well with the experimental ones [294]. In principle such an approach requires little or no detailed structural information about the P450 active site. However, this technique does require significant experimental binding data in order to parameterize the appropriate free energy relationship. The method employed by Szklarz is intermediate in complexity and in the amount of experimental data required. Like the approach of Wade and coworkers [295], a detailed three-dimensional structure for the enzyme is required. However, as Szklarz results indicate, a homology model rather than a crystal structure can be sufficient to use his model. With only a single adjustable parameter, the approach described requires less experimental data than that of Lewis. Furthermore, Szklarz results indicate that the value of α does not need to be re-parameterized for each P450 isozyme but rather that the same value of α can be used for several different isozymes.

Lizunov et al. analyzed the frequency with which intraligand contacts occurred in a set of 1300 published protein−ligand complexes [296]. Their analysis showed that flexible ligands often form intraligand hydrophobic contacts, while intraligand hydrogen bonds are rare. The test set was also thoroughly investigated and classified. They suggested a universal method for enhancement of a scoring function based on a potential of mean force (PMF-based score) by adding a term accounting for intraligand interactions. The method was implemented via in-house developed program, utilizing an Algo_score scoring function [297] based on the Tarasov-Muryshev PMF [298]. The enhancement of the scoring function was shown to significantly improve the docking and scoring quality for flexible ligands in the test set of 1300 protein−ligand complexes [296]. They also investigated the correlation of the docking results with two parameters of intraligand interactions estimation, the weight of intraligand interactions and the minimum number of bonds between the ligand atoms required to take their interaction into account.

Fragment-based drug design (FBDD) has evolved from a niche technique 20 years ago into a powerful approach used throughout the pharmaceutical and biotech industry. [299-304] Multiple fragment derived compounds are currently in clinical trials, with one compound, vemurafenib, gaining FDA approval in 2011. [305] In brief, the FBDD approach uses libraries of small molecules (fragments, often defined as following the "rule of three" [306] with molecular weights < 300 and ClopP < 3 instead of the "rule of five" in use for drug-like compounds [307] to find hits that bind with a high degree of ligand efficiency, which can then be optimized. Due to the extraordinarily rapid increase in the size of the accessible chemical space as the molecular size is increased, [308] a library of thousands of fragments will offer much higher coverage of the chemical space in its size class than an HTS library with millions of drug-like compounds.

FBDD typically aims for the discovery of weak (e.g., millimolar to high micromolar) inhibitors via biochemical assays or biophysical techniques such as SPR, TSA, ITC, and NMR. Weak initial binders can then be optimized through multiple rounds of chemical modifications, ideally resulting in larger and more drug-like lead compounds with high affinity and specificity. Two possible strategies for fragment optimization involve either growing a single fragment via introduction of chemical R-groups to fill empty regions of a binding site or linking two or more fragment inhibitors with nonoverlapping binding modes via appropriate linking groups. Both techniques make heavy use of structural information, which explains why FBDD programs typically contain an aggressive structural aspect via NMR or X-ray crystallography. The FBDD approach has produced an impressive and growing number of success stories in recent years, with dozens of compounds from fragment studies in clinical trials [309] and multiple mature commercial and academic FBDD programs in existence today.

While the idea to use computational techniques to predict the binding of fragment probes to binding sites was explored from the very start, [171,310,311,312] modern in silico structure-based drug design approaches to optimize the affinity of fragment hits have not played a major role in most FBDD efforts. There are several potential reasons for this, including the following: 1) ligand docking algorithms and scoring functions that have been parametrized to describe the binding of drug-like compounds may not perform as well for fragment-sized compound; 2) the binding of fragment inhibitors, which often exhibit weak directional interactions and more dynamic binding modes, may pose particularly challenging

for an accurate description using fast, empirical methods; 3) since FBDD programs have typically grown around the application of particular experimental techniques, their underutilization of computational approaches might be based on their evolutionary history.

It has also been applied free energy calculation techniques (the FEP methodology in particular) to the study of fragment-sized molecules. FEP calculations are based on molecular dynamics simulations that explicitly consider conformational flexibility and entropy effects through the use of a physics-based force field to describe molecular interactions and explicit solvent to model the real environment of the protein binding site. One of the goals of this study was to explore whether a rigorous free energy approach, such as FEP, can provide predictive power and insights for FBDD without any special parametrization or customization, but results obtained are not yet conclusive. [313]

Since attempts to improve ligand potency usually go hand in hand with increases in molecular weight, FBDD studies often use various metrics of ligand efficiency to judge if a gain in potency is worth the required addition of new functional groups. [314,315] This concept of focusing on the effect that small chemical changes have on the thermodynamics of binding suggests that the results of FEP calculations, which likewise deal with the effect that small chemical perturbations have on the binding free energy, could be well suited for practical FBDD applications.

Relative binding free energies for ligands within a series of drugs can be computed directly from FEP+ calculations, and absolute binding free energy values can be derived from the FEP+ results (relative free energies, i.e. $\Delta\Delta G^0$) by adding a single free energy offset to all members of each series (this offset is chosen to minimize the mean unsigned error of the absolute free energy predictions and is added merely to make comparison of experimental and computational results easier; see refs 316 and 317 for details). This uniform offset value facilitates plotting of the predicted versus measured free energies of binding of the individual ligands and does not alter the mean unsigned error or the R^2-value of the predictions.

Using the approach described to compute absolute binding free energies, the calculated affinities can be directly compared between different systems, unlike many other SBDD methods (e.g., ligand docking scores and approximate free energy methods). The ability to compute a true binding free energy has significant implications, such as being able to compare energies of ligands binding to different targets (binding selectivity) or the effects of putative drug resistance mutations on ligand binding. In addition, accurately modeling the underling physics should result in transferability across a broad range of targets without the need for parametrization. [313]

6. Computing Methodologies

6.1. Techniques Used for Free Energy Calculations

Noncovalent interactions are crucial in chemistry, biochemistry, and materials science as they govern the structure and conformational dynamics of molecular systems and are, therefore, also crucial to reactive properties. The ability to understand and predict noncovalent interactions is thus indispensable to theoretical and computational studies of complex molecules.

The accurate calculation of absolute binding affinities of protein-ligand complexes is a very important goal in the study of biomolecular recognition [318] and computational drug design [319]. However, the currently available computational methods often require some knowledge of experimental binding affinities to calibrate parameters for a particular protein target [162]. The free-energy techniques known as double decoupling methods [320,321,322], have been developed to calculate the absolute binding affinities of complexes without a priori experimental information. These methods involve calculating the free-energy cycle for decoupling the protein and ligand, and then reintroducing the ligand to the bulk solvent. This rigorous technique has only been used for very small ligands [320] or with simplistic implicit solvent models [322], because one of the difficulties involved in this approach is that the ligand must be decoupled slowly enough from the binding pocket such that the mechanical work associated with the process can be performed reversibly. Nowadays, new techniques have been developed that can obtain free energies from repeated nonequilibrium simulations [323,324] and may help make double decoupling applications more efficient. Using a different strategy, Chang et al. enumerated the configuration integrals of the bound and unbound state of simple host-guest complexes to calculate the free energy of association [325].

Any alchemical pathway between bound and unbound states can, in principle, be used to obtain free energies for complex formation. One of the most obvious pathways is to simply pull out the ligand from the active site of the protein by a potential of mean force (PMF) approach. The PMF approach has existed since the early days of molecular mechanics and is well grounded in the statistical mechanics of liquids. The exponential improvements in computer hardware as well as enhanced molecular dynamics algorithms make the PMF approach a reality for protein-ligand systems. Nevertheless, the computational requirements are still quite demanding. Izrailev et al. [326] have been using several pulling methods for over a decade to study the nature of molecular recognition in protein-protein complexes. Fukunishi et al. [327] devised an approach to estimate the free energy of binding in protein-ligand complexes utilizing a self-avoiding random walk procedure. Also Woo and Roux [328] successfully applied a PMF approach to the calculation of the equilibrium binding constant of the phosphotyrosine peptide pYEEI to the Src homology 2 domain of human Lck.

A commonly used approximate method for the calculation of absolute binding affinities is the so-called molecular mechanics-Poisson-Boltzmann-surface area (MM/PB-SA) method [149,329]. In this approach, an explicit solvent simulation of the bound state is carried out and the solvation free energy of binding is obtained from a Poisson-based solvation model [330]. Then the simulation is postprocessed to determine the enthalpic differences between the bound and unbound solute states. Separately, the binding entropy is estimated by harmonic analysis

using a simple r-dielectric function to approximate solvent screening of charge-charge interactions. The free energy of binding, ΔG_{bind} is calculated as [81,331]:

$$\Delta G_{bind} = \Delta E + \Delta G_{solv} + \Delta G_{SA} \quad (7)$$

$$\Delta E = E_{complex} - E_{protein} - E_{ligand} \quad (8)$$

where $E_{complex}$, $E_{protein}$, and E_{ligand} are the minimized energies of the protein–inhibitor complex, protein, and inhibitor, respectively.

$$\Delta G_{solv} = G_{solv(complex)} - G_{solv(protein)} - G_{solv(ligand)} \quad (9)$$

where $G_{solv(complex)}$, $G_{solv(protein)}$, and $G_{solv(ligand)}$ are the solvation free energies of the complex, protein, and inhibitor, respectively:

$$\Delta G_{SA} = G_{SA(complex)} - G_{SA(protein)} - G_{SA(ligand)} \quad (10)$$

where $G_{SA(complex)}$, $G_{SA(protein)}$, and $G_{SA(ligand)}$ are the surface area energies for the complex, protein and inhibitor, respectively.

It is well known that MM-PBSA free energies do not usually replicate the experimental free energy in absolute value, but this approach calculates binding affinity ranking and exhibits good correlation with experiments in certain cases, but only provides modest accuracy for relative binding affinities in systems dominated by electrostatics

The simplest estimation of the free energy of binding $\Delta G_{complexation}$ is that it can be approximated by the binding energy BE

$$BE = H_{complex} - \left(H_{receptor} + H_{ligand}\right) \quad (11)$$

where H represents the total (internal) energy from molecular mechanics for the individual components subtracted from the internal energy of the system in the complex. This approach completely ignores the contribution of entropy as well as solvation, although corrections for entropic effects can be incorporated using a wide range of sampling methods. [332,333] In many computational studies, a classical potential function (i.e., a molecular mechanical force field) is used to describe noncovalent interactions. This is based on the assumption that, in the absence of chemical reactivity and therefore any change in covalent bonding, the potential function can be expressed as a sum of a set of relatively simple functional forms. For the noncovalent component, for example, the typical force field includes Coulombic terms between point charges or higher-order multipoles, [334] Lennard-Jones terms for van der Waals interactions, and sometimes polarizable dipoles, [334,335] fluctuating charges, [336] or charge transfer terms. [337] Classical force fields are vital for condensed-phase simulations due to their computational efficiency; their accuracy for certain properties (e.g., population of various conformations) can be rather high for well-calibrated systems.

Despite the success of force fields, there is still tremendous interest in developing efficient quantum mechanics (QM) based methods for treating noncovalent interactions due to several considerations. First, the parametrization of a force field is often a laborious process that requires extensive tests and refinement of parameters that are not easily decoupled. In the recent years, there has been progress regarding the development of “ab initio” force fields in which parameters are computed rather than fitted. [338-340] Although this is an exciting and promising direction, there are still technical challenges, such as the balance of bonded and nonbonded contributions in the treatment of polymeric systems. Second, most force fields, including those based on first-principles calculations, use rather simple functional forms, which may not be able to capture subtle effects such as hyperconjugation, charge transfers, and other many-body effects. [48,341,342] Third, due to the various approximations in classical force fields, they are likely most suitable for a particular set of molecules under a specific range of conditions. For example, the stability of ion-pair interactions in a protein’s interior is likely overestimated by typical nonpolarizable force fields. [343]

Reddy and Erion have used QM/MM-FES to calculate the binding free energies of five inhibitors of fructose-1,6-bisphosphatase. [344] They used AM1 for the ligand and MM for the protein and solvent. With standard EA, they obtained results with a suspiciously high accuracy: Five relative binding affinities were reproduced with an error of less than 1.4 kJ/mol, although the reported statistical uncertainties of the calculated affinities were 1.9−2.5 kJ/mol. The results obtained at the MM level were almost equally good, with a maximum error of 2.1 kJ/mol. The SEQM calculations increased the computational time by a factor of 5. The same approach was later used to study the binding of some other inhibitors to the same enzyme. [345] Again, excellent results were obtained for 22 relative binding affinities, with maximum errors of 1.7 and 3.3 kJ/mol for SEQM and pure MM, respectively, although the reported uncertainty in the relative energies is 3.8 kJ/mol.

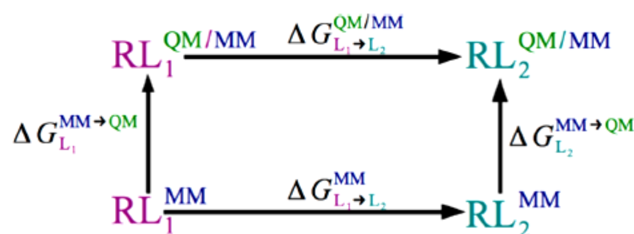

Figure 2. Thermodynamic cycle to obtain relative QM/MM binding free energies. The desired binding energy can be obtained by $\Delta G^{QM/MM}_{L1\to L2} = \Delta G^{MM}_{L1-L2} - \Delta G^{MM\to QM/MM}_{L1} + \Delta G^{MM\to QM/MM}_{L2}$ (and similar for the free ligands in water solution). If the QM/MM and MM potentials are similar enough, it may be enough to use sampling at the MM level [193]

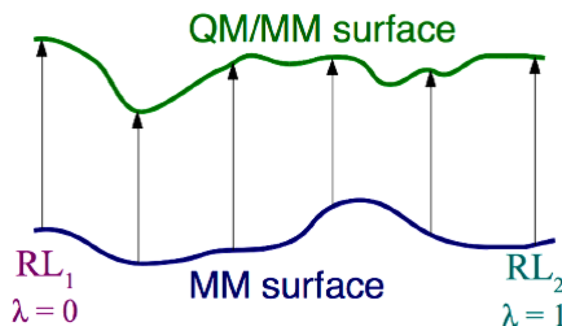

Figure 3. Method of reweighting MM simulations to conform to the QM potential surface. In many applications, only the first and last points are reweighted to the QM/MM surface, whereas the remainder of the binding is studied only on the MM surface. [193]

A slightly more common approach has been to perform the sampling at the MM level and then evaluate QM/MM energies only for a restricted number of snapshots. The problem with such an approach is that the energy functions used for the simulations and the perturbations are not the same, so that normal equations cannot directly be applied to the QM/MM energies. Instead, valid QM/MM free energies should be obtained either by an MM → QM/MM FES calculation, employing the thermodynamic cycle in Figure 2, [346,347,348] or by reweighting the MM snapshots with the QM/MM energy function (Figure 3), e.g., by the non-Boltzmann BAR approach (NBB). [349] Such approaches have been used quite extensively for enzyme reactions, [346,347,348,350,351,352] for solvation free energies, [353-358] and in a few cases also for ligand binding. [359-363] The challenge with these approaches is to obtain converged results for the MM→QM/MM perturbation, which must be performed in a single step to avoid the need of QM/MM sampling, i.e., to ensure that the overlap of the distributions generated by the MM and QM/MM potentials is large enough. For enzyme reactions, proper convergence has been obtained by keeping the QM system fixed, [346,347,348] but for binding affinities, such an approximation seems inappropriate, because the entropy and reorganization of the ligand is crucial for the binding.

Alternatively, binding affinities can be estimated as averages of interaction energies over molecular dynamics (MD) or Monte Carlo (MC) simulations. Such calculations reduce the local minimum problem, but they are also much more time consuming. Many variants have been suggested, but the two most used are the linear interaction energy (LIE) and the MM/PBSA or MM/GBSA (MM combined with Poisson−Boltzmann or generalized Born and surface area) approaches. [58,153]

However, a more strict statistical mechanical way to obtain binding free energies is by free-energy simulation (FES) techniques. [364,365] These also involve MD or MC sampling, but also the conversion of the ligand to either another ligand (giving the difference in ΔG_{bind} between the two ligands) or a noninteracting ligand (giving the absolute ΔG_{bind}). Such conversions need to be performed in many small steps to give a proper convergence, so the FES approaches are computationally expensive.

$$K_{bind} = e^{-\Delta G_{bind}/RT}. \tag{12}$$

A problem with these calculations is that the estimated binding affinities need to be very accurate. Equation (12) shows that a difference in binding constants of 1 order of magnitude translates to a difference of only 6 kJ/mol in ΔG_{bind}. Thus, the accuracy of a computational method needs to be better than this to be useful in drug development. Unfortunately, very few computational methods have such an accuracy, especially not MM methods, with their lacking description of polarization, charge transfer, many-body effects, etc. Therefore, there has lately been quite some interest in improving ligand-binding estimates by using quantum-mechanical (QM) methods. They can in principle include all contributions to the receptor−ligand interaction energy and therefore provide an ideal energy function. However, in practice, QM calculations are also approximate, and depending on the level of theory used, sometimes the approximations may deteriorate the results below the level obtained by MM methods, e.g., because you cannot afford a proper sampling of the phase space. Moreover, often only a part of the receptor−ligand complex is used in the QM calculation. Therefore, it is not certain that QM calculations will automatically improve calculated binding affinities

In 1994, Åqvist et al. suggested the linear interaction-energy (LIE) method. [147,149] It is based on two MD simulations, one of the receptor−ligand complex (RL) and one of the free ligand in water (L). The binding free energy is estimated from the difference in the average electrostatic and van der Waals interaction energies between the ligand and the surroundings (E_{el}^{L-S} and E_{vdW}^{L-S}) in the two simulations:

$$\begin{aligned}\Delta G_{bind} &= \alpha\left(E_{vdw\ RL}^{L-S} - E_{vdw\ L}^{L-S}\right) \\ &\quad + \beta\left(E_{veldwRL}^{L-S} - E_{veldwL}^{L-S}\right)\end{aligned} \tag{13}$$

The two terms are scaled by empirical constants, α and β. β should be 0.5 according to the linear-response approximation, but it often is assigned values of 0.3−0.5 depending on the chemical nature of the ligand. [366,367] The other parameter is truly empirical.

Over the last three decades, Quantum Monte Carlo (QMC) methods were applied to numerous applications that are directly or indirectly related to systems where accurate description of noncovalent interactions is crucial. These applications include biomolecular systems of intermediate and large sizes. For example, Korth et al. performed large-scale calculations to obtain interaction energies of adenine-thymine and cytosine-guanine base pairs, [368] a benchmark FNDMC interaction energy of a large supramolecular host−guest complex, glycine anhydride interacting with amide macrocycle, was calculated by Tkatchenko et al. [369], the interaction of adenine-thymine step in B-DNA was estimated by FNDMC in the work of Hongo et al. [370] and Benalit et al. presented a FNDMC estimate of interaction energy for the fragment of DNA intercalated with anticancer drug, ellipticine. [371]

QMC has a number of important ingredients that make it attractive for a variety of many-body quantum problems, including challenges such as calculating very small interaction energies in noncovalent systems. Dubecký et al. emphasized that the real-space Fixed Node Diffusion Monte Carlo (FNDMC) methods recover all possible many-body correlations within the nodal constraint $\Psi_T = 0$. This property makes QMC an interesting choice for noncovalent interactions in particular, since noncovalent interactions result from subtle long-range correlations that are spread across sizable parts of the configuration space. Fittingly, these are exactly the types of many-body effects where QMC is especially effective. Indeed, the demonstration that QMC is able to produce results competitive with coupled-cluster calculations with single, double, and perturbatively treated triple excitations, [372-378] and in some cases calculations with coupled-cluster singles, doubles, triples, and quadruples CCSDT(Q), [379] is already convincing for small systems, and similar performance is expected for larger systems.

The main source of FNDMC errors is the approximate Ψ_T which determines the fixed-node bias and in the presence of effective core potentials, also bias from the localization approximation. Besides that, further sources of possible errors exist: population control bias, DMC time step bias, or treatment of noninteracting structures reference energy (it matters whether one obtains the total energies of structures separately or in a single run with structures separated reasonably far from each other [380]). As Dubecký et al. indicated, the calculations should be organized in such a manner that the interacting system and the structures are described on the same footing, and in addition, it is necessary to check that the technical errors are under control while the cancellation in energy differences is maximized

For covalently bonded systems, the thermochemical accuracy of 1 kcal/mol is usually satisfactory [381-385] and provides useful insights for processes such as bond formation/breaking in reactions and similar phenomena. However, benchmarks for noncovalent interactions require typically an order of magnitude higher accuracy, on par with the subchemical threshold value of 0.1 kcal/mol that poses a steep challenge for any computational method, because technical parameters used in QMC calculations that make no or very little qualitative difference at the scale of 1 kcal/mol may play a decisive role in case of noncovalent interactions. The quality of QMC results depends considerably on parameters that enter the typical multistage computational sequence, since each step depends on its own set of technical parameters and/or user choices. One such set of parameters and choices includes selection, construction, and variational optimization of trial wave function ansatz. The FNDMC production step requires sufficiently long projection times that reach the desired states. Another relevant DMC aspect is checking the time step and population control biases. In addition, the biases coming from the treatment of ECPs must be kept under control as well.

It is reasonable to assume that an explicit construction of exact eigenstate(s) for a large interacting quantum system is both unnecessary and not really useful. What is really "only" needed is the necessary amount of information that enables evaluations of accurate expectation values (e.g., energy differences and other quantities of interest within a desired error margin). This is in fact the basic premise of the reductionism paradigm that underlies a vast number of methods based on mean-fields, DFT, reduced density matrices, [386-388] etc. The challenge comes from the fact that the reduced quantity is actually a very complicated object since all the many-body effects have been folded into it during the reduction (for example, by integrating over (N−2) particles when getting the two-body density matrix). However, this is in general a very difficult task due to its inverse nature: one wants to reconstruct correlations from the particles that are already integrated out. Advanced methods based on density matrix renormalization [389-391] and other renormalization group methods or sophisticated perturbation approaches try to address exactly this critical issue by carefully guiding the reduction in an appropriate and presumably efficient manner. However, beyond a certain level the systematic improvements of reduced quantities often becomes very difficult, sometimes perhaps almost as difficult as solving the full many-body problem in the first place.

In this respect, QMC appears to be a unique methodology that combines known analytical insights and direct constructions with the robustness of the stochastic methods in order to capture the many-body effects efficiently. It seems that this combination offers somewhat of a sweet spot between the fully explicit and the fully reduced descriptions. In particular, the uninteresting and heavy load of reducing (i.e., calculating expectation values) is left to the machine.

The value of the QMC method comes also from its new insights that reveal the nature of quantum correlations that are stimulating for correlated wave function methods in general. In particular, (i) stochastic sampling is carried out from a complete basis, and the extent of sampling is determined automatically by the desired error bar, largely avoiding thus the basis set issues; (ii) to the leading order, the explicit inclusion of exact nonanalytical behavior, such as electron−electron cusps, eliminates another issue that is difficult in many other approaches; and (iii) in addition, the correlation factor captures the smooth, medium-, and long-range correlations with remarkable efficiency, with just a few to a few tens of variational parameters.

On the other hand, as expected, QMC has its own limitations and challenges. The key challenge that has a long history and that has to do with the fundamental and infamous Fermion sign problem can be formulated also as a construction of an optimal effective Hamiltonian. The actual Hamiltonian that is solved exactly by the fixed-node DMC can be written as

$$H_{\mathrm{FN}} = H + V_{\infty}(\Gamma)(17) \tag{14}$$

where Γ is the location of the nodal hypersurface and $V_{\infty}(\Gamma)$ is an infinite barrier at this subset of configurations. Improvement of $V_{\infty}(\Gamma)$ is done indirectly through more accurate trial functions, and it is therefore important to understand the sources of the nodal errors. As described above, the fixed-node bias grows with the electronic density as well as with the complexity of the bonds. [392] This finding has implications also for noncovalent systems since dispersion interactions result in low densities and mainly σ-like character of single-reference bonds, hence the fixed-node errors are much less pronounced in energy differences. However, this finding has a significant importance also for other electronic structure problems and possibly beyond. Clearly, these arguments need to be further refined and quantified so that much remains to be done in this direction.

Another "technical" issue of importance is the construction and testing of accurate ECPs (Effective Core Potentials): since QMC can provide benchmark accuracy for many systems, the quality of ECPs for heavier elements becomes crucial. In particular, new opportunities would open up with ECPs that would enable valence only calculations with an accuracy target, say, 0.1 kcal/mol (≈ 0.005 eV) or so for energy differences. Furthermore, overall applicability of QMC methods could be significantly enhanced with more systematic benchmarking, testing, and providing data sets of calculations for a broad use.

Perhaps one of the greatest challenges is better understanding of errors related to current QMC procedures.

So far most of the calculations focused on energy differences such as cohesion, gaps, and similar quantities that are larger than 0.1 eV or so, rather than on tiny differences important for intermolecular bonding. Even though some sources of errors have been already identified, it is still not fully understood to what extent they are systematic, how they scale with the system size, and what are the best ways of reducing them. In addition, it is still possible that more sources of errors will be uncovered. This state of the matter is strikingly different from the mainstream WFT, where sources of errors are very well mapped out and many solutions exist that may be routinely applied. Much more detailed work will be necessary to achieve this level of routine with the QMC methods.

One of the open challenges is also optimization of the QMC codes and development of new fast algorithms. Whereas mainstream methods and codes were optimized for performance over many decades, both on the side of more efficient programming and method development, at present the QMC codes are mostly basic research tools. Obviously, this is an opportunity for future as it is likely that QMC codes may become more efficient once they become more widely used.

These considerations have led to the development of various linear-scaling QM methods, [393,393,395] which hold the promise to treat both covalent and noncovalent interactions for large molecules. In practice, however, linear-scaling QM calculations remain computationally expensive whenever ab initio QM or density functional theory (DFT) methods are used. This is a particularly serious limitation for the study of biomolecules and other soft matter, where adequate conformational sampling is imperative. For many biological applications, [396,397] for example, molecular dynamics simulations on a nanosecond to microsecond scale are required, which involve millions to billions of energy and force evaluations.

This review has illustrated the great interest of using QM calculations to improve estimates of binding affinities. This is quite natural, considering the increasing awareness of the shortcomings of standard MM force fields. In particular, QM calculations automatically include effects of polarization, charge transfer, charge penetration, and the coupling of the various terms. Moreover, QM avoids the need of parametrization of force fields for the ligands, which is a tedious and time-consuming procedure, if you aim at accurate results. QM can also consistently treat the formation of covalent or metal-coordination bonds.

From reading this paper, you can easily get the impression that inclusion of QM methods nearly always improves calculated binding affinities. But this is not an accurate picture as there are several examples of QM methods giving comparable or even worse results than MM methods. [359,398,399] The apparent success is probably an effect of the fact that it is easier to publish success stories. [400,401] Moreover, the performance of a method depends on the quality measures used: Methods not based on FES or fitting tend to overestimate energies and energy differences, and QM methods often increase this overestimation. However, this typically improves the correlation between experimental and calculated affinities. Therefore, a correlation coefficient should always be supplemented by a measure of the agreement in terms of the free energy (MAD or RMSD), but such results are much more seldom reported.

Thus, it can be concluded that QM methods do not currently automatically provide a clear improvement in computational estimates of binding affinities. This is probably caused by a combination of several effects: insufficient sampling and treatment of entropy effects, the use of rather crude continuum-solvation methods, and the fact that a QM treatment of most interactions (e.g., electrostatics, polarization, and repulsion) gives rise to larger energy components (of opposite signs) that require a higher precision to give accurate final results (MM methods gain much from error cancellation). However, this may change as it becomes possible to perform valid FES simulations at the QM(/MM) level.

It is in this context that semiempirical (SE) methods, which have a long history in quantum chemistry, [402] have come back into the spotlight in recent years. The most prevalent SE methods are those based on approximations (e.g., neglect of diatomic differential overlap, NDDO) to the Hartree−Fock (HF) theory, leading to methods such as AM1, [403] PM3, [404] MNDO/d, [405] and OMx. [406] Another approach that has become popular in the past decade is the density functional tight binding (DFTB) approach, [407,408,409] which was derived in the framework of DFT based on a Taylor expansion of the energy with respect to a reference density. Both sets of SE methods use minimal basis sets and involve various approximations to electron integrals, leading to an increase of computational efficiency by a factor of 100 to 1000 over typical implementations of ab initio QM and DFT methods. As a result, with the same computational resources, SE methods can be used to study systems 10 times larger or to carry out 1000 times longer sampling. [410] These enhancements can be further improved by integrating SE approaches with modern computational architectures (e.g., GPUs) [411,412] and computational algorithms (e.g., linear-scaling/fragmentation techniques, [393,394,395,413,414] faster diagonalizations, and/or extended Lagrangian MD algorithm). [415]

These considerations, however, raise the following critical question: Are the SE methods sufficiently accurate for the description of structure, dynamics, and reactivity of complex molecular systems? The development of SE methods has focused on the description of the chemical bond traditionally; therefore, there is vast literature on the parametrization and benchmark of SE methods for heats of formation, structures, and other properties of mostly small molecules. [416,417,418,419] The description of larger systems, in which noncovalent interactions like van der Waals forces and hydrogen bonds are important, poses different challenges. In this review, we focus on this aspect of SE methods.

As discussed earlier in this review, the lack of polarizing functions in a valence-only AO basis set causes the intermolecular polarization to be underestimated. A direct consequence is that the strength of an intermolecular interaction between polar functional groups is underestimated. During the past decade, a number of post-SCF hydrogen bonding corrections have been proposed to alleviate this shortcoming of SE (NDDO) methods.

Halogen bonds are noncovalent interactions between an electron donor and a halogen atom covalently bound to an electron acceptor. Compared to the hydrogen atom in a hydrogen bond, the electron density around a halogen atom is much more anisotropic, and a positively charged region exists along the covalent bonding axis of the halogen atom, the so called σ-hole. This region interacts with the lone pair of the electron donor, forming the halogen bond. Riley and Hobza used symmetry-adapted perturbation theory to show that this interaction accounts for approximately half of the halogen bond energy, while the rest of the interaction energy is mostly due to dispersion interaction. [420] Halogen bonds are badly reproduced by minimal basis set SE methods, which systematically overestimate the interactions. Řezáč and Hobza have devised the halogen bond "X-correction" term for PM6, which is combined with Grimme's D2 dispersion corrections to yield the D2X correction. [421] The halogen bond correction adds a repulsive potential to alleviate the overestimated interaction energy.

Noncovalent interactions are most frequently evaluated by calculating the dissociation energy between two or more individual molecules in a cluster. Accurate estimates of these energies can be done by using extrapolation techniques: these approaches exploit the fact that the remaining correlation energy beyond MP2 converges quite rapidly. [422] Thus, the CCSD(T)/CBS (coupled-cluster calculations with single, double, and perturbately treated triple excitations/complete basis set) energy can be approximated as the sum of the MP2/CBS energy and a small CCSD(T) correction

$$E_{CCSD(T)}^{CBS} \approx E_{MP2}^{CBS} + \Delta E_{CCSD(T)}^{small-basis} \quad (15)$$

$$\Delta E_{CCSD(T)}^{small-basis} = E_{CCSD(T)}^{small-basis} - E_{MP2}^{small-basis} \quad (16)$$

where E_{MP2}^{CBS} is the MP2 energy extrapolated to the CBS limit, and $E_{CCSD(T)}^{small-basis}$ and $E_{MP2}^{small-basis}$ are the CCSD(T) and MP2 energies evaluated with a smaller basis set. Extrapolation schemes such as that by Halkier and co-workers can be used to extrapolate the MP2 energy to the CBS limit. [423] Řezáč and coworkers recommended using at least aug-cc-pVDZ as the small basis set and preferably aug-cc-pVTZ and aug-cc-pVQZ to extrapolate the MP2 energy to the CBS limit, and the Boys−Bernardi counterpoise correction [424] should be used to remove any basis set superposition error. [425] For dimer interaction energies calculated using equation (15), the error is estimated to be around 1%, and such energies are thus appropriate as validation and benchmark data for SE methods. [426]

Christensen et al. have pointed out that a high accuracy of an SE method in the gas-phase does not necessarily guarantee the transferability of the method to the condensed phase. [427] For this reason, it is necessary to create data sets containing larger molecular clusters, such as trimers, host−guest complexes, and water clusters. In addition, a treatment such as equation (14) is trivial for smaller molecules, but the steep scaling of the CCSD(T) correction becomes prohibitive for large complexes quickly.

The use of localized molecular orbitals (LMOs) with cutoffs in coupled cluster methods have led to methods with greatly reduced scaling compared to canonical coupled clusters, and these methods are very attractive for use in the study of large molecular assemblies. For instance, the DLPNO-CCSD(T) (domain-based local-pair natural-orbital coupled-cluster) method [428] has recently been applied to a data set containing complexes with up to 112 atoms, [39] being expected for these LMO methods to constitute the basis of future data sets with larger molecular assemblies.

All of the various scoring functions (SFs) used in computer-aided drug design to evaluate drug candidates rely on a rapid and accurate evaluation of intermolecular interactions between the drug and the receptor. Therefore, the SF must be able to describe all relevant types of noncovalent interactions, [429] such as dispersion, hydrogen bonding, [430] and halogen bonding. [431] Application of calibrated SE methods is a good way to meet these conditions. [432] For recent dedicated reviews on the application of SE methods in ligand scoring, refer to refs [433,434,435].

The possibility of applying a QM method in the framework of an SF was investigated by Vasilyev and Bliznyuk in 2004. [32] They applied an efficient implementation of AM1/COSMO to estimate the binding energies for a series of RNA···theophylline-analog complexes, and the approach was found useful, although preliminary. Another SE study of ligand−protein interactions aimed to reproduce experimental binding enthalpies of several complexes using PM3. [436] The results from calculations agreed well with the experiment, with an error of up to 2 kcal/mol. A different idea was exploited in a docking study in which PM6-based atomic charges instead of empirical charges were used in the SF where scoring results improved markedly. [437]

The SE-based scoring has been developed further by Hobza and co-workers, who have complemented the SE-based SF with advanced empirical corrections to the SE methods. [438] Much like the previous work by Merz and co-workers, [439,440] the SF by Hobza and co-workers relies on a phenomenological, idealized decomposition of the process of binding, which in turn leads to the approximation of binding free energy by the sum of several contributions with a clear physical meaning. These individual contributions are the gas-phase interaction energy, the solvation/desolvation free energy, the change of the conformational free energies of the protein and ligand, and the entropy change upon binding. The values of the contributions and their comparison may provide additional insight into the nature of the protein−ligand binding. Unlike the case of an empirical SF, none of the contributions are fitted, for example, to experimental data in any way. The idea is to use the most accurate methods for the respective terms, however, in such a way that they are balanced with the other terms, in terms of both accuracy and computational cost. Still, two of the terms in the SF, the protein deformation energy and the binding entropy, would need prohibitively long calculation times to be evaluated accurately, and thus, they are evaluated largely approximately. Consequently, the SF shall provide merely relative values of the binding free energy, and the goal is a (linear) correlation of the score with the real, experimental binding free energy.

The gas-phase interaction energy is calculated using geometries obtained with SE energy minimization in the

solution phase, and its magnitude is usually the largest among all of the contributions to the SF, amounting to as much as several hundred kcal/mol (for charged ligands). It is, however, compensated for largely by the change of solvation free energies upon binding. Consequently, even small deviations may lead to large deviations in the final score, and special care is needed here. The gas-phase interaction energy is calculated with a corrected PM6 method—most recently, PM6-D3H4X—which was shown to be the most accurate among SE methods, outperforming even the costly ab initio MP2 method.

Although the main motivation to use an SE method in the scoring may be the improved accuracy with respect to empirical schemes, there are also situations in which any empirical scheme would come into trouble for principal reasons: (i) There is a chemical element (or binding situation) that is difficult to parametrize in MM. (ii) A covalent bond is formed between the receptor and the ligand. (iii) Another process that cannot be described with MM takes place; an example is a halogen bond due to the presence of a σ-hole. In any of these cases, the application of an SE-based SF is clearly superior.

In ligand docking and scoring studies, it is valuable to combine SE methods with computational approaches of miscellaneous sorts. To illustrate such possibilities, some other applications are mentioned below.

It was investigated how the conformational and desolvation free energies are affected if only a single conformer of a flexible inhibitor of HIV protease was considered. [253] To provide reference data, sufficient sampling was guaranteed by means of extensive MD simulation (using an MM force field). The error caused by the limited treatment of ligand flexibility amounted to ca. 5% of the total range of the scores. It was concluded that an approach restricted to a single conformer represents a reasonable, viable compromise between accuracy and computational efficiency.

The statistical mechanically strict way to obtain binding free energies is to perform free-energy simulations (FESs). However, these require extensive sampling of the receptor−ligand complex and various intermediate states, making them computationally very demanding. Therefore, full FES simulations for ligand binding have been performed only at the SEQM/MM level and in only a few cases. It has been more common to perform the sampling at the MM level and then try to extrapolate these results to the QM level.

Simulation of phenomenological models constitutes a powerful tool to explore binding processes. This procedure allows generating equilibrium binding isotherms from the time courses of the binding reaction and, from these data, the Hill coefficients can be determined. We show that Hill coefficients estimated from these isotherms have a bi-univocal relation with the Gibbs free energy of interaction among binding sites and their values are independent of the free energy of ligand association to the empty sites. Furthermore, a careful exploration of the simulated data shows distinctive features between the binding time courses corresponding to negative cooperativity and different classes of binding sites, although they are undistinguishable at equilibrium. In this way, our results also highlight the usefulness of preequilibrium time-resolved strategies to explore binding models as a key complement of equilibrium studies. Additionally, simulated results show that under conditions of strong negative cooperativity, the existence of some binding sites can be overlooked. We show that experiments at very high ligand concentrations (when compatible with solubility and stability conditions) are a valuable tool to unmask such sites. [441]

Specific recognition and interaction between macromolecules and ligands determines the fate of most cellular processes and thus the behavior, response and regulation of essential functions in all living organisms [442,443]. Regulation of gene expression [444], enzyme activities [445,446], protein stabilization [447,448], cell membrane electrochemical potential [449], oxygen transport [450] and neural proliferation [451] are only a few examples of the great diversity of phenomena that occur in biological systems as a consequence of intermolecular interactions. Despite its key role, the underlying relationships and mechanisms are still a matter of debate.

Cooperative binding represents perhaps one of the most interesting, and not fully understood, types of molecular interactions observed in nature [452]. For macromolecules having two or more binding sites, cooperativity is characterized by a change of the intrinsic (site specific) equilibrium binding constant as a function of the reaction progress (i.e. the affinity of a given binding site for a ligand will be affected by the occupancy of other sites by the same or different ligands). The first modeling approaches describing cooperative binding were proposed by A.V. Hill at the beginning of the twentieth century by analyzing the binding of oxygen to human hemoglobin [453,454]. At that time, hemoglobin was thought to be a monomeric molecule containing one atom of iron [455]. To conciliate this data with the sigmoidal shape of the oxygen binding curve, Hill proposed that these monomers aggregate in groups of n units, and that this 'aggregate' bound n molecules of oxygen simultaneously [455]. From this model a mathematical expression (equation (17)) was derived [454], where the fractional saturation of hemoglobin by O_2 (θ) is expressed as a function of the partial pressure of O_2 or, for the general case, the free ligand concentration [L]. This equation includes two parameters: an association constant (K) and an exponent affecting the ligand concentration today denoted as the Hill coefficient (n_H).

$$\theta = \frac{K \cdot [L]^{nH}}{1 + K1 + K.[L]^{nH}}. \quad (17)$$

Despite the fact that these two hypotheses (hemoglobin aggregation and infinite interaction) are perhaps unrealistic, equation (17) is still nowadays the most used mathematical expression to describe cooperative binding in the scientific literature. In this scenario, it is commonly accepted that the Hill coefficient provides a criterion to determine the type of interaction between binding sites in a macromolecule. When the value of the Hill coefficient is 1, the Hill equation becomes a rectangular hyperbola indicating that there is no interaction among binding sites (referred hereafter as identical and independent sites). If n_H takes values higher than 1, it is said that the system shows positive cooperativity; this could be the result of an

increase in the affinity of a binding site due to the previous binding of a ligand to another site. Instead, n_H values lower than 1 would indicate negative cooperativity (also called antagonism) and, in this case, the binding of the first ligand molecule diminishes the probability of binding for a second molecule. However, the condition $n_H < 1$ while being necessary, is not sufficient to probe the existence of negative cooperativity, since macromolecules with multiple binding sites and different ligand affinities will also depict n_H values lower than the unity [456]. Although there are few well-documented cases of negative cooperativity (see Ruzicka & Frey [457], Abeliovich [458] and references therein), its relevance cannot be underestimated. Indeed, Koshland and Hamadani suggested that both, positive and negative cooperativity, are part of a phenomenon of universal importance in biological systems, and have about equal evolutionary relevance [459].

6.2. Alchemical Models

Estimating binding free energies accurately is a very time-consuming process. The most accurate results are obtained with methods such as Free Energy Perturbation (FEP) / Thermodynamics Integration (TI), and similar results can be obtained at a lower computational cost with methods such as MM-PBSA/MM-GBSA or Linear Interaction Energy (LIE). These methods are however still considered too computationally intensive to be of much use in virtual screening approaches. [460]

Low-throughput computational approaches for the calculation of ligand binding free energies can be divided into “pathway” and “endpoint” methods [461,462]. In pathway methods, the system is converted from one state (e.g., the complex) to the other (e.g., the unbound protein/ligand). This can be achieved by introducing a set of finite or infinitesimal “alchemical” changes to the energy function (the Hamiltonian) of the system through free-energy perturbation (FEP) or thermodynamic integration (TI), respectively [152,463]. The fundamentals of FEP and TI methods were introduced many decades ago by John Kirkwood [464] and Robert Zwanzig [465]. In recent years, their use in the computation of absolute binding affinities has become feasible due to increases in computational power, the development of more accurate models of atomic interactions [466,467,468], the clarification of the underlying theoretical framework and the introduction of methodological advances [322,461,462,469,470,471]. Combined with atomistic molecular dynamics (MD) or Monte Carlo (MC) simulations in explicit water solvent models, they are arguably the most accurate methods for calculating absolute or relative ligand 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. FEP and TI methods are often referred to as "computational alchemy", in the sense that they evaluate the difference between the binding energy of two similar ligands by using pathways to compute the change in free energy when ligand A is changed to ligand B within the binding site and in solution [472]. The horizontal legs describe the experimentally accessible actual binding processes, with free energies $\Delta G_{bind}(L1)$ and $\Delta G_{bind}(L2)$. Since the free energy is a state function, the relative binding free energy $\Delta\Delta G_{bind}$ is exactly equal to the difference of the free energies in the horizontal or vertical legs.

These methods generally give a very good estimate of the binding energy, with errors below 1kcal/mol [473,474]. By slowly "growing" the ligand into the binding site it is also possible to calculate the absolute binding free energy, but this is a very time consuming process [295]. Figure 4 shows the thermodynamic cycle that is used for chemical alchemy calculations.

$$\Delta\Delta G_{\text{bind}} = \Delta G_{\text{bind}}(L2) - \Delta G_{\text{bind}}(L1) = \tag{18}$$

$$= \Delta G_{\text{complex}}(L1 \to L2) - \Delta G_{\text{free}}(L1 \to L2). \tag{19}$$

The simulations follow the vertical steps (equation (19)) 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 4, 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 $\Delta G_{free}(L1 \to L2)$ and $\Delta G_{complex}(L1 \to L2)$, respectively. Averaging over both transformation directions is often used to improve the free-energy estimates, although this is not always the case [475]. 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 is in [476].

MM-GB(PB)SA methods are widely recognized as valuable tools in 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 [84,477].The inclusion of entropic contributions brings the MM-GB(PB)SA values somewhat closer to experimental absolute affinities [462]. However, such entropic terms are costly and contain large uncertainties. Force-field 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 [83,84].

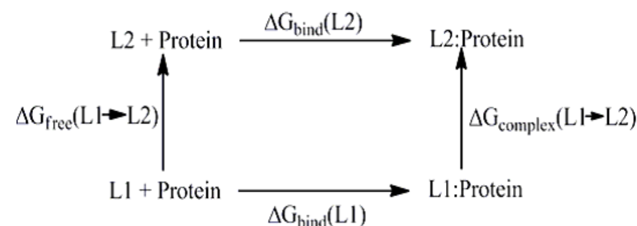

Figure 4. Thermodynamic cycle linking the binding of two ligands L1 and L2 to a protein in solution

Continuum electrostatics models ignore the molecular structure of the solvent; in some cases this might affect the results, particularly when key receptor-ligand interactions are bridged by water molecules, *c.f.* Section 4.1 [478]. 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 [479,480]. Hou and coworkers suggested in a recent MM-PBSA study that the use of ε_{in} = 4 for a highly charged protein-ligand binding interface, ε_{in} = 2 for a moderately charged binding interface and ε_{in} = 1 for a hydrophobic binding interface may improve ligand ranking [84]. The lack of a consistent optimum dielectric constant for MM-PBSA calculations has been noted by other workers (see for e.g. [481]), although generally a value ε_{in} =4 often gives satisfactory results [481-483]. As a final note, MM-GB(PB)SA calculations require some degree of user expertise and planning, from the initial set-up and analysis of the MD simulations through to the binding free energy calculations themselves.

In the MM-GB(PB)SA formulation, the binding free energy of a ligand (L) to a protein (P) to form the complex (PL) is obtained as the difference [484]:

$$\Delta G_{\text{bind}} = G(PL) - G(P) - G(L) \tag{20}$$

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

$$G(X) = E_{MM}(X) + G_{\text{solv}}(X) - TS(X) \tag{21}$$

In equation (21), E_{MM} is the total molecular mechanics energy of molecular system X in the gas phase, G_{solv} is a correction term (solvation free energy) accounting for the fact that X is surrounded by solvent, and S is the entropy of X, E_{MM} is the sum of the bonded (internal), and non-bonded electrostatic and van der Waals energies

$$E_{\text{MM}} = E_{\text{bonded}} + E_{\text{elec}} + E_{\text{vdw}} \tag{22}$$

These energy contributions are computed from the atomic coordinates of the protein, ligand and complex using the (gas phase) molecular mechanics energy function (or forcefield). The solvation free energy term G_{solv} contains both polar and non-polar contributions. The polar contributions are accounted for by the generalized Born, Poisson, or Poisson-Boltzmann model, and the non-polar are assumed proportional to the solvent-accessible surface area (SASA):

$$G_{\text{solv}} = G_{\text{PB(GB)}} + G_{SASA}. \tag{23}$$

Finally, the entropy S is decomposed into translational, rotational and vibrational contributions. The first two are computed by standard statistical-mechanical expressions, and the last is typically estimated from a normal-mode (harmonic or quasiharmonic) analysis [485,486,487]. In practice, current software implementations normally determine all three contributions to S as part of a normal-mode analysis.

Usually, to improve the accuracy of the computed binding free energies, the various terms of equation (21) 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.

This procedure is costly and prevents the consideration of a large number of conformations; insufficient sampling can therefore sometimes an issue. 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 [488]. However, a large variation of the entropy term often results from these 'free' minimizations. Including a fixed buffer region (with water molecules) beyond the cut-off can lead to more stable entropy predictions [488].

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 [489,490], which employs simulations of a single state (the complex) to generate conformations for all three states (complex, protein and ligand). 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 equation (20). Hence, effectively only the protein-ligand (non-bonded) interaction energies of the E_{MM} term in equation (22) contribute to ΔG_{bind}. 'Single-trajectory' simulations significantly reduce computational effort and are generally sufficiently accurate for most applications. The downside of the approximation is 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 ε_{in} = 1-2 [479,480,482,491], the neglect of explicit structural relaxation may introduce errors depending on the system [492]. Separate MD simulations for the complex, and 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 co-workers 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 XIAP [493]. In certain cases, therefore, the added expense of separate simulations may be justified.

Proteins function usually inside aqueous solutions or in membrane environments, which are in the vicinity of an aqueous medium. The surrounding solvent can influence protein stability and function, ligand binding and protein-protein association. Since the solvent modifies in a non-trivial manner the intramolecular and intermolecular interactions, an accurate inclusion of solvent effects in biomolecular modeling and simulation is a challenging task. Currently, the most accurate treatment in molecular simulations is achieved by atomic-detail models that represent explicitly the biomolecule and its surrounding environment.

Several water models are used successfully to describe the aqueous environment in atomistic simulations; examples include SPC [494], SPC/E [495], TIP3P and TIP4P [496], and TIP5P [497]. In practice, the explicit inclusion of water leads to a considerable increase in both the size of the simulation system and the computational cost of the simulation itself. Furthermore, the computation of solvation or binding free energies requires an exhaustive sampling of the solvent degrees of freedom. The energies are normally averaged over conformations sampled from molecular dynamics simulations, often with explicit water. These methods have been applied to a variety of ligand-protein complexes and proven to give good estimates of the binding energy [498]. Despite the differences observed by Michel et al for various sets of enzyme inhibitors, the use of an implicit solvent framework to predict the ranking of congeneric inhibitors to a protein is shown to be faster, as accurate or more accurate than the explicit solvent protocol, and superior to empirical scoring schemes. [499]

A much less costly approach is to represent the solvent implicitly in the simulation, through the incorporation of additional "potential of mean force" terms [500,501] in the gas-phase energy function (e.g., equation (24) below). These terms depend only on the atomic coordinates of the solute, and express the solute free energy for a given configuration, after the solvent degrees of freedom have been "integrated out" [500, 501]. Thus, the simulation system has the same number of degrees of freedom as in the gas phase and there is no need for explicit sampling over solvent degrees of freedom.

$$G_{SASA}(X) = \gamma \cdot SASA + \beta. \qquad (24)$$

MM-PBSA and MM-GBSA are so called end-point methods in the sense that they only evaluate the initial and final states of the system instead of the path between the states. They can use molecular dynamics or Monte Carlo simulations to obtain snapshots of the protein-ligand complex which are used to calculate the average binding free energy. If the configurational entropy is included it is estimated by minimizing a small number of snapshots and from them calculate the entropy with a rigid-rotor/harmonic-oscillator approximation [462], but often this term is neglected if only relative binding affinities are required because it is very time consuming.

To examine the potential of computationally designing SHC ligands and study their thermodynamics of binding, the relative free energies of binding of a series of structurally similar anticholesteremic inhibitors of SHC were calculated by Schwab et al. using single-step perturbation (SSP) and thermodynamic integration (TI) techniques. While neither technique succeeds in quantitatively matching the relatively small experimental values, TI qualitatively reproduces the relative order of the experimental affinities, but SSP does not. [502]

To conclude, the emerging implementation of biomolecular codes on GPU architectures [503,504] and the development of simple free-energy protocols [505] make atomistic methods of absolute or relative affinities very promising for larger-scale calculations in the near future. Nevertheless, at present they are still relatively time-consuming, and require considerable expertise and planning. They preclude the consideration of more than a few complexes per day on a dedicated CPU cluster with a few tens of nodes. 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) [162] or the molecular mechanics – Poisson Boltzmann (MM-PBSA) [58] and the related molecular mechanics – generalized Born (MM-GBSA) approximation [506]. All these methods compute binding free energies along the horizontal legs of Figure 4, but use only models for the "endpoints" (bound and unbound states).

6.3. Molecular Dynamics Simulations

Computational chemistry and molecular modeling have become an indispensable part of the modern drug design process. However, the calculation of absolute binding affinities for protein−ligand complexes remains as one of the main challenges in computational structure-based ligand design. At the same time, estimating differences in free energy is also central to the process of rational molecular design. This is because all equilibrium properties of a system such as phase behavior, association-dissociation constants, solubilities, adsorption coefficients and conformational equilibria depend on differences in free energy between alternative states. Free energy differences are essentially related to the relative probability of finding a system in a given microscopic state. Many empirical approaches have been developed to estimate interaction or binding free energies between proteins and ligands. However, only by using an approach that samples an appropriate thermodynamic ensemble of states, such as Molecular Dynamics (MD) and Monte Carlo (MC) simulation techniques, from which it is possible to get thermal averages over microscopic configurations at an atomic level, can differences in free energy between two states of a system be estimated directly [190,507-514]. The difficulty is that the computational cost of obtaining sufficient sampling and converged results has made the routine application of free energy calculations for estimating binding free energies impractical. This situation is, however, rapidly changing. The use of modified intermediate potentials has been shown to improve sampling dramatically and the rapid advance of computer power means that the utility of free energy calculation in molecular design must be constantly re-evaluated.

Several approaches to direct evaluation of binding affinities from molecular modeling have been developed, ranging from empirical and "'knowledge-based"' scoring functions, using end point energy difference methods such as MD MM-PB/GBSA methods [58,82,514,330] to those based on free energies calculations, such as the rigorous free energy perturbation (FEP) [465] and thermodynamic integration (TI) methods.[153,190,464]. It is known that the accuracy of MD based methods depends on sufficient sampling for convergence, as well as on the force field used. In the end point energy difference methodology, the binding energy is obtained by calculating individual components of binding energy, which are the enthalpy and

entropy. The components of binding energy are estimated through the energy difference between the bound and unbound states of the protein–ligand complexes. By contrast, the TI methods directly estimate the binding energy by integrating the differential changes in the binding energy with respect to the relative presence or absence of the ligand in the system. Although FEP and TI approaches have shown to be accurate for binding affinity calculation, they are exceedingly demanding in computational power if applied to large molecular data sets. [461,515] Thus, further development of fast and accurate methods for structure-based drug design is still needed. Åqvist et al. [147] developed a semi-empirical method, coined the linear interaction energy (LIE) method, [147,159] for absolute binding free energy calculation. This method is based on conformational sampling of receptor-bound and unbound drug states with molecular dynamics (MD) or Monte Carlo (MC) simulations. By virtue of sampling two endpoints of the process only, LIE is considerably faster than either FEP or TI simulations. However, it is considerably slower than single conformation scoring functions methods, thus presenting a very useful alternative (to FEP simulations) for already generated and curated sets of protein−drug states from docking simulations. [147] The LIE method is based on the linear response assumption [516] for electrostatic interactions combined with an empirical expression for nonpolar contributions to drug solvation and binding. The binding free energy in LIE is expressed according to equation (25):

$$\Delta G_{bind} = \beta \cdot V_{l-s}^{el} + \alpha \cdot V_{l-s}^{vdw} + \gamma \quad (25)$$

where V_{l-s}^{el} and V_{l-s}^{vdw} are MD or MC-generated interaction energy averages from the nonbonded electrostatic and van der Waals interactions of the ligand (l) with its surrounding environment (s), respectively. The symbol Δ denotes the change in average values when transferring the ligand from solution (free state) into the binding site of the solvated receptor (bound state). The coefficients α and β are scaling factors for these energy terms, while γ is an empirical constant. The linear response (LR) approximation theory provides a physical basis for treating the electrostatics contribution to the binding free energy, predicting a value of $\beta = 0.5$. The formalization of other terms related to nonelectrostatic effects in the protein−drug associations is, however, more challenging. Accordingly, the automatic workflows to facilitate the set up and execution of LIE-based binding free energy calculations for protein−ligand complexes have been developed. [517,518] The van der Waals interactions between the ligand and its environment, represented by a Lennard−Jones potential, are commonly used to calculate the nonpolar contribution to binding free energy. [147] However, an a priori prediction of α or γ parameters remains a challenge. The commonly used strategy builds on the empirical parameters fitting by available experimental data on a small set of receptor−ligand complexes and then extending it to the test sets of interest. [159,160,162,163,167,514] For example, the empirical α coefficient was initially calibrated by using $\beta = 0.5$ and four endothiapepsin inhibitors experimental binding data yielding the value of $\alpha = 0.161$. [147] Next, series of MD simulations were performed with the Gromos 96 force field furthering the proposed value of the α. [519] The original model showed reasonable predictive power for different proteins in complexes with ligands with different structural scaffolds such as endothiapepsin, [147] HIV-1 protease, [158,516] glucose binding protein, [520] and trypsin. [157] Later, this parameterization was refined by Åqvist et al., [159,514] who used 18 protein−ligand complexes comprising endothiapepsin, [147] HIV-1 protease, [158,516] glucose binding protein, [520] and trypsin [157] as the training set. The authors determined more accurate values of β by using FEP calculations. The deviation from the linear-response regime is especially pronounced for relatively polar compounds. [159,367] This resulted in an improved LIE model known as the standard model, referred here as LIE-S. This model uses β values, ranging between 0.33 and 0.5, along with $\alpha = 0.18$ and $\gamma = 0$. Further studies have shown that binding free energies calculated with LIE-S are in good agreement with experimental data for several protein−ligand systems. [168,521,522,523,524]

Nevertheless, for proteins containing hydrophobic binding sites, a non-zero γ constant is required to reproduce observed trends in the absolute binding free energies. The commonly used value of $\alpha = 0.18$ downscales the nonpolar contribution ($\alpha\Delta\langle V_{l-s}^{vdw}\rangle$), thus leading to a significant underestimation of binding energies. [160,163,167] Some notable cases where LIE fails to predict are to be found in binding of retinoids to retinol binding protein (RBP), [167] biotin analogs to avidin, [160] substrates to cytochrome P450 (P450cam), [163] and inhibitors to human thrombin. [525] For these systems, reported γ values range from −2.9 to −7.0 kcal/mol. [162] Almlöf et al. [168] found a clear relationship between the ranking of these binding sites hydrophobicity (RBP > P450cam > thrombin > trypsin) and the γ value. To some extent this is similar to the idea developed by Wang et al., [160] who investigated variations of the nonpolar coefficient α in the absence of the constant term γ, as a way to distinguish among different binding-site types. The main outcome of their research is the linear correlation ($R^2 = 0.96$) obtained between the weighted nonpolar desolvation ratio (WNDR) and the α values in the LIE method. [160] Valiente et al. [526] established a linear relationship ($R^2 = 0.85$) between WNDR and γ using Wang's training set complexes. This novel model, termed LIE-C, was successfully applied to predict the binding free energy of five PlmII-Inhibitor complexes ($\langle|error|\rangle = 1.47$ kcal/mol). [526] Nevertheless, the main disadvantage of LIE-C is the use of atomic desolvation parameters for WNDR calculation. Furthermore, the only six atom types (C, S, N, N^+, O, and O^-) considered in that report [526] exclude other heteroatoms (Fe, Zn, Cu, F, Br, I, etc.) that could also be present in protein−ligand complexes.

Miranda et al. developed LIE-D, a novel LIE parameterization model that accurately predicts the γ coefficient based on the balance between polar and nonpolar contributions to binding free energy (D parameter) extracted from MD simulations. Leave-one-out assessment showed that LIE-D accurately reproduced our training data set experimental binding free energies. The model robustness was demonstrated by reproducing the binding free energies of two of the three protein−ligand sets outside the training data set, using the reported electrostatic and van der Waals interaction energies

calculated with different force fields. Thus, LIE-D can be useful for lead optimization phases where computational methodologies more accurate than scoring functions will be needed to predict absolute binding free energies of protein−ligand complexes. However, to conclude which force field is the best choice for free energy calculations with the LIE-D model, further studies will be needed. [527]

A study for the binding of seven biotin analogues to avidin suggested that to obtain statistically converged MM-GBSA results, several independent simulations each with sampling times of 20-200 ps (averaging the results) is more effective than a single long simulation [252]. 'Single-trajectory' simulations of the complex are generally sufficiently accurate for most applications, and while MD simulation length does have an obvious impact on the accuracy of predictions, longer MD simulations doesn't necessarily mean better predictions [84]. For the calculations of the ΔE_{MM} and ΔG_{solv} terms (equation 20), a large ensemble (e.g. several hundred) conformations are typically extracted in small intervals from the single MD trajectory of the complex. Alternatively, averaging over a select few receptor-ligand binding conformations from the MD trajectory via clustering has proved effective, as well as more time efficient [478]. MM-GB(PS)SA calculations on single (minimized) structures has also been proposed and validated [83,528], but not necessarily for structures generated from MD simulations. Meanwhile, for the entropy term calculated using normal mode analysis, fewer snapshots (typically less than a 100) are employed, due to the computational cost involved. As a larger number of snapshots may be required for more stable and accurate predictions, these calculations, however, are computationally expensive and often not feasible with limited computational resources. Consequently, neglect of the entropy term can in some cases lead to sufficient or more accurate predictions for ranking of ligand binding affinities in certain macromolecular systems [84,478,528].

The MM/PBSA method [58,205] was introduced by Srinivasan et al. [330] It combines molecular mechanics (MM) and continuum solvent approaches to estimate binding energies. An initial MD simulation in explicit solvent provides a thermally average ensemble of structures. Several snapshots are then processed, removing all water and counterion molecules, and used to calculate the total binding free energy of the system with the equation

$$\Delta G_{bind} = \bar{G}_{complex} - \left[\bar{G}_{protein} + \bar{G}_{ligand}\right] \quad (26)$$

where the average free energy $\bar{G}$ of the complex, protein, and ligand, are calculated according to the following equations:

$$\bar{G} = \bar{E}_{MM} + \bar{G}_{solvation} + T\bar{S} \quad (27)$$

$$\bar{E}_{MM} = \bar{E}_{\text{int}} + \bar{E}_{elec} + \bar{E}_{vdw} \quad (28)$$

$$\bar{G}_{solvation} = \bar{G}_{polar} + \bar{G}_{non-polar}. \quad (29)$$

$\bar{E}_{MM}$ is the average MM energy in the gas phase, calculated for each desolvated snapshot with the same MM potential used during the simulation but with no cut-offs. $\bar{G}_{solvation}$, the solvation free energy, is calculated in two parts, the electrostatic component $\bar{G}_{polar}$ using a Poisson–Boltzmann approach, and a non-polar part using the solvent-accessible surface area (SASA) model. [529] The entropy ($T\bar{S}$) is the most difficult term to evaluate; it can be estimated by quasi-harmonic analysis [485,530,531] of the trajectory or using normal mode analysis. [330,530,531] The entropy change can be omitted if only the relative binding energies of a series of structurally similar compounds is required, but if the absolute energy is important, or if the compounds are notably different, then its contribution to the final free energy cannot be ignored. A quite recent study by Kuhn et al. [83] suggests that the MM-PBSA function could be used as a post-docking filter during the virtual screening of compounds, as their use of a single relaxed structure provided better results than usual averaging over MD simulation snapshots. However, as the simulation conditions used in this work were not optimal, improved calculations could lead to significantly different conclusions.

Although only a single MD simulation of the complex is commonly used to determine the conformational free energy, [161] as the structures for both the free ligand and ligand-free protein molecules are extracted from the simulation for the protein–ligand complex, this approach might not be the best, as a study by Pearlman [484] showed that using a single simulation to generate all structures for a series of complexes of p38 MAP kinase and 16 different ligands provides final results that are significantly worse than those from separate simulations, and that savings achieved in computing time are minimal and do not justify the simplification.

Application of the MM-PBSA approach has produced reasonable binding energies for several systems, [330,532,533,534] but not for others. [484] Evaluation of the MM-PBSA method using a series of p38 MAP kinase complexes resulted in very poor results compared with other approaches, and at an appreciably larger computational cost. [484] Kollman and co-workers [151] presented a combined approach that implements docking, MD simulations, and MM-PBSA, and used it to predict the binding mode of the inhibitor efavirenz to HIV-1 reverse transcriptase. Initially, they evaluated the capacity of combined MD simulations and MM-PBSA to reproduce binding free energies of 12 crystal structures ofHIV-1 RT complexed with different TIBO-like inhibitors. They found that both relative and absolute free energies were correctly predicted with an error of 1.0 kcal/mol. For the docking of efavirenz, five different binding modes were submitted to MD simulation and further processed using the MM-PBSA approach. The most stable binding mode was clearly identified, with a binding free energy of -13.2 kcal/mol in good agreement with the experimental value of -11.6 kcal/mol. The final structure was found to be in very good agreement with a crystal structure of the complex, not initially available to the authors. They concluded that molecular docking combined with MD simulations followed by MM-PBSA analysis presented a reasonable approach for modeling protein complexes a priori.

Ahmed et al. utilized a much longer MD simulations because certain conformational changes are not accessible at the conventional MD time scale. In addition to MD simulations, they carried out binding free energy calculations based either on a single snapshot (Prime-MM/GBSA) or

ensemble of snapshots (AMBER-MM/PB(GB)SA). To account for the protein polarization effect, they have also rescored the relaxed MD complexes using QM/MM rescoring.

For those complexes that are subjected to the MD simulations, physics based rescoring was carried out applying both MM and QM/MM techniques. For MM rescoring, they applied a single, end point rescoring for the final trajectory snapshot produced from the MD simulations and using the Prime/MM-GBSA in the Schrodinger suite [61,535,536]. The complexes were also rescored applying the much more sophisticated AMBER/GB(PB)SA module in AMBER12 [537,538]. The AMBER/GB(PB)SA module takes the advantage of statistical averaging over many potential configurations. For the QM/MM rescoring and because of the prohibitive computational expense, in DFT/MM rescoring they use only the ligand as the QM subsystem as implemented in the QM/MM-PBSA script of Schrodinger and solvation effect is ignored in that case. These procedures seemed to provide plausible explanations for the observed inhibitory effects.

Liu et al. have proposed an MM/PBSA-based free energy estimator (PBSA_E) for efficient prediction of protein−ligand binding affinity. The method involves MM/PBSA calculation of the protein−ligand binding energy using a single protein−ligand complex configuration that is optimized from the crystal structure, and no MD simulation is needed. The calculated PBSA energies were fitted to a training set to obtain optimized coefficients for use in a formula (equation 29) to give the free energy prediction (PBSA_E). The performance of this method was validated on two test sets. Explicit comparison with three popular scoring function methods (GlideSP, GlideXP, and Sybyl_F) demonstrated that the PBSA_E is superior to all three of these methods in reliability and accuracy, but it takes more computational effort. The study also showed that using a single optimized protein−ligand configuration is preferred over using multiple configurations generated from MD simulations.

$$PBSA - E = a_1\left(\Delta E_{ele} + \Delta G_{pb}\right) + a_2\Delta E_{vdw} + a_3\Delta G_{pbsur} + a_4 N_{rot} \quad (30)$$

It is useful to give a brief discussion of their computational timing. The single-configuration MM/PBSA calculation normally takes 1−2 min for each protein system on one CPU of a standard workstation. If optimization of the complex structure is done in explicit water, it generally takes about 20 min on an eight-CPU workstation (Intel Xeon E5620 2.4 GHz processor). The long optimization time is mostly due to optimization of water molecules. However, their study showed that if this optimization is carried out either in the gas phase or using an implicit water model (the GB model), the optimization time could be cut to under 2 min. It should also be mentioned that newer versions of the currently used scoring functions could achieve better correlations than the ones obtained in their study, [539] and also using MM-GBSA and MM-PBSA methods, relative binding affinities for a set of ligands to a given target can often be reproduced with good accuracy and considerable less computational effort compared to full-scale molecular dynamics FEP/TI simulations. Furthermore, free-energies can be decomposed into insightful interaction and desolvation components [478,482,540].

The EEMD method (expanded ensemble molecular dynamics), successfully used earlier for accurate calculation of solvation energies for organic molecules [541-543] has also been used for calculating binding energies. Using this method it is also possible to obtain the thermodynamic components of binding energy which has not been attempted using the TI methodology (details of MD end point energy difference and TI methods can be found elsewhere [482,544,545]).

The EEMD simulations face a complication that, during the sampling in subensembles, there is a risk that the ligand might drift away from the active site of the protein. This occurs when the partial representation of the ligand no longer contains the key interactions that are necessary for its binding to the active site residues. Once the ligand drifts away from the active site of the protein, the energy estimations might no longer represent the binding energy of the ligand. In order to avoid this complication, the ligand needs to be restrained in the active site of the protein, which can be done. [546]

Zhao et al. proposed MD as a tool to analyze the free-energy surface and pathways of (un)binding of small molecules from/to proteins [547,548]. The network analysis revealed multiple binding modes characterized by distinct intermolecular hydrogen bonds and hydrophobic contacts. Moreover, the unbinding kinetics showed single-exponential time dependence which indicates that the barrier for full dissociation is significantly higher than the barriers between different binding modes. It is instructive to compare experimental and simulation approaches. The aforementioned biophysical techniques for the analysis of fragment binding to proteins have limitations in temporal and/or spatial resolution. In contrast, the MD simulations of (un)binding generate a complete picture of the free-energy surface and (un)binding pathways at atomic level of detail [114,547,548].

Recently, motivated by the atomic force microscopy (AFM) experiments the steered molecular dynamics (SMD) [549,550] has developed and applied to study mechanical unfolding of biomolecules [551,552], transportation of ions [553,554] and organic compounds through membrane channels [555,556]. This method is also employed to probe unbinding pathways of ligand from its receptor [557,558]. Li and Mai published that SMD is one of possible candidates to achieve two goals of drug design: more accurate than docking and more efficient than MD methods. Application of basic concepts of the SMD to study ligand binding affinity of various complexes shown SMD to be as powerful as MM-PBSA method in predicting binding affinity but about one order of magnitude faster.

Inter- and intra-molecular forces are key to the stability of biomolecules. Up to now understanding of these forces is possible through indirect physical and thermodynamical measurements like crystallography, light scattering, nuclear magnetic resonance spectroscopy etc. In the case of ligand binding to receptor the binding energy E_{bind} is estimated via equilibrium association constant K_a constants using the equation E_{bind} =.$-RT\ ln(K_a)$. Single molecule force spectroscopy experiments such as AFM

[559], laser optical tweezers [560] and magnetic tweezers [561] can directly probe molecular forces and provide unexpected insights into the strength of forces driving various interactions responsible for the mechanical stability of bio-systems.

In standard single molecule force spectroscopy experiments one terminal (end) of a biomolecule is anchored to a surface, while another terminal is attached to a force sensor. A biomolecule is stretched by increasing the distance between the surface and the force sensor. In the AFM case, the force sensor is a micron-sized cantilever. After the pioneering AFM experiment of Gaub *et al.* [562], a lot of experimental as well as theoretical works have been done on proteins, DNA and RNA [563,564]. Biomolecules are pulled either with a constant force or by a force ramped with a constant loading rate. The force-extension curve of proteins under constant rate pulling has a saw-tooth shape due to domain by domain unfolding [565]. Schulten's group [327] was first to reproduce this remarkable result by SMD simulations.

To probe the binding affinity of ligand to receptor, one applies the external force to pull ligand from the binding pocket without fixing any atom of ligand (Figure 5). As in the single molecule force spectroscopy the force experienced by ligand is proportional to the displacement of the cantilever. Under the force loaded with a constant rate the total energy of the receptor-ligand complex is as follows

$$E = E_{receptor} + E_{ligand} + E_{receptor-ligand} + E_{force} \quad (31)$$

$$E_{force} = \frac{k}{2}(x - vt)^2 \quad (32)$$

Here $E_{receptor}$, E_{ligand} and $E_{receptor\text{-}ligand}$ are energies of receptor, ligand and receptor-ligand interaction, respectively.

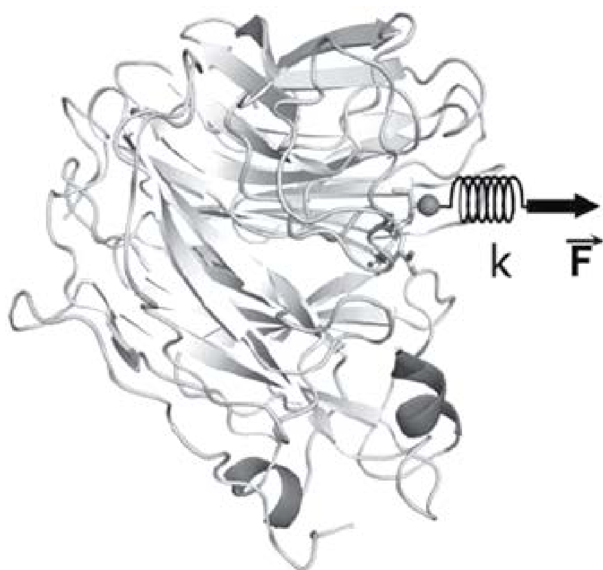

Figure 5. Schematic plot for pulling ligand from a receptor by applying the force $\vec{F}$ via the cantilever with the spring constant k [566]

Therefore, it is important to establish what pulling path should be chosen. If biomolecule is stretched with one point fixed, the force is directed along the vector connecting the anchored point and the point which the force is applied to. The situation becomes more complicated when a ligand is pulled from the binding site not keeping any atom fixed. In this case, one can not pull along arbitrary direction and an optimal pathway can be chosen [567] using, e.g. Caver 2.0 [568], a plugin of Pymol. The optimal path goes through the widest tunnel which is directed outside the binding pocket. The corresponding rupture force would be smallest among all possible pathways.

The Linear Interaction Energy (LIE) methods [147] are end-point method like MM-PBSA/GBSA and also use averaged conformations from molecular dynamics simulations. The binding free energy is estimated as:

$$\Delta G_{bind} \approx \alpha\left(E_{elecbound} - E_{elec\,free}\right) + \beta\left(E_{vdwbound} - E_{vdw\,free}\right) \quad (33)$$

where the brackets denote averages from molecular dynamics trajectory. The factors α and β account for changes in the internal energy of the solvent and protein and are determined empirically [147]. This method has been shown to give accurate results [147] and newer implementation include a solvation energy term to increase accuracy [569]. A drawback of this method is that there is no universal value for the factors α and β, they have to be determined independently for each case and require experimental data.

Van Vuong et al. presented a new method for finding the optimal path for pulling a ligand from the binding pocket using steered molecular dynamics (SMD), defining scoring function as the steric hindrance caused by a receptor to ligand movement. Then the optimal path corresponds to the minimum of this scoring function, calling the new method MSH (Minimal Steric Hindrance). Contrary to existing navigation methods, their approach takes into account the geometry of the ligand while other methods including CAVER only consider the ligand as a sphere with a given radius. Using three different target + receptor sets, they had shown that the rupture force F_{max} and nonequilibrium work W_{pull} obtained based on the MSH method show a much higher correlation with experimental data on binding free energies compared to CAVER. Furthermore, W_{pull} was found to be a better indicator for binding affinity than F_{max}. Thus, they claimed the new MSH method as a reliable tool for obtaining the best direction for ligand exiting from the binding site. Its combination with the standard SMD technique can provide reasonable results for ranking binding affinities using W_{pull} as a scoring function. [570]

7. Conclusions

An accurate knowledge of interactions in biomolecules is of utmost importance for understanding of their structure and function and high quality calculations have a potential to provide important insights in this area. In addition, modeling of biomolecular structure and dynamics profits from accurate reference values needed for calibration of less demanding quantum chemical, semiempirical, and empirical methods used in molecular dynamics simulations. Typical size of biomolecules is,

however, often relatively large and out of reach of the current WFT benchmark methods. Therefore, reference calculations on biomolecular complexes are challenging mainly because of the size that is required for obtaining useful results. For these types of problems, QMC calculations may offer a significant advantage.

A number of studies have shown that refining docking and scoring calculations by performing molecular dynamics (MD) and free energy calculations starting from docked poses can greatly increase the accuracy of binding affinity predictions [266,571,572]. Due to the much more elaborate procedure and the simulation time needed for each compound, only a small set of compounds, up to ~50, can be predicted at the same time. This scheme is therefore only useful in the lead optimization stage, where accurate binding affinity predictions of a smaller set of similar compounds are needed for selection of compounds to be synthesized and for rationalization of particularly interesting interactions between the compounds and the binding site. The improved accuracy of the simulations is mainly due to the increased level of molecular detail, using a flexible and explicitly solvated protein. Problems do remain, such as the restriction of sampling time, as no major structural changes will take place during the simulation time that can realistically be used for efficient binding affinity predictions. In connection with this is the importance of using accurate starting structures, which has been reported in a number of studies [266,572,573,574].

MM-GBSA and MM-PBSA are computationally efficient, end-point free energy methods that have been widely used to study protein-ligand binding affinities. Even though they lack the sound theoretical foundations of recently developed computationally demanding absolute-affinity free-energy methods [322,461,462,470], their connection with statistical thermodynamics has been established [161]. Due to the approximations inherent in MM-GB(PB)SA methods, they are more applicable for ranking ("scoring") of ligand binding affinities rather than to quantitatively predicted absolute binding free energies. They should be regarded as approximate, as they combine a molecular mechanics energy function with a continuum-electrostatics treatment of solvation effects; they include solute conformational entropy effects in an approximate manner [477]; and ignore the solvent molecular structure. Accurate incorporation of solute entropy [575] and solvent effects in binding affinity calculations is challenging, but future extensions and development of MM-GB(PB)SA methods will undoubtedly serve to address these limitations. It is important to point out that the binding energy scores for various sets of compounds under study at different levels of QM/MM and pure-MM levels of theory allow Ahmed et al. to conclude that, in general, MM based methods perform well in comparison with the QM/MM based methods. [576]

MM-GBSA and MM-PBSA are computationally efficient, end-point free energy methods that have been widely used to study protein-ligand binding affinities. Even though they lack the sound theoretical foundations of recently developed computationally demanding absolute-affinity free-energy methods [322,461,462,470], their connection with statistical thermodynamics has been established [161]. Due to the approximations inherent in MM-GB(PB)SA methods, they are more applicable for ranking ("scoring") of ligand binding affinities rather than to quantitatively predicted absolute binding free energies.

All-atom MD simulations of pharmaceutically relevant protein targets in their apo state or in complex with (putative) ligands have provided useful information at the beginning and final phase, respectively, of high-throughput docking campaigns. Explicit solvent MD runs of the apo structure are becoming more frequent for the selection of one or more snapshots for docking of large libraries of compounds. At the final stage of ranking, MD simulations starting from the predicted binding mode are performed for the in silico validation of the top ranking compounds. Moreover, atomistic MD simulations have provided insights into the mechanism of the antiviral drug darunavir (a dimerization inhibitor of the HIV protease); have helped in the interpretation of the selectivity profile of two kinase inhibitors used in the clinics as anticancer drugs; and have revealed that acetyl-lysine (the natural ligand of bromodomains) has an alternative binding mode which is more buried than the one observed in the available crystal structures [577]. These and several other examples (some of which are reviewed in [578]) suggest that the applications of MD will play an even more important role in drug design in the future. There are three main obstacles to the identification of potent (i.e., nanomolar) inhibitors by high-throughput docking: the very small chemical space of the libraries of compounds, the approximations used for scoring, and the use of a (mainly) rigid protein structure [579]. The future developments of MD-based methods for in silico screening will help to remove the two latter obstacles.

The accuracy of free energy calculations depends on two factors. First, the accuracy of the force field. This can only be confirmed by comparison to experimental results. Second, free energy calculations depend on the degree of sampling and convergence. This is in principle independent of whether we know the 'true'.

The force field might be responsible for not being able to correctly discriminate between inhibitors with similar binding free energies (with a difference in the order of a few kJ/mol) [580]. However, when dealing with force field inaccuracies, errors should be systematic. Although the ranking of the inhibitors is expected to be different, the spread of the values should be similar in the experimental values and the values calculated from the simulations.

Pieffet results [580] suggest clearly that in fact nanosecond simulations are too short to yield accurate estimates of the binding free energies in cases relevant to drug design and that closures of thermodynamic cycles are useful but by no mean sufficient to ensure that convergence has been reached. They can also suggest that the different orientations of the inhibitors inside the binding pocket of the receptor might need to be taken into account and not only the most favorable one. In this case the correct distribution would be required and much longer simulations would be needed.

The discrepancy between the values obtained from experiment and the values calculated from the simulations can have different origins. The force field is potentially the primary source of error. The accuracy of the force field can vary depending on the type of residue under consideration [581]. For this reason it is not straightforward to quantify the accuracy of a force field in

a global sense [582]. Furthermore, the results observed in various tests can be greatly influenced by the protocols used to perform the simulations and these protocols might be different from the one used initially for the force field parameterization.

In the case of the calculation of the (non-physical) free energy associated with the mutation of a residue into another one, it is clear that an accurate description of both the wild type and the mutant protein is required to obtain the correct answer. The solvation free energies for analogs of several residues involved in the mutations have been shown to be quite accurate in different environments [581] using the GROMOS96 force field. The solvation free energies of analogs of hydrophobic amino acids such as Ala, Val and Leu are within 2 kJ/mol of the experimental value in both water and cyclohexane. Thus, from a force field perspective the mutations of Val or Leu into Ala were expected to perform well irrespective of whether the residues are exposed to solvent or buried inside the protein. However, the results show discrepancies with experiment much larger than 2 kJ/mol and are randomly distributed clearly indicating that the major source of error is not directly related to the accuracy of the force field for these specific amino acids. [580]

The accuracy of the force field can also have indirect effects on the result. The force field determines which conformations are sampled during the simulations and therefore determines the local environment around the mutation site. The force field must be able to yield the correct structure if the free energy associated with a given mutation is to be estimated correctly.

The accuracy of the starting structures used for the simulations is also a crucial factor affecting the reliability of the calculations. This is especially true if the relaxation of the protein towards its lowest free energy conformations is not possible within the time scale of the simulation. If the starting structure was incorrect it would be expected to deviate significantly during the course of a MD simulation. Since the results of free energy calculations are directly related to the local interactions around the mutation site it is extremely important that the structure remains close enough to the native structure so that the local environment around the mutation site is appropriate.

Numerical errors can also arise from the integration over a small number of discrete values of the free energy derivative. Integrating using the trapezoidal rule which corresponds to using a linear interpolation scheme gives similar results as using a cubic spline interpolation scheme. [580]

Free energy calculations in proteins using the thermodynamic integration method still remain extremely difficult even for very simple mutation such as the transformation of Leu into Ala. Although the size of the mutation seems to be related to the discrepancy observed between simulation and experiment, the results obtained for other mutations suggests that other factors such as the rigidity of the side-chain might help convergence. Mutations of residues associated with a change in net charge whether positive or negative do not give accurate results due to the large amount of work done against the system. [580]

8. Future Perspectives

In this review, there has been highlighted some of the main limitations of current docking programs and methodologies. Basically, none of them is perfect but, collectively, it is true that most of the key challenges in protein-ligand docking seem to be correctly addressed in the most used programs. Therefore, while waiting for the perfect program, it must be encouraged additional efforts to extend the use of docking tools beyond the confines of bioinformatic groups. Undoubtedly, challenges still remain, especially for issues involving the accuracy of the available scoring functions, which are in fact classical approximations of events ruled by quantum mechanics. Most molecular docking programs successfully predict the binding modes of small-molecule ligands within receptor binding sites. However, the current algorithms do not estimate the absolute energy associated with the intermolecular interaction with satisfactory accuracy. The appropriate handling of issues such as solvent effects, entropic effects, and receptor flexibility are major challenges that require attention. Successful molecular docking protocols require a solid knowledge of the fundamentals of the applied methods. Understanding these principles is essential in the production of meaningful results.

Molecular docking has several strengths, among which the method's ability to screen large compound databases at low cost compared to experimental techniques such as HTS is particularly notable. In the current panorama of drug discovery, where high attrition rates are a major concern, properly designed VS strategies are time-saving, cost-effective and productive alternatives. Existing methods, which often rely on expertise knowledge and thus may have limited applications in real practice. A universally accurate and reliable solution is still far from reach in the near future. Revolutionary innovations are definitely in urgent need and thus highly encouraged to address the fundamental challenges such as target flexibility and water molecules.

Competing Interests

The author declare no competing financial interest.

Acknowledgement

The autor thanks the MCINN Project Consolider Ingenio CSD2010-00065 for financial support.

References

[1] Vindigni, A. Energetic dissection of specificity in serine proteases. *Comb. Chem. High Throughput Screen,* 1999, 2, 139-153.

[2] Cheng, A. C.; Calabro, V.; Frankel, A. D. Design of RNA binding proteins and ligands. *Curr. Opin. Struct. Biol.,* 2001, 11, 478-484.

[3] Garvie, C. W.; Wolberger, C. Recognition of specific DNA sequences. *Mol. Cell.,* 2001, 8, 937-946.

[4] Autumn, K.; Sitti, M.; Liang, Y. A.; Peattie, A. M.; Kenny, T. W.; Fearing, R.; Israelachvili, J. N.; Hansen, W. R.; Sponberg, S.; Full,

R. J. Evidence for van der Waals adhesion in gecko setae. *Proc. Natl. Acad. Sci. U. S. A.*, 2002, 99, 12252-12256.

[5] Riley, K. E.; Hobza, P. On the importance and origin of aromatic interactions in chemistry and biodisciplines. *Acc. Chem. Res.*, 2013, 46, 927-936.

[6] Grimme, S. Supramolecular binding thermodynamics by dispersion-corrected density functional theory. *Chem. - Eur. J.*, 2012, 18, 9955-9964.

[7] Juríček, M.; Strutt, N. L.; Barnes, J. C.; Butterfield, A. M.; Dale, E. J.; Baldridge, K. K.; Stoddart, J. F.; Siegel, J. S. Induced-fit catalysis of corannulene bowl-to-bowl inversion. *Nat. Chem,*. 2014, 6, 222-228.

[8] Liu, W.; Tkatchenko, A.; Scheffler, M. Modeling adsorption and reactions of organic molecules at metal surfaces. *Acc. Chem. Res.,* 2014, 47, 3369-3377.

[9] Kronik, L.; Tkatchenko, A. Understanding molecular crystals with dispersion-inclusive density functional theory: pairwise corrections and beyond. *Acc. Chem. Res*., 2014, 47, 3208-3216.

[10] Sponer, J.; Sponer, J. E.; Mládek, A.; Jurečka, P.; Banáš, P.; Otyepka, M. Nature and magnitude of aromatic base stacking in DNA and RNA: quantum chemistry, molecular mechanics and experiment. *Biopolymers*, 2013, 99, 978.

[11] Riley, K. E.; Hobza, P. Noncovalent interactions in biochemistry. *Comp. Mol. Sci.*, 2011, 1, 3-17.

[12] Murthy, P. S. Molecular handshake: recognition through weak noncovalent interactions. *J. Chem. Educ*., 2006, 83, 1010.

[13] Rubbiani, R.; Can, S.; Kitanovic, I.; Alborzinia, H.; Stefanopoulou, M.; Kokoschka, M.; Mönchgesang, S.; Sheldrick, W. S.; Wölfl, S.; Ott, I. Comparative in vitro evaluation of N-heterocyclic carbene gold(I) complexes of the benzimidazolylidene type. *J. Med. Chem*., 2011, 54, 8646-8657.

[14] Meyer, A.; Bagowski, C. P.; Kokoschka, M.; Stefanopoulou, M.; Alborzinia, H.; Can, S.; Vlecken, D. H.; Sheldrick, W. S.; Wölfl, S.; Ott, I. On the biological properties of alkynyl phosphine gold(I) complexes. *Angew. Chem., Int. Ed*., 2012, 51, 8895-8899.

[15] Beno, B. R.; Yeung, K.-S.; Bartberger, M. D.; Pennington, L. D.; Meanwell, N. A. A Survey of the role of noncovalent sulfur interactions in drug design. *J. Med. Chem*., 2015, 58, 4383-4438.

[16] Riley, K. E.; Pitoňák, M.; Jurečka, P.; Hobza, P. Stabilization and structure calculations for noncovalent interactions in extended molecular systems based on wave function and density functional theories. *Chem. Rev.*, 2010, 110, 5023-5063.

[17] Georgakilas, V.; Otyepka, M.; Bourlinos, A. B.; Chandra, V.; Kim, N.; Kemp, K. C.; Hobza, P.; Zbořil, R.; Kim, K. S. Functionalization of graphene: covalent and non-covalent approaches, derivatives and applications. *Chem. Rev*., 2012, 112, 6156-6214.

[18] DiStasio, R. A.; Gobre, V. V.; Tkatchenko, A. Many-body van der Waals interactions in molecules and condensed matter. *J. Phys.: Condens. Matter*, 2014, 26, 213202.

[19] Schneider, G.; Bohm, H. J. Virtual screening and fast automated docking methods. *Drug Discov. Today*., 2002, 7, 64-70.

[20] Shoichet, B. K.; Leach, A. R.; Kuntz, I. D. Ligand solvation in molecular docking. *Proteins.*, 1999, 34, 4-16.

[21] Chothia, C. Hydrophobic bonding and accessible surface area in proteins. Nature, 1974, 248, 338-9.

[22] 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.

[23] Gohlke, H.; Klebe, G. Approaches to the description and prediction of the binding affinity of small-molecule ligands to macromolecular receptors. *Angew. Chem., Int. Ed*., 2002, 41, 2645-2676.

[24] Homans, S. W. Water, water everywhere-except where it matters? *Drug Discovery today*, 2007, 12, 534-539.

[25] Jeffrey, G.; Saenger, W. Hydrogen bonding in biological structures. 1991, Springer, Berlin.

[26] Stahl, M.; Bohm, H. J. Development of filter functions for protein-ligand docking. *J. Mol. Graph. Mod.*, 1998, 16, 121-132.

[27] Brady, G. P.; Sharp, K. A. Entropy in protein folding and in protein-protein interactions. *Curr. Opin. Struct. Biol* ., 1997, 7, 215-21.

[28] Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymatic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. USA*, 1971, 68, 1678-83.

[29] Boehr, D. D.; Nussinov, R.; Wright, P. E. The role of dynamic conformational ensembles in biomolecular recognition. *Nat. Chem. Biol*., 2009, 5, 789-796.

[30] Risthaus, T.; Grimme, S. Benchmarking of london dispersion-accounting density functional theory methods on very large molecular complexes. *J. Chem. Theory Comput.*, 2013, 9, 1580-1591.

[31] Sure, R.; Grimme, S. Comprehensive benchmark of association (free) energies of realistic host-guest complexes. *J. Chem. Theory Comput*., 2015, 11, 3785-3801.

[32] Cao, L.; Skalamera, D.; Zavalij, P. Y.; Hostaš, J.; Hobza, P.; Mlinarič-Majerski, K.; Glaser, R.; Isaacs, L. Influence of hydrophobic residues on the binding of CB[7] toward diammonium ions of common ammonium·ammonium distance. *Org. Biomol. Chem*., 2015, 13, 6249-6254.

[33] Lee, J.; Sorescu, D. C.; Deng, X.; Jordan, K. D. Water chain formation on TiO_2 rutile(110). *J. Phys. Chem. Lett*., 2013, 4, 53-57.

[34] Bartels, L. Tailoring molecular layers at metal surfaces. *Nat. Chem*., 2010, 2, 87-95.

[35] Auwärter, W.; Écija, D.; Klappenberger, F.; Barth, J. V. Porphyrins at interfaces. *Nat. Chem.*, 2015, 7, 105-120.

[36] Vítek, A.; Ofiala, A.; Kalus, R. Thermodynamics of water clusters under high pressures. A case study for $(H_2O)_{15}$ and $(H_2O)_{15}CH_4$. *Phys. Chem. Chem. Phys.*, 2012, 14, 15509-15519.

[37] Chattoraj, J.; Risthaus, T.; Rubner, O.; Heuer, A.; Grimme, S. A multi-scale approach to characterize pure CH_4, CF_4, and CH_4/CF_4 mixtures. *J. Chem. Phys*., 2015, 142, 164508.

[38] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical quantum chemical PM6 method augmented by dispersion and h-bonding correction terms reliably describes various types of noncovalent complexes. *J. Chem. Theory Comput.,* 2009, 5, 1749-1760.

[39] Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-cost quantum chemical methods for noncovalent interactions. *J. Phys. Chem. Lett.,* 2014, 5, 4275-4284.

[40] Grimme, S.; Brandenburg, J. G.; Bannwarth, C.; Hansen, A. Consistent structures and interactions by density functional theory with small atomic orbital basis sets. *J. Chem. Phys.*, 2015, 143, 054107.

[41] Grimme, S. Density functional theory with London dispersion corrections. *WIREs*, 2011, 1, 211-228.

[42] Waller, M. P.; Kruse, H.; Muck-Lichtenfeld, C.; Grimme, S. Investigating inclusion complexes using quantum chemical methods. *Chem. Soc. Rev*., 2012, 41, 3119-3128.

[43] Chałasiński, G.; Szczeşńiak,M.M. State of the art and challenges of the ab initio theory of intermolecular interactions. *Chem. Rev.,* 2000, 100, 4227-4252.

[44] Pykal, M.; Jurečka, P.; Karlický, F.; Otyepka, M. Modelling of graphene functionalization. *Phys. Chem. Chem. Phys*., 2016, 18, 6351-6372.

[45] Hobza, P.; Mű́ller-Dethlefs, K. Non-covalent interactions; *RSC Theoretical and Computational Chemistry Series*; RSC Publishing, 2010.

[46] Lazar, P.; Karlický, F.; Jurečka, P.; Kocman, M.; Otyepková, E.; Safářová, K.; Otyepka, M. Adsorption of small organic molecules on graphene. *J. Am. Chem. Soc*., 2013, 135, 6372-6377.

[47] Karlický, F.; Otyepková, E.; Banáš, P.; Lazar, P.; Kocman, M.; Otyepka, M. Interplay between ethanol adsorption to high-energy sites and clustering on graphene and graphite alters the measured isosteric adsorption enthalpies. *J. Phys. Chem.* C, 2015, 119, 20535-20543.

[48] Stone, A. J. The theory of intermolecular forces. Calerdon Press, Oxford, 2002.

[49] Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation theory approach to intermolecular potential energy surfaces of van der waals complexes. *Chem. Rev*., 1994, 94, 1887-1930.

[50] Bale, T. L.; Vale, W. W. CRF and CRF receptors: role in stress responsivity and other behaviors. *Annu. Rev. Pharmacol. Toxicol*., 2004, 44, 525-557.

[51] López de Victoria A, Tamamis P, Kieslich CA, Morikis D. Insights into the structure, correlated motions, and electrostatic properties of two HIV-1 gp120 V3 loops., *PLOS ONE* 2012, 7, e49925.

[52] Lou J, Smith RJ. Modelling the effects of adherence to the HIV fusion inhibitor enfuvirtide. *J. Theor. Biol*., 2011, 268, 1-13.

[53] Cosconati, S.; Marinelli, L.; Lavecchia, A.; Novellino, E. Characterizing the 1,4-dihydropyridines binding interactions in the l-type ca^{2+} channel: model construction and docking calculations *J. Med. Chem.,* 2007, *50*, 1504-1513.

[54] Storici, P.; De Biase, D.; Bossa, F.; Bruno, S.; Mozzarelli, A.; Pene_, C. Structures of γ-aminobutyric acid (GABA) aminotransferase, a pyridoxal 5′-phosphate, and [2Fe-2S] cluster-containing enzyme, complexed with γ-ethynyl-GABA and with the antiepilepsy drug vigabatrin. *J. Biol. Chem.,* 2004, *279*, 363-373.

[55] Huey, R.; Morris, G. M. Using AutoDock 4 with AutoDocktools: a tutorial. La Jolla, CA, USA: The Scripps Research Institute, Molecular Graphics Laboratory, 2008, pp. 54-56.

[56] Davood, A.; Nematollahi, A.; Iman, M.; Shafiee, A. Synthesis and docking studies of new 1,4-dihydropyridines containing 4-(5)-Chloro-2-ethyl-5-(4)-imidazolyl substituent as novel calcium channel agonist. *Arch. Pharm. Res.,* 2009, *32*, 481-487.

[57] O'Reilly, A. O.; Khambay, B. P. S.; Williamson, M. S.; Field, L. M.; Wallace, B. A. Modelling insecticide-binding sites in the voltage-gated sodium channel. *Biochem. J.,* 2006, *396*, 255-263.

[58] 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. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res*., 2000, 33, 889-897.

[59] Case, D. A.; Cheathham, T. E., III; Darden, T.; Gohlke, H.; Luo, R.; Merz, K. M., Jr.; Onufriev, A.; Simmerling, C.; Wang, B., Woods, R. J. Comput. Chem., 2005, 26, 1668-1688.

[60] 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.

[61] 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/MM-GBSA calculation. *J. Comput. Chem*., 2011, 32, 2800-2809.

[62] Zheng, F.; Zhan, C.-G. Computational modeling of solvent effects on protein-ligand interactions using fully polarizable continuum model and rational drug design**.** Commun. Comput. Phys., 2013, 13, 31-60.

[63] 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.

[64] 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

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

[66] 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.

[67] Wang, R. X.; Lu, Y. P.; Wang, S. M. Comparative evaluation of 11 scoring functions for molecular docking. *J. Med. Chem*., 2003, 46, 2287-2303.

[68] Cummings, M. D.; DesJarlais, R. L.; Gibbs, A. C.; Mohan, V.; Jaeger, E. P. Comparison of automated docking programs as virtual screening tools. *J. Med. Chem.,* 2005, 48, 962-976.

[69] Kontoyianni, M.; McClellan, L. M.; Sokol, G. S. Evaluation of docking performance: comparative data on docking algorithms. *J. Med. Chem*., 2004, 47, 558-565.

[70] Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of library ranking efficacy in virtual screening. *J. Comput. Chem.* 2005, 26, 11-22.

[71] Perola, E.; Walters, W. P.; Charifson, P. S. A detailed comparison of current docking and scoring methods on systems of pharmaceutical relevance. *Proteins,* 2004, 56, 235-249.

[72] Wang, R. X.; Lu, Y. P.; Fang, X. L.; Wang, S. M. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.,* 2004, 44, 2114-2125.

[73] Bissantz, C.; Folkers, G.; Rognan, D. Protein-based virtual screening of chemical databases. 1. Evaluation of different docking/scoring combinations. *J. Med. Chem*., 2000, 43, 4759-4767.

[74] Paul, N.; Rognan, D. ConsDock, a new program for the consensus analysis of protein-ligand interactions. *Proteins*, 2002, 47, 521-533.

[75] Weis, A.; Katebzadeh, K.; Söderhjelm, P.; Nilsson, I.; Ryde, U. Ligand affinities predicted with the mm/pbsa method: dependence on the simulation method and the force field. *J. Med. Chem.,* 2006, 49, 6596-6606.

[76] Wang, J. M.; Hou, T. J.; Xu, X. J. Recent advances in free energy calculations with a combination of molecular mechanics and continuum models. *Curr. Comput.-Aided Drug Des*., 2006, 2, 287-306.

[77] 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.

[78] Muegge, I. Effect of ligand volume correction on PMF scoring. *J. Comput. Chem*., 2001, 22, 418-425.

[79] Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: A simplified potential approach. *J. Med. Chem*., 1999, 42, 791-804.

[80] Gohlke, H.; Hendlich, M.; Klebe, G. Knowledge-based scoring function to predict protein−ligand interactions. *J. Mol. Biol*., 2000, 295, 337-356.

[81] Lyne, P. D.; Lamb, M. L.; Saeh, J. C. Accurate prediction of the relative potencies of members of a series of kinase inhibitors using molecular docking and MM-GBSA scoring. *J. Med. Chem*., 2006, 49, 4805-4808.

[82] Guimaraes, C. R. W.; Cardozo, M. MM-GB/SA rescoring of docking poses in structure-based lead optimization. *J. Chem. Inf. Model*., 2008, 48, 958-970.

[83] Kuhn, B.; Gerber, P.; Schultz-Gasch, T.; Stahl, M. Validation and use of the MM-PBSA approach for drug discovery. *J. Med. Chem*., 2005, 48, 4040-4048.

[84] Hou, T. J.; Wang, J. M.; Li, Y. 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. Inf. Model*., 2011, 51, 69-82.

[85] Hou, T. J.; Wang, J. M.; Li, Y. Y.; Wang, W. Assessing the performance of the molecular mechanics Poisson Boltzmann surface area and molecular mechanics generalized Born surface area methods. II. The accuracy of ranking poses generated from docking. *J. Comput. Chem*., 2011, 32, 866-877.

[86] Zhang, D. W.; Zhang, J. Z. H. Molecular fractionation with conjugate caps for full quantum mechanical calculation of protein−molecule interaction energy. *J. Chem. Phys*., 2003, 119, 3599-3606.

[87] Gao, A. M.; Zhang, D. W.; Zhang, J. Z. H.; Zhang, Y. K. An efficient linear scaling method for ab initio calculation of electron density of proteins. *Chem. Phys. Lett*., 2004, 394, 293-297.

[88] Mei, Y.; Ji, C. G.; Zhang, J. Z. H. A new quantum method for electrostatic solvation energy of protein. *J. Chem. Phys*., 2006, 125, No. 094906.

[89] Ji, C. G.; Mei, Y.; Zhang, J. Z. H. Developing polarized protein specific charges for protein dynamics: MD free energy calculation of pK_a shifts for Asp26/Asp20 in thioredoxin. *Biophys. J*., 2008, 95, 1080-1088.

[90] Tong, Y.; Mei, Y.; Zhang, J. Z. H.; Duan, L. L.; Zhang, Q. G. Quantum calculation of protein solvation and protein-ligand binding free energy for HIV-1 protease/water complex. *J. Theor. Comput. Chem*., 2009, 8, 1265-1279.

[91] McConkey, B. J.; Sobolev, V.; Edelman, M. The performance of current methods in ligand-protein docking. *Curr. Sci*., 2002, 83, 845-856.

[92] de Graaf, C.; Pospisil, P.; Pos, W.; Folkers, G.; Vermeulen, N. P. E. Binding mode prediction of cytochrome P450 and thymidine kinase protein-ligand complexes by consideration of water and rescoring in automated docking. *J. Med. Chem*., 2005, 48, 2308-2318.

[93] Poornima, C. S.; Dean, P. M. Hydration in drug design. 1. Multiple hydrogen-bonding features of water molecules in mediating protein-ligand interactions. *J. Comput.-Aided Mol. Des*., 1995, 9, 500-512.

[94] Poornima, C. S.; Dean, P. M. Hydration in drug design. 2. Influence of local site surface shape on water binding. *J. Comput.-Aided Mol.Des*., 1995, 9, 513-520.

[95] Poornima, C. S.; Dean, P. M. Hydration in drug design. 3. Conserved water molecules at the ligand-binding sites of homologous proteins. *J. Comput.-Aided Mol. Des*., 1995, 9, 521-531.

[96] 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.

[97] Young, T.; Abel, R.; Kim, B.; Berne, B. J.; Friesner, R. A. Motifs for molecular recognition exploiting hydrophobic enclosure in protein−ligand binding. *Proc. Natl. Acad. Sci. U.S.A.*, 2007, 104, 808-813.

[98] Abel, R.; Young, T.; Farid, R.; Berne, B. J.; Friesner, R. A. Role of the active-site solvent in the thermodynamics of factor Xa ligand binding. *J. Am. Chem. Soc*., 2008, 130, 2817-2831.

[99] Rarey, M.; Kramer, B.; Lengauer, T.; Klebe, G. The particle concept: placing discrete water molecules during protein-ligand docking predictions. *Proteins*, 1999, 34, 17-28.

[100] Schnecke, V.; Kuhn, L. A. Virtual screening with alvation and ligand-induced complementarity. *Perspect. Drug Discov. Des*., 2000, 20,171-190.

[101] Osterberg, F.; Morris, G. M.; Sanner, M. F.; Olson, A. J.; Goodsell, D. S. Automated docking to multiple target structures: incorporation of protein mobility and structural water heterogeneity in AutoDock. *Proteins*, 2002, 46, 34-40.

[102] Verdonk, M. L.; Chessari, G.; Cole, J. C.; Hartshorn, M. J.; Murray, C. W. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem*., 2005, 48, 6504-6515.

[103] Böhm, H.-J.;Klebe, G. What can we learn from molecular recognition in protein-ligand complexes for the design of new drugs*? Angew. Chem. Int. Ed. Engl.,* 1996, 3S, 2588-2614.

[104] Uchida, T.; Ishimori, K.; Morishima, I. The effects of heme pocket hydrophobicity on the ligand binding dynamics in myoglobin as studied with leucine 29 mutants. *J. Biol. Chem*., 1997, 272, 30108-30114.

[105] Setny, P.; Baron,R.; Kekenes-Huskey, P. M.; McCammon, J. A.; Dzubiella, J. Solvent fluctuations in hydrophobic cavity-ligand binding kinetics. *Proc. Natl. Acad. Sci. U. S. A.*, 2013, 110, 1197-1202.

[106] Ahmad, M.; Gu, W.; Helms, V. Mechanism of fast peptide recognition by SH3 domains. *Angew. Chem. Int. Ed. Engl*., 2008, 47, 7626-7630.

[107] Setny, P.; Wang, Z.; Cheng, L.-T.; Li, B.; McCammon, J. A.; Dzubiella, J. Dewetting-controlled binding of ligands to hydrophobic pockets. *Phys. Rev. Lett.,* 2009, 103, 187801.

[108] Han, S.; Zaniewski, R. P.; Marr, E. S.; Lacey, B. M.; Tomaras, A. P.; Evdokimov, A.; Miller,J. R.; Shanmugasundaram, V. Structural basis for effectiveness of siderophore-conjugated monocarbams against clinically relevant strains of Pseudomonas aeruginosa. *Proc. Natl. Acad. Sci. U. S. A*,. 2010, 107, 22002-22007.

[109] Liu, L.; Michelsen, K.; Kitova, E. N.; Schnier, P. D.; Klassen, J. S. Evidence that water can reduce the kinetic stability of protein-hydrophobic ligand interactions. *J. Am. Chem. Soc*., 2010, 132, 17658-17660.

[110] Dror, R. O. C. Pan, A. C.; Arlow, D. H.; Borhani, D. W.; Maragakis, P.; Shan, Y.; Xu, H.; Shaw, D. E. Pathway and mechanism of drug binding to G-protein-coupled receptors. *Proc. Natl. Acad. Sci. U. S. A*., 2011, 108, 13118-13123.

[111] Schmidtke, P.; Luque, F. J.; Murray, J. B.; Barril, X. Shielded hydrogen bonds as structural determinants of binding kinetics: application in drug design. *J. Am. Chem. Soc.*, 2011, 133, 18903-18910.

[112] Chandler, D. Interfaces and the driving force of hydrophobic assembly. *Nature,* 2005, 437, 640-647.

[113] Dror, R.O.; Jensen, M. Ø.; Borhani, D. W.; Shaw, D. E. Exploring atomic resolution physiology on a femtosecond to millisecond timescale using molecular dynamics simulations. *J. Gen. Phys*., 2010, 135, 555-562.

[114] Buch, I.; Giorgino, T.; De Fabritiis, G. Complete reconstruction of an enzyme-inhibitor binding process by molecular dynamics simulations. *Proc. Natl. Acad. Sci. U. S. A.*, 2011, 108, 10184-10189.

[115] Shan, Y.; Kim, E. T.; Eastwood, M. P.; Dror, R. O.; Seeliger, M. A.; Shaw, D. E. How does a drug molecule find its target binding site? *J. Am. Chem. Soc*., 2011, 133, 9181-9183.

[116] Chang, C.E. A.; Trylska, J.; Tozzini, V.; McCammon, J. Binding pathways of ligands to HIV-1 protease: coarse grained and atomistic simulations. *Chem. Biol. Drug Des*., 2007, 69, 5-13.

[117] Li, X.; Latour, R. A.; Stuart, S. J. TIGER2: an improved algorithm for temperature intervals with global exchange of replicas. *J. Chem. Phys*., 2009. 130, 174106.

[118] Kang, M. Roberts, C.; Cheng, Y.; Chang, C.-E. A. Gating and intermolecular interactions in ligand-protein association: coarse-grained modeling of HIV-1 protease. J. Chem. Theory Comput., 2011, 7, 3438-3446.

[119] Kruse, A.C.; Hu, J.; Pan, A. C.; Arlow, D. H.; Rosenbaum, D. M.; Rosemond, E.; Green, H. F.; Liu, T.; Chae, P. S.; Dror, R. O.; Shaw, D.E.; Weis, W. I.; Wess, J.; Kobilka, B. K. Structure and dynamics of the M3 muscarinic acetylcholine receptor. *Nature*, 2012, 482, 552-556.

[120] Carlsson, P.; Burendahl, S.; Nilsson, L. Unbinding of retinoic acid from the retinoic acid receptor by random expulsion molecular dynamics. *Biophys. J*., 2006, 91, 3151-3161.

[121] Wang, T.; Duan, Y. Ligand entry and exit pathways in the b2-adrenergic receptor. *J. Mol. Biol*., 2009, 392, 1102-1115.

[122] Pietrucci, F.; Marinelli, F.; Carloni, P.; Laio, A. Substrate binding mechanism of HIV-1 protease from explicit-solvent atomistic simulations. *J. Am. Chem. Soc*., 2009, 131, 11811-11818.

[123] Colizzi, F.; Perozzo, R.; Scapozza, L.; Recanatini, M.; Cavalli, A. Single-molecule pulling simulation can discern active from inactive enzyme inhibitors. *J. Am. Chem. Soc*., 2010, 132, 7361-7371.

[124] Limongelli, V.; Bonomi, M.; Marinelli, L.; Gervasio, F. L.; Cavalli, A.; Novellino, E.; Parrinello, M. Molecular basis of cyclooxygenase enzymes (COXs) selective inhibition. *Proc. Natl. Acad. Sci. U. S. A.,* 2010, 107, 5411-5416.

[125] Bortolato, A.; Deflorian, F.; Weiss, D. R.; Mason, J. S. Decoding the role of water dynamics in ligand−protein unbinding: CRF_1R as a test case. *J. Chem. Inf. Model*., 2015, 55, 1857-1866

[126] Luckner, S.R.; Liu, N.; am Ende, C. W.; Tonge, P. J.; Kisker, C. A slow, tight binding inhibitor of InhA, the enoyl-acyl carrier protein reductase from Mycobacterium tuberculosis. *J. Biol. Chem*., 2010, 285, 14330-14337

[127] Ferrara, P.; Gohlke, H.; Price, D. J.; Klebe, G.; Brooks, C. L. III. Assessing scoring functions for protein-ligand interactions. *J. Med. Chem*., 2004, 47, 3032-3047.

[128] Bliznyuk, A A.; Rendell, A. P. Electronic effects in biomolecular simulations: Investigation of the KcsA potassium ion channel. *J. Phys. Chem. B*, 2004, 108, 13866-13873.

[129] Bliznyuk, A. A.; Rendell, A. P.; Allen, T. W.; Chung, S. H. The potassium ion channel: comparison of linear scaling semiempirical and molecular mechanics representations of the electrostatic potential. *J. Phys. Chem. B*, 2001, 105, 12674-12679.

[130] Zuegg, J.; Bliznyuk, A. A.; Gready, J. E. Comparison of electrostatic potential around proteins calculated from Amber and AM1 charges: application to mutants of prion protein. *Mol. Phys*., 2003, 101, 2437-2450.

[131] Raha, K.; Merz, K. M. Large-scale validation of a quantum mechanics based scoring function: predicting the binding affinity and the binding mode of a diverse set of protein-ligand complexes. *J. Med. Chem*., 2005, 48, 4558-4575.

[132] Vasilyev, V.; Bliznyuk, A. Application of semiempirical quantum chemical methods as a scoring function in docking. *Theor. Chem. Acc*., 2004, 112, 313-317.

[133] Villar, R.; Gil, M. J.; Garcia, J. I.; Martinez-Merino, V. Are AM1 ligand-protein binding enthalpies good enough for use in the rational design of new drugs? J. Comput. Chem., 2005, 26, 1347-1358.

[134] Charifson, P. S.; Corkery, J. J.; Murcko, M. A.; Walters, W. P. Consensus scoring: a method for obtaining improved hit rates from docking databases of three-dimensional structures into proteins. *J. Med. Chem*., 1999, 42, 5100-5109.

[135] Clark, R. D.; Strizhev, A.; Leonard, J. M.; Blake, J. F.; Matthew, J. B. Consensus scoring for ligand/protein interactions. *J. Mol. Graph. Model*., 2002, 20, 281-295.

[136] Perez, C.; Ortiz, A. R. Evaluation of docking functions for protein-ligand docking. *J. Med. Chem*., 2001, 44, 3768-3785.

[137] Czerminski, R.; Elber, R. Computational studies of ligand diffusion in globins: I. Leghemoglobin. *Proteins*, 1991, 10, 70-80.

[138] Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chem. Phys. Lett*., 1999, 314, 141-151.

[139] Berndt, K. D.; Guntert, P.; Wuthrich, K. Conformational sampling by NMR solution structures calculated with the program DIANA evaluated by comparison with long-time molecular dynamics calculations in explicit water. *Proteins*, 1996, 24, 304-313.

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

[141] Apostolakis, J.; Pluckthun, A.; Caflisch, A. Docking small ligands in flexible binding sites. *J. Comput. Chem*., 1998, 19, 21-37.

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

[143] Trosset, J. Y.; Scheraga, H. A. PRODOCK: Software package for protein modeling and docking. *J. Comput. Chem*., 1999, 20, 412-427.

[144] Carlson, H. A.; Masukawa, K. M.; Rubins, K.; Bushman, F. D.; Jorgensen, W. L.; Lins, R. D.; Briggs, J. M.; McCammon, J. A. Developing a dynamic pharmacophore model for HIV-1 integrase. *J. Med. Chem*., 2000, 43, 2100-2114.

[145] Alonso, H.; Bliznyuk, A. A.; Gready, J. E. Combining docking and molecular dynamic simulations in drug design. *Med. Res. Rev*., 2006, 26, 531-568.

[146] Leach, A. R. Ligand docking to proteins with discrete side-chain flexibility. *J. Mol. Biol*., 1994, 235, 345-356.

[147] Aqvist, J.; Medina, C.; Sammuelsson, J. E. A new method for predicting binding affinity in computer-aided drug design. *Protein Eng.,* 1994, 7, 385-391.

[148] Huo, S.; Wang, J.; Cieplak, P.; Kollman, P. A.; Kuntz, I. D. Molecular dynamics and free energy analyses of cathepsin D-inhibitor interactions: insight into structure-based ligand design. *J. Med. Chem*., 2002, 45, 1412-1419.

[149] Kuhn, B.; Kollman, P. A. Binding of a diverse set of ligands to avidin and streptavidin: An accurate quantitative prediction of their relative affinities by a combination of molecular mechanics and continuum solvent models. *J. Med. Chem*., 2000, 43, 3786-3791.

[150] Smith, R. H. Jr; Jorgensen, W. L.; Tirado-Rives, J.; Lamb, M. L.; Janssen, P. A.; Michejda, C. J.; Kroeger Smith, M. B. Prediction of binding affinities for TIBO inhibitors of HIV-1 reverse transcriptase using Monte Carlo simulations in a linear response method. *J. Med. Chem*., 1998, 41, 5272-5286.

[151] Wang, J.; 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.

[152] Straatsma, T. P.; McCammon J. A. Computational alchemy. *Ann. Rev.Phys. Chem.*, 1992, 43, 407-435.

[153] Brandsdal, B. O.; Ơsterberg, F.; Almlöf, M.; Feierberg, I.; Luzhkov, V. B.; Åqvist, J. Free energy calculations and ligand binding. *Adv. Protein Chem*., 2003, 66, 123-158.

[154] Beveridge, D. L.; DiCapua, F. M. Free energy via molecular simulation: Applications to chemical and biomolecular systems. *Annu. Rev. Biophys. Biophys. Chem*., 1989, 18, 431-492.

[155] Gready, J. E.; Cummins, P. L. Mechanism-based substrates and inhibitors of dihydrofolate reductase. In: Reddy, M. R.; Erion, M. D., editors. *Free energy calculations in rational drug design*. New York: Kluwer Academic/Plenum, 2001, p 343-364.

[156] Bash, P. A.; Singh, U. C.; Brown, F. K.; Langridge, R.; Kollman, P. A. Calculation of the relative change in binding free energy of a protein-inhibitor complex. *Science*, 1987, 235, 574-576.

[157] Aqvist ,J. Calculation of absolute binding free energies for charged ligands and effects of long-range electrostatic interactions. *J. Comput. Chem*., 1996, 17, 1587-1597.

[158] Hansson, T.; Aqvist, J. Estimation of binding free energies for HIV proteinase inhibitors by molecular dynamics simulations. *Protein. Eng*., 1995, 8, 1137-1144.

[159] Hansson, T.; Marelius, J.; Aqvist, J. Ligand binding affinity prediction by linear interaction energy methods. *J. Comput. Aided Mol. Des*., 1998, 12, 27-35.

[160] Wang, W.; Wang, J.; Kollman, P. A. What determines the van der Waals coefficient beta in the LIE (linear interaction energy) method to estimate binding free energies using molecular dynamics simulations? *Proteins*, 1999, 34, 395-402.

[161] Swanson, J. M.; Henchman, R. H.; McCammon, J. A. Revisiting free energy calculations: a theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophys. J*., 2004, 86, 67-74.

[162] Aqvist, J.; Luzhkov, V. B.; Brandsal, B. O. Ligand binding affinities from MD simulations. *Acc. Chem. Res*., 2002, 35, 358-365.

[163] Paulsen, M. D.; Ornstein, R. L. Binding free energy calculations for P450cam-substrate complexes. *Protein. Eng*., 1996, 9, 567-571.

[164] Wang, J.; Dixon, R.; Kollman, P. A. Ranking ligand binding affinities with avidin: a molecular dynamics-based interaction energy study. *Proteins: Struct. Funct. Genet*., 1999, 34, 69-81.

[165] Jones-Hertzog, D. K.; Jorgensen, W. L. Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *J. Med. Chem*., 1997, 40, 1539-1549.

[166] Carlson, H. A.; Jorgensen, W. L. An extended linear-response method for determining free-energies of hydration. *J. Phys. Chem*., 1995, 99, 10667-10673.

[167] Aqvist, J.; Marelius, J. The linear interaction energy method for predicting ligand binding free energies. *Comb. Chem. High Throughput Screen*, 2001, 4, 613-626.

[168] Almlof, M.; Brandsdal, B. O.; Aqvist, J. Binding affinity prediction with different force fields: examination of the linear interaction energy method. *J. Comput. Chem*., 2004, 25, 1242-1254.

[169] Norberg, J.; Nilsson, L. Advances in bimolecular simulations: methodology and recent applications. *Q. Rev. Biophys*., 2003, 36, 257-306.

[170] Tai, K. Conformational sampling for the impatient. *Biophys. Chem*., 2004, 107, 213-220.

[171] Miranker, A.; Karplus, M. Functionality maps of binding sites: A multiple copy simultaneous search method. *Proteins*, 1991, 11, 29-34.

[172] Huber, T.; Torda, A. E.; van Gunsteren, W. F. Local elevation: A method for improving the searching properties of molecular dynamics simulation. *J. Comput. Aided Mol. Des*., 1994, 8, 695-708.

[173] Schulze, B. G.; Grubmuller, H.; Evanseck, J. D. Functional significance of hierarchical tiers in carbon monoxy myoglobin: conformational substates and transitions studied by conformational flooding simulations. *J. Am. Chem. Soc*., 2000, 122, 8700-8711.

[174] Guvench, O.; MacKerell, A. D. Jr. Computational fragment-based binding site identification by ligand competitive saturation. *PLoS Comput. Biol*., 2009, 5, e1000435.

[175] Seco, J.; Luque, F.J.; Barril, X. Binding site detection and druggability index from first principles. J. Med. Chem., 2009, 52, 2363-2371.

[176] Raman, E.P.; Yu, W.; Guvench, O.; Mackerell, A.D. Reproducing crystal binding modes of ligand functional groups using site-identification by ligand competitive saturation (SILCS) simulations. J. Chem. Inf. Model., 2011, 51, 877-896.

[177] Gervasio, F. L.; Laio, A.; Parrinello, M. Flexible docking in solution using metadynamics. *J. Am. Chem. Soc*., 2005, 127, 2600-2607.

[178] Laio, A.; Parrinello, M. Escaping free-energy minima. *Proc. Natl. Acad. Sci. USA,* 2002, 99, 12562-12566.

[179] Bieri, M.; Kwan, A. H.; Mobli, M.; King, G. F.; Mackay, J. P.; Gooley, P. R. Macromolecular NMR spectroscopy for the non-spectroscopist: beyond macromolecular solution structure determination. *FEBS J*., 2011, 278, 704-715

[180] Kwan, A. H.; Mobli, M.; Gooley, P. R.; King, G. F.; Mackay, J. P. Macromolecular NMR spectroscopy for the non-spectroscopist. *FEBS J.* ,2011, 278, 687-703

[181] Kasai, K.; Ishii, S. J. Quantitative analysis of affinity chromatography of trypsin. A new technique for investigation of protein-ligand interaction. *Biochem*, 1975, 77, 261-264.

[182] Eldridge, M. D.; Murray, C. W.; Auton, T. R.; Paolini, G. V.; Mee, R. P. Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput.Aided Mol. Des*., 1997, 11, 425-445.

[183] Richarme, G.; Kepes, A. Study of binding protein-ligand interaction by ammonium sulfate-assisted adsorption on cellulose

esters filters. *Biochim. Biophys. Acta, Protein Struct. Mol. Enzymol*. 1983, 742, 16-24.

[184] Muegge, I.; Martin, Y. C. A general and fast scoring function for protein-ligand interactions: a simplified potential approach. *J. Med. Chem*., 1999, 42, 791-804.

[185] Olsson, T. S. G.; Williams, M. A.; Pitt, W. R.; Ladbury, J. E. The thermodynamics of protein-ligand interaction and solvation: insights for ligand design. *J. Mol. Biol*., 2008, 384, 1002-1017.

[186] Bash, P. A.; Field, M. J.; Karplus, M. Free energy perturbation method for chemical reactions in the condensed phase: a dynamic approach based on a combined quantum and molecular mechanics potential. *J. Am. Chem. Soc*., 1987, 109, 8092-8094.

[187] Rao, S. N.; Singh, U. C.; Bash, P. A.; Kollman, P. A. Free energy perturbation calculations on binding and catalysis after mutating Asn 155 in subtilisin. *Nature*, 1987, 328, 551-554.

[188] Kita, Y.; Arakawa, T.; Lin, T. Y.; Timasheff, S. N. Contribution of the surface free energy perturbation to protein-solvent interactions. *Biochemistry*, 1994, 33, 15178-15189.

[189] Rao, B. G.; Singh, U. C. A free energy perturbation study of solvation in methanol and dimethyl sulfoxide. *J. Am. Chem. Soc*., 1990, 112, 3803-3811.

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

[191] Jorgensen, W. L.; Thomas, L. L. Perspective on Free-Energy Perturbation Calculations for Chemical Equilibria. *J. Chem. Theory Comput*., 2008, 4, 869-876.

[192] Zacharias, M.; Straatsma, T. P.; McCammon, J. A. Separation-shifted scaling, a new scaling method for Lennard-Jones interactions in thermodynamic integration. *J. Chem. Phys.*, 1994, 100, 9025-9031.

[193] Ryde, U.; Sӧderhjelm, P. Ligand-binding affinity estimates supported by quantum-mechanical methods. *Chem. Rev*., 2016, 116, 5520-5566.

[194] Massova, I.; Kollman, P. A. Combined molecular mechanical and continuum solvent approach (MM-PBSA/GBSA) to predict ligand binding. *Perspect. Drug Discovery Des*., 2000, 18, 113-135.

[195] Luo, R.; David, L.; Gilson, M. K. Accelerated Poisson-Boltzmann calculations for static and dynamic systems. *J. Comput. Chem*., 2002, 23, 1244-1253.

[196] Nicholls, A.; Honig, B. A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem*., 1991, 12, 435-445.

[197] Chen, Z.; Baker, N. A.; Wei, G. W. J. Comput. Phys., 2010, 229, 8231-8258.

[198] Warren, G. L.; Andrews, C. W.; Capelli, A.-M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.;, Head, M. S. A critical assessment of docking programs and scoring functions. *J. Med. Chem*., 2006, 49, 5912-5931.

[199] Cross, J. B.; Thompson, D. C.; Rai, B. K.; Baber, J. C.; Fan, K. Y.; Hu, Y.; Humblet, C. Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J. Chem. Inf. Model*., 2009, 49, 1455-1474.

[200] Lengauer, T.; Rarey, M. Computational methods for biomolecular docking. *Curr. Opin. Struct. Biol*., 1996, 6, 402-406.

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

[202] Rosenfeld, R. J.; Goodsell, D. S.; Musah, R. A.; Morris, G. M.; Goodin, D. B.; Olson A. J. Automated docking of ligands to an artificial active site: augmenting crystallographic analysis with computer modeling. *J. Comput. Aided Mol. Des*., 2003, 17, 525-536.

[203] Goodsell, D. S.; Morris, G. M.; Olson, A. J. Automated docking of flexible ligands: applications of AutoDock. *J. Mol. Recognit*., 1996, 9, 1-5.

[204] Ewing, T. J. A.; Makino, S.; Skillman, A. G.; Kuntz, I. D. DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J. Comput. Aided Mol. Des*., 2001, 15, 411-428.

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

[206] Yang, J. M. Development and evaluation of a generic evolutionary method for protein-ligand docking. J. *Comput. Chem*., 2004, 25, 843-857.

[207] Yang, J. M.; Chen, C. C. GEMDOCK: a generic evolutionary method for molecular docking. *Proteins*, 2004, 55, 288-304.

[208] 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.

[209] Kirkpatrick, P. Virtual screening: gliding to success. *Nat. Rev. Drug. Discov.,* 2004, 3, 299-299.

[210] Krovat, E. M.; Steindl, T.; Langer, T. Recent advances in docking and scoring. *Curr. Comput. Aided Drug Des*., 2005, 1, 93-102.

[211] Verdonk, M. L.;, Cole, J. C.; Hartshorn, M. J.; Murray, C. W.; Taylor, R. D. Improved protein-ligand docking using GOLD. *Proteins*, 2003, 52, 609-623.

[212] Joy, S.; Nair, P. S.; Hariharan, R.; Pillai, M. R. Detailed comparison of theprotein-ligand docking effciencies of GOLD, a commercial package and ArgusLab, a licensable freeware. *In Silico Biol*., 2006, 6, 601-605.

[213] Thomsen, R.; Christensen, M. H. MolDock: A new technique for high-accuracy molecular docking. *J. Med. Chem*., 2006, 49, 3315-3321.

[214] Jain, A. N. Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J. Med. Chem*., 2003, 46, 499-511.

[215] Mitrasinovic, P. M. On the structure-based design of novel inhibitors of H5N1 influenza A virus neuraminidase (NA). *Biophys. Chem*., 2009, 140, 35-38.

[216] Mihajlovic, M. L.; Mitrasinovic, P. M. Applications of the ArgusLab 4/A Score protocol in the structure-based binding affinity prediction of various inhibitors of group-1 and group-2 influenza virus neuraminidases (NAs). *Mol. Simul*., 2009, 35, 311-324.

[217] Ravindranathan, K. P.; Mandiyan, V.; Ekkati, A. R.; Bae, J. H.; Schlessinger, J,; Jorgensen, W. L. Discovery of novel fibroblast growth factor receptor 1 kinase inhibitors by structure-based virtual screening. *J. Med. Chem*., 2010, 53, 1662-1672.

[218] Johnson, B. C.; Métifiot, M.; Pommier, Y.; Hughesa, S. H. Molecular dynamics approaches estimate the binding energy of HIV-1 integrase inhibitors and correlate with *in vitro* activity. *Antimic. Ag. Chemother*., 2012, 56, 411-419.

[219] Langley, J. N. On the reaction of cells and of nerve-endings to certain poisons, cheifly as regards the reaction of striated muscle to nicotine and to curari. *J. Physiol*., 1905, 33, 374-413.

[220] Copeland, R. A.; Pompliano, D. L.; Meek, T. D. Drug-target residence time and its implications for lead optimization. *Nat. Rev. Drug Discov*., 2006, 5, 730-739.

[221] Swinney, D. C. The role of binding kinetics in therapeutically useful drug action. *Curr. Opin. Drug Discov. Dev*., 2009, 12, 31-39.

[222] Lu, H.; Tonge, P. J. Drug-target residence time: critical information for lead optimization. *Curr. Opin. Chem. Biol.*, 2010, 14, 467-474.

[223] Waszkowycz, B.; Clark, D. E.; Gancia, E. Outstanding challenges in protein-ligand docking and structure-based virtual screening. *WIREs Comput. Mol. Sci*., 2011, 1, 229-259.

[224] Coupez, B.; Lewis, R. Docking and scoring-theoretically easy, practically impossible? *Curr. Med. Chem*., 2006, 13, 2995-3003.

[225] Alberts, B. The cell as a collection overview of protein machines: preparing the next generation of molecular biologists. *Cell,* 1998, 92, 291-294.

[226] Seifriz, W. A materialistic interpretation of life. *Phil. Sci*., 1939, 6, 266-284.

[227] Cavasotto, C. N.; Abagyan, R. A. Protein flexibility in ligand docking and virtual screening to protein kinases. *J. Mol. Biol*., 2004, 337, 209-226.

[228] Chandrika, B.; Subramanian, J.; Sharma, S. D. Managing protein flexibility in docking and its applications. *Drug Discov. Today*, 2009, 14, 394-400.

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

[230] Tuffery, P.; Derreumaux, P. Flexibility and binding affinity in protein-ligand, protein-protein and multi-component protein interactions: limitations of current computational approaches. *J. R. Soc. Interface*, 2012, 9, 20-33.

[231] Guo, D.; Thea Mulder-Krieger, T.; Jzerman, A. P. I.; Heitman, L. H. Functional efficacy of adenosine A2A receptor agonists is positively correlated to their receptor residence time. *Br. J. Pharmacol*., 2012, 166, 1846-1859.

[232] Lipton, S. A. Turning down, but not off: neuroprotection requires a paradigm shift in drug development. *Nature*, 2004, 428, 473-474.

[233] Ohlson, S. Designing transient binding drugs: a new concept for drug discovery. *Drug Discov. Today*, 2008, 13, 433-439.

[234] Vauquelin, G.; Bostoen, S.; Vanderheyden, P.; Seeman, P. Clozapine, atypical antipsychotics, and the benefits of fast-off D2 dopamine receptor antagonism. *Naunyn-Schmiedeberg's Arch. Pharmacol.,* 2012, 385, 337-372.

[235] Keighley, W. The need for high throughput kinetics early in the drug discovery process. *Drug Disc. World Summer*, 2011, 39-45.

[236] Frauenfelder, H.; Sligar, S. G.; Wolynes, P. G. The energy landscapes and motions of proteins. *Science*, 1991, 254, 1598-1603.

[237] Uchida, T.; Unno, M.; Ishimori, K.; Morishima, I. Effects of the intramolecular disulfide bond on ligand binding dynamics in myoglobin. *Biochemistry*, 1997, 36, 324-332.

[238] Pargellis, C.; Tong, L.; Churchill, L.; Cirillo, P.F.; Gilmore, T.; Graham; A. G.; Grob, P. M.; Hickey, E. R.; Moss, N.; Pav, S.; Regan, J. Inhibition of p38 MAP kinase by utilizing a novel allosteric binding site. *Nat. Struct. Biol*., 2002, 9, 268-272.

[239] Millan, D.S. Mark E. Bunnage, M. E.; Burrows, J. L.; Butcher, K. J.; Dodd, P. G.; Evans, T. J.; Fairman, D. A.; Hughes, S. J.; Kilty, I. C.; Lemaitre, A.; Lewthwaite, R. A.; Mahnke, A.; Mathias, J. P.; Philip, J.; Smith, R. T.; Stefaniak, M. H.; Yeadon, M.; Phillips, C. Design and synthesis of inhaled p38 inhibitors for the treatment of chronic obstructive pulmonary disease. *J. Med. Chem*., 2011, 54, 7797-7814.

[240] Robinson, R.A. and Stokes, R.H. *Electrolyte Solutions* (2nd revised edn.), 2002, Dover Publications Inc., Mineola NY, USA.

[241] Sun, S.F., ed. Physical Chemistry of Macromolecules: Basic Principles and Issues. Second Edition. John Wiley and Sons, Inc., 2004, Hoboken, New Jersey.

[242] Schreiber, G.; Fersht, A. R. Rapid, electrostatically assisted association of proteins. *Nat. Struct. Biol.,* 1996, 3, 427-431.

[243] Fedosova, N. U.; Champeil, P.; Esmann, M. Nucleotide binding to Na, K-ATPase: the role of electrostatic interactions. *Biochemistry*, 2002, 41, 1267-1273.

[244] Radić, Z.; Kirchhoff, P. D.; Quinn, D. M.; McCammon, J. A.; Taylor, P. Electrostatic influence on the kinetics of ligand binding to acetylcholinesterase. Distinctions between active center ligands and fasciculin. *J. Biol. Chem*., 1997, 272, 23265-23277.

[245] Collins, K. D. Why continuum electrostatics theories cannot explain biological structure, polyelectrolytes or ionic strength effects in ion-protein interactions. *Biophys. Chem*., 2012, 167, 43-59.

[246] Ledvina, P. S.; Tsai, A. L.; Wang, Z.; Koehl, E.; Quiocho, F. A. Dominant role of local dipolar interactions in phosphate binding to a receptor cleft with an electronegative charge surface: equilibrium, kinetic, and crystallographic studies. *Protein Science*, 1998, *7*, 2550-2559.

[247] Krishnamurthy, V. M.; Kaufman, G. K.; Urbach, A. R.; Gitlin, I.; Gudiksen, K. L.; Weibel, D. B.; Whitesides, G. M. Carbonic anhydrase as a model for biophysical and physical-organic studies of proteins and protein-ligand binding. *Chem. Rev*., 2008, 103, 946-1051.

[248] Ryde, U. On the role of covalent strain in protein function. *Recent Research Developments in Protein Engineering*, Research Signpost: Trivandrum, India, 2002; 2, 65-91.

[249] Boström, J.; Norrby, P.-O.; Liljefors, T. Conformational energy penalties of protein-bound ligands. *J. Comput.-Aided Mol. Des*. ,1998, 12, 383-396.

[250] Sitzmann, M.; Weidlich, I. E.; Filippov, I. V.; Liao, C.; Peach, M. L.; Ihlenfeldt, W.-D.; Karki, R. G.; Borodina, Y. V.; Cachau, R. E.; Nicklaus, M. C. PDB ligand conformational energies calculated quantum-mechanically. *J. Chem. Inf. Model.*, 2012, 52, 739-756.

[251] Ryde, U.; Olsen, L.; Nilsson, K. Quantum chemical geometry optimisations in proteins using crystallographic raw data. J. Comput.Chem., 2002, 23, 1058-1070.

[252] Fu, Z.; Li, X.; Merz, K. M. Accurate assessment of the strain energy in a protein-bound drug using QM/MM X-ray refinement and converged quantum chemistry. *J. Comput. Chem.*, 2011, 32, 2587-2597.

[253] Genheden, S.; Ryde, U. How to obtain statistically converged MM/GBSA results. *J. Comput. Chem.*, 2010, 31, 837-846.

[254] Kolář, M.; Fanfrlík, J.; Hobza, P. Ligand conformational and solvation/desolvation free energy in protein-ligand complex formation. *J. Phys. Chem. B*, 2011, 115, 4718-4724.

[255] Moghaddam, S.; Yang, C.; Rekharsky, M.; Ko, Y. H.; Kim, K.; Inoue, Y.; Gilson, M. K. New ultrahigh affinity host-guest complexes of cucurbit[7]uril with bicyclo[2.2.2]octane and adamantane guests: thermodynamic analysis and evaluation of M2 affinity calculations. *J. Am. Chem. Soc*., 2011, 133, 3570-3581.

[256] Chen, W.; Chang, C.-E.; Gilson, M. K. Calculation of cyclodextrin binding affinities: energy, entropy, and implications for drug design. *Biophys. J*., 2004, 87, 3035-3049.

[257] Zhou, H.-X.; Gilson, M. K. Theory of free energy and entropy in noncovalent binding. *Chem. Rev*., 2009, 109, 4092-4107.

[258] Liu, L.; Guo, Q.-X. Isokinetic relationship, isoequilibrium relationship, and enthalpy−entropy compensation. Chem. Rev., 2001, 101, 673-696.

[259] Sharp, K. Entropy−enthalpy compensation: fact or artifact? *Protein Sci*., 2001, 10, 661-667.

[260] Chodera, J. D.; Mobley, D. L. Entropy-enthalpy compensation: role and ramifications in biomolecular ligand recognition and design. *Annu. Rev. Biophys*., 2013, 42, 121-142.

[261] Ryde, U. A Fundamental view of enthalpy−entropy compensation. *Med. Chem. Comm*., 2014, 5, 1324-1336.

[262] Moitessier, N.; Englebienne, P.; Corbeil, C. R. Towards the development of universal, fast and highly accurate docking/scoring methods: a long way to go. *Br. J. Pharmacol*., 2008, 153, S7-S26.

[263] Leach, A. R.; Shoichet, B. K.; Peishoff, C. E. Prediction of protein-ligand interactions. Docking and scoring: successes and gaps. *J. Med. Chem*., 2006, 49, 5851-5855.

[264] Cheng, T.; Li, X.; Wang, R. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model*., 2009, 49, 1079-1093.

[265] Stjernschantz, E.; Marelius, J.; Oostenbrink, C. Are automated molecular dynamics simulations and binding free energy calculations realistic tools in lead optimization? An evaluation of the linear interaction energy (LIE) method. *J. Chem. Inf. Model*., 2006, 46, 1972-1983.

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

[267] Sherman, W.; Day, T.; Farid, R. Novel procedure for modeling ligand/receptor induced fit effects. *J. Med. Chem*., 2006, 49, 534-553.

[268] Bottegoni, G.; Kufareva, I.; Abagyan, R. A new method for ligand docking to flexible receptors by dual alanine scanning and refinement (SCARE). *J. Comput. Aided Mol. Des*., 2008, 22, 311-325.

[269] Totrov, M.; Abagyan, R. Flexible ligand docking to multiple receptor conformations: a practical alternative. *Curr. Opin. Struct. Biol*., 2008, 18, 178-184.

[270] Verdonk, M. L.; Chessari, G.; Taylor, R. Modeling water molecules in protein-ligand docking using GOLD. *J. Med. Chem*., 2005, 48, 6504-6515.

[271] Roberts, B. C.; Mancera, R. L. Ligand-protein docking with water molecules. *J. Chem. Inf. Model*., 2008, 48, 397-408.

[272] Englebienne, P.; Moitessier, N. Docking ligands into flexible and solvated macromolecules. 4. Are popular scoring functions accurate for this class of proteins? *J. Chem. Inf. Model*., 2009, 49, 1568-1580.

[273] Vasanthanathan, P.; Hritz, J.; Oostenbrink, C. Virtual screening and prediction of site of metabolism for cytochrome P450 1A2 ligands. *J. Chem. Inf. Model*., 2009, 49, 43-52.

[274] Santos, R.; Hritz, J.; Oostenbrink, C. Role of water in molecular docking simulations of cytochrome P450 2D6. *J. Chem. Inf. Model*., 2010, 50, 146-154.

[275] Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomole cules. *Nat. Struct. Biol*., 2002, 9, 646-652.

[276] Kitchen, D. B.; Decornez, H.; Furr, J. R.; Bajorath, J. Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat. Rev. Drug Discov*., 2004, 3, 935-949.

[277] Lin, J. H.; Perryman, A. L.; Schames, J. R.; McCammon, J. A. Computational drug design accommodating receptor flexibility: the relaxed complex scheme. *J. Am. Chem. Soc.*, 2002, 124, 5632-5633.

[278] Schames, J. R.; Henchman, R. H.; Siegel, J. S.; Sotriffer, C. A.; Ni, H.; McCammon, J. A. Discovery of a novel binding trench in HIV integrase. *J. Med. Chem*., 2004, 47, 1879-1881.

[279] Boehm, H. J.; Boehringer, M.; Bur, D.; Gmuender, H.; Huber, W.; Klaus, W.; Kostrewa, D.; Kuehne, H.; Luebbers, 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.

[280] Böhm, H. J.; Banner, D. W.; Weber, L. Combinatorial docking and combinatorial chemistry: design of potent non-peptide thrombin inhibitors. *J. Comput. Aided Mol. Des.,* 1999, 13, 51-56.

[281] Bressi, J. C.; Verlinde, C. L.; Aronov, A. M.; Shaw, M. L.; Shin, S. S.; Nguyen, L. N.; Suresh, S.; Buckner, F. S.; van Voorhis, W. C.; Kuntz, I. D.; Hol, W. G.; Gelb, M. H. Adenosine analogues as selective inhibitors of glyceraldehyde-3-phosphate dehydrogenase of Trypanosomatidae via structure-based drug design. *J. Med. Chem.*, 2001, 44, 2080-2093.

[282] Burkhard, P.; Hommel, U.; Sanner, M.; Walkinshaw, M. D. The discovery of steroids and other novel FKBP inhibitors using a molecular docking program. *J. Mol. Biol*., 1999, 287, 853-858.

[283] Filikov, A. V.; Mohan, V.; Vickers, T. A.; Griffey, R. H.; Cook, P. D.; Abagyan, R. A.; James, T. L. Identification of ligands for RNA targets via structure-based virtual screening: HIV-1 TAR. *J. Comput. Aided Mol. Des*., 2000, 14, 593-610.

[284] Iwata, Y.; Arisawa, M.; Hamada, R.; Kita, Y.; Mizutani, M. Y.; Tomioka, N.; Itai, A.; Miyamoto, S. Discovery of novel aldose reductase inhibitors using a protein structure-based approach: 3D-database search followed by design and synthesis. *J. Med. Chem*., 2001, 44, 1718-1728.

[285] Schneider, G.; Clement-Chomienne, O.; Hilfiger, L.; Schneider, P.; Kirsch, S.;, Bohm, H, J.; Neidhart, W. Virtual screening for bioactive molecules by evolutionary de novo design. *Angew. Chem. Int. Ed. Engl*., 2000, 39, 4130-4133.

[286] Schneider, G.;, Lee, M. L.; Stahl, M.; Schneider, P. De novo design of molecular architectures by evolutionary assembly of drug-derived building blocks. *J. Comput. Aided Mol. Des.*, 2000, 14, 487-494.

[287] Kubinyi, H. Drug research: myths, hype and reality. *Nat. Rev. Drug Discov*., 2003, 2, 665-668.

[288] Martin, Y. C. A bioavailability score. *J. Med. Chem*., 2005, 48, 3164-3170.

[289] Blundell, T. L.; Jhoti, H.; Abell, C. High-throughput crystallography for lead discovery in drug design. *Nat. Rev. Drug Discov*., 2002, 1, 45-54.

[290] Congreve, M.; Murray, C. W.; Blundell, T. L. Structural biology and drug discovery. *Drug Discov. Today*, 2005, 10, 895-907.

[291] Thorsteinsdottir, H. B.; Schwede, T.; Zoete, V.; Meuwly, M. How innacuracies in protein structure models affect estimates of protein-ligands interactions: computational analysis of HIV-I protease inhibitor binding. *Proteins Struct. Funct. Bioinf*., 2006, 65, 407-423.

[292] Paulsen, M. D.; Ornstein, R. L., Binding free energy calculations for P450cam-substrate complexes. *Protein Eng.* 1996, 9, 567-571.

[293] Lewis, D. F. V.; Eddershaw, P. J.; Dickins, M.; Tarbit, M. H.; Goldfarb, P. S. Structural determinants of cytochrome P450 substrate specificity, binding affinity and catalytic rate. *Chem.-Biol.Interact.,* 1998, 115, 175-199.

[294] Lewis, D. F. V. Homology modelling of human cytochromes P450 involved in xenobiotic metabolism and rationalization of substrate selectivity. *Exp. Toxic. Pathol.,* 1999, 51, 369-374.

[295] Helms, V.; Wade, R. C. Computational alchemy to calculate absolute protein-ligand binding free energy. *J. Am. Chem. Soc.,* 1998, 120, 2710-2713.

[296] Plewczynski, D.; Łaźniewski, M.; Augustyniak, R.; Ginalski, K. Can we trust docking results? Evaluation of seven commonly used programs on PDBbind database. *J. Comput. Chem.*, 2011, 32, 742-755.

[297] Ramensky, V.; Sobol, A.; Zaitseva, N.; Rubinov, A.; Zosimov, V. A novel approach to local similarity of protein binding sites substantially improves computational drug design results. *Proteins: Struct., Funct., Genet*., 2007, 69, 349-357.

[298] Muryshev, A. E.; Tarasov, D.; Butygin, A.; Butygina, O. Y.; Aleksandrov, A.; Nikitin, S. A novel scoring function for molecular docking. *J. Comput. Aided Mol. Des*., 2003, 17, 597-605.

[299] Loving, K.; Alberts, I.; Sherman, W. Computational approaches for fragment-based and de novo design. *Curr. Top. Med. Chem*., 2010, 10, 14-32.

[300] Hajduk, P. J.; Greer, J. A decade of fragment-based drug design: strategic advances and lessons learned. *Nat. Rev. Drug Discovery*, 2007, 6, 211-219.

[301] Erlanson, D. A.; McDowell, R. S.; O'Brien, T. Fragment-Based Drug Discovery. *J. Med. Chem*., 2004, 47, 3463-3482.

[302] Carr, R. A. E.; Congreve, M.; Murray, C. W.; Rees, D. C. Fragment-based lead discovery: leads by design. *Drug Discovery Today*, 2005, 10, 987-992.

[303] Murray, C. W.; Blundell, T. L. Structural biology in fragment based drug design. *Curr. Opin. Struct. Biol.*, 2010, 20, 497-507.

[304] Scott, D. E.; Coyne, A. G.; Hudson, S. A.; Abell, C. Fragment based approaches in drug discovery and chemical biology. *Biochemistry*, 2012, 51, 4990-5003.

[305] Erlanson, D. Practical Fragments. http://practicalfragments.blogspot.de/2015/01/fragments-in-clinic-2015-edition.html (accessed 13.03.2016).

[306] Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A 'rule of three'for fragment-based lead discovery*? Drug Discovery Today*, 2003, 8, 876-877.

[307] Lipinski, C. A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discovery Today: Techno*l., 2004, 1, 337-341.

[308] Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J. L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. *J. Chem. Inf. Model*., 2012, 52, 2864-75.

[309] Baker, M. Fragment-based lead discovery grows up. *Nat. Rev. Drug Discovery*, 2013, 12, 5-7.

[310] Goodford, P. J. A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *J. Med. Chem*., 1985, 28, 849-857.

[311] Lauri, G.; Bartlett, P. CAVEAT: A program to facilitate the design of organic molecules. *J. Comput. Aided Mol. Des*., 1994, 8, 51-66.

[312] Eisen, M. B.; Wiley, D. C.; Karplus, M.; Hubbard, R. E. HOOK: a program for finding novel molecular architectures that satisfy the chemical and steric requirements of a macromolecule binding site. Proteins. Struct. Funct. Genet., 1994, 19, 199-221.

[313] Steinbrecher, T. B.; Dahlgren, M.; Cappel, D.; Lin, T.; Wang, L.; Krilov, G.; Abel, R.; Friesner, R.; Sherman, W. Accurate Binding Free Energy Predictions in Fragment Optimization. *J. Chem. Inf. Model*., 2015, 55, 2411-2420

[314] Murray, C. W.; Erlanson, D. A.; Hopkins, A. L.; Keserü, G. M.; Leeson, P. D.; Rees, D. C.; Reynolds, C. H.; Richmond, N. J. Validity of ligand efficiency metrics. *Med. Chem. Lett*., 2014, 5, 616-618.

[315] Shultz, M. D. Improving the plausibility of success with inefficient metrics. *Med. Chem. Lett*., 2014, 5, 2-5.

[316] Wang, L.; Deng, Y.; Knight, J. L.; Wu, Y.; Kim, B.; Sherman, W.; Shelley, J. C.; Lin, T.; Abel, R. Modeling local structural rearrangements using FEP/REST: application to relative binding affinity predictions of CDK2 inhibitors. *J. Chem. Theory Comput*., 2013, 9, 1282-1293.

[317] Wang, L.; Berne, B. J.; Friesner, R. A. On achieving high accuracy and reliability in the calculation of relative protein−ligand binding affinities. *Proc. Natl. Acad. Sci. USA*., 2012, 109, 1937-1942.

[318] McCammon, J. A. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol*., 1998, 8, 245-249.

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

[320] 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.

[321] 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.

[322] Boresch, S.; Tettinger, F.; Leitgeb, M. Absolute binding free energies: a quantitative approach to their calculation. *J. Phys. Chem. B*., 2003, 107, 9535-9551.

[323] Jarzynski, C. Equilibrium free energy differences from nonequilibrium measurements: a master equation approach. Phys. Rev. E., 1997, 56, 5018.

[324] Ytreberg, F. M.; Zuckerman, D. Efficient use of nonequilibrium measurement to estimate free energy differences for molecular systems. *J. Comput. Chem*., 2004, 25, 1749-1759.

[325] Chang, C.E.; Gilson, M. K. Free energy, entropy, and induced fit in host-guest recognition: calculations with the second generation mining minima algorithm. J. Am. Chem. Soc., 2004, 126, 13156-13164.

[326] Izrailev, S.; Stepaniants, S.; Balsera, M.; Oono, Y.; Schulten, K. Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys. J.*, 1997, 72, 1568-1581.

[327] Fukunishi, Y.; Mikami, Y.; Nakamura, H. The filling potential method: a method for estimating the free energy surface for protein-ligand docking. *J. Phys. Chem. B*., 2003, 107, 13201-13210.

[328] Woo, H.J.; Roux, B. Calculation of absolute protein-ligand binding free energy from computer simulations. Proc. Natl. Acad. Sci. USA., 2005, 102, 6825-6830.

[329] Srinivasan, J.; Cheatham, T. E. III; 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.

[330] Gilson, M. K.; Rashin, A.; Fine, R.; Honig, B. On the calculation of electrostatic interactions in proteins. *J. Mol. Biol*., 1985, 184, 503-516.

[331] Das, D.; Koh, Y.; Tojo,Y.; Ghosh, A. K.; Mitsuya, H. Prediction of potency of protease inhibitors using free energy simulations with polarizable quantum mechanics-based ligand charges and a hybrid water model. *J. Chem. Inf. Model*., 2009, 49, 2851-2862.

[332] Meng, H.; Li, C.; Wang, Y.; Chen, G. Molecular dynamics simulation of the allosteric regulation of eIF4A protein from the open to closed state, induced by ATP and RNA substrates. *PLoS One*, 2014, 9, e86104.

[333] Cheatham, T. E. 3rd; Srinivasan, J.; Case, D. A.; Kollman, P. A. Molecular dynamics and continuum solvent studies of the stability of polyG-polyC and polyA-polyT DNA duplexes in solution. *J. Biomol. Struct. Dyn*., 1998, 16, 265-280.

[334] Ponder, J. W.; Wu, C. J.; Ren, P. Y.; Pande, V. S.; Chodera, J. D.; Schnieders, M. J.; Haque, I.; Mobley, D. L.; Lambrecht, D. S.; DiStasio, R. A. Jr.; Head-Gordon, M.; Clark, G. N. I.; Johnson, M. E.; Head-Gordon, T. Current status of the AMOEBA polarizable force field. *J. Phys. Chem. B*, 2010, 114, 2549-2564.

[335] Lopes, P. E. M.; Huang, J.; Shim, J.; Luo, Y.; Li, H.; Roux, B.; MacKerell, A. D., Jr. Polarizable force field for peptides and proteins based on the classical drude oscillator. J. Chem. Theory Comput., 2013, 9, 5430-5449.

[336] Banks, J. L.; Kaminski, G. A.; Zhou, R. H.; Mainz, D. T.; Berne, B. J.; Friesner, R. A. Parametrizing a polarizable force field from Ab Initio data. I. The fluctuating point charge model. J. Chem. Phys., 1999, 110, 741-754.

[337] Gresh, N.; Cisneros, G. A.; Darden, T. A.; Piquemal, J. P. Anisotropic, polarizable molecular mechanics studies of inter- and intramoecular interactions and ligand-macromolecule complexes. A bottom-up strategy. *J. Chem. Theory Comput*., 2007, 3, 1960-1986.

[338] Schmidt, J. R.; Yu, K.; McDaniel, J. G. Transferable next-generation force fields from simple liquids to complex materials. Acc. Chem. Res., 2015, 48, 548-556.

[339] Rotenberg, B.; Salanne, M.; Simon, C.; Vuilleumier, R. From localized orbitals to material properties: building classical force fields for nonmetallic condensed matter systems. *Phys. Rev. Lett*., 2010, 104, 138301.

[340] Grimme, S. A. General quantum mechanically derived force field (QMDFF) for molecules and condensed phase simulations. J. Chem. Theory Comput., 2014, 10, 4497-4514.

[341] Weinhold, F.; Landis, C. R. Valency and Bonding. Cambridge University Press, Cambridge, UK, 2005.

[342] Medders, G. R.; Paesani, F. On the interplay of the potential energy and dipole moment surfaces in controlling the infrared activity of liquid water. J. Chem. Phys., 2015, 142, 212411.

[343] Leontyev, I. V.; Stuchebrukhov, A. A. Accounting for Electronic Polarization in Non-Polarizable Force Fields. *Phys. Chem. Chem. Phys*., 2011, 13, 2613-2626.

[344] Reddy, M. R.; Erion, M. D. Relative binding affinities of fructose-1,6-bisphosphatase inhibitors calculated using a quantum mechanics-based free energy perturbation method. *J. Am. Chem. Soc*., 2007, 129, 9296-9297.

[345] Rathore, R. S.; Reddy, R. N.; Kondapi, A. K.; Reddanna, P.; Reddy, M. R. Use of quantum mechanics/molecular mechanics-based FEP method for calculating relative binding affinities of FBPase inhibitors for type-2 diabetes. *Theor. Chem. Acc*., 2012, 131, 1096.

[346] Rod, T. H.; Ryde, U. Quantum mechanical free energy barrier for an enzymatic reaction. *Phys. Rev. Lett*., 2005, 94, 138302.

[347] Luzhkov, V.; Warshel, A. Microscopic models for quantum mechanical calculations of chemical processes in solutions: LD/AMPAC and SCAAS/AMPAC calculations of solvation energies. *J. Comput. Chem*., 1992, 13, 199-213.

[348] Duarte, F.; Amrein, B. A.; Blaha-Nelson, D.; Kamerlin, S. C. L. Recent advances in QM/mm free energy calculations using reference potentials. *Biochim. Biophys. Acta, Gen. Subj*., 2015, 1850, 954-965.

[349] König, G.; Boresch, S. Non-Boltzmann sampling and Bennett's acceptance ratio method: how to profit from bending the rules. *J. Comput. Chem*., 2011, 32, 1082-1090.

[350] Senn, H. M.; Thiel, W. QM/MM methods for biomolecular systems. *Angew. Chem., Int. Ed*., 2009, 48, 1198-1229.

[351] Hu, H.; Yang, W. Free energies of chemical reactions in solution and in enzymes with Ab Initio quantum mechanics/molecular mechanics methods. *Annu. Rev. Phys. Chem*., 2008, 59, 573-601.

[352] Heimdal, J.; Ryde, U. Convergence of QM/MM free-energy perturbations based on molecular-mechanics or semiempirical simulations. *Phys. Chem. Chem. Phys*., 2012, 14, 12592-12604.

[353] Woods, C. J.; Manby, F. R.; Mulholland, A. J. An efficient method for the calculation of quantum mechanics/molecular mechanics free energies. *J. Chem. Phys*., 2008, 128, 014109.

[354] Fox, S. J.; Pittock, C.; Tautermann, C. S.; Fox, T.; Christ, C.; Malcolm, N. O. J.; Essex, J. W.; Skylaris, C.-K. Free energies of binding from large-scale first-principles quantum mechanical calculations: application to ligand hydration energies. *J. Phys. Chem. B*, 2013, 117, 9478-9485.

[355] Genheden, S.; Cabedo Martinez, A. I.; Criddle, M. P.; Essex, J. W. Extensive all-atom Monte Carlo sampling and QM/MM corrections in the SAMPL4 hydration free energy challenge. J. Comput. Aided Mol. Des., 2014, 28, 187-200.

[356] König, G.; Pickard, F. C.; Mei, Y.; Brooks, B. R. Predicting hydration free energies with a hybrid QM/MM approach: an evaluation of implicit and explicit solvation models in SAMPL4. *J. Comput. Aided Mol. Des*., 2014, 28, 245-257.

[357] König, G.; Hudson, P. S.; Boresch, S.; Woodcock, H. L. Multiscale free energy simulations: an efficient method for connecting classical MD simulations to QM or QM/MM free energies using non-Boltzmann Bennett reweighting schemes. J. *Chem. Theory Comput.*, 2014, 10, 1406-1419.

[358] Sampson, C.; Fox, T.; Tautermann, C. S.; Woods, C.; Skylaris, C.-K. A "Stepping stone" approach for obtaining quantum free energies of hydration. *J. Phys. Chem. B*, 2015, 119, 7030-7040.

[359] Mikulskis, P.; Cioloboc, D.; Andrejic, M.; Khare, S.; Brorsson, J.; Genheden, S.; Mata, R. A.; Söderhjelm, P.; Ryde, U. Free-energy perturbation and quantum mechanical study of SAMPL4 octa-acid host-guest binding energies. J. Comput. Aided Mol. Des., 2014, 28, 375-400.

[360] Beierlein, F. R.; Michel, J.; Essex, J. W. A simple QM/MM approach for capturing polarization effects in protein ligand binding free energy calculations. *J. Phys. Chem. B*, 2011, 115, 4911-4926.

[361] Woods, C. J.; Shaw, K. E.; Mulholland, A. J. Combined quantum mechanics/molecular mechanics (QM/MM) simulations for protein−ligand complexes: free energies of binding of water molecules in influenza neuraminidase. *J. Phys. Chem. B*, 2015, 119, 997-1001.

[362] Genheden, S.; Ryde, U.; Söderhjelm, P. Binding affinities by free-energy perturbation using QM/MM with a large QM system and polarizable MM model. J. *Comput. Chem*., 2015, 36, 2114-2124.

[363] Olsson, M. A.; Söderhjelm, P.; Ryde, U. Converging ligand-binding free energies obtained with free-energy perturbations at the quantum mechanical level. *J. Comput. Chem*., 2016.

[364] Hansen, N.; van Gunsteren, W. F. Practical aspects of free-energy calculations: a review. *J. Chem. Theory Comput*., 2014, 10, 2632-2647.

[365] Mobley, D. L.; Klimovich, P. V. Perspective: alchemical free energy calculations for drug discovery. *J. Chem. Phys*., 2012, 137, 230901.

[366] Aqvist, J.; Hansson, T. On the validity of electrostatic linear response in polar solvents. *J. Phys. Chem*., 1996, 100, 9512-9521.

[367] Almlöf, M.; Carlsson, J.; Åqvist, J. Improving the accuracy of the linear interaction energy method for solvation free energies. *J. Chem. Theory Comput*., 2007, 3, 2162-2175.

[368] Korth, M.; Lüchow, A.; Grimme, S. Toward the exact solution of the electronic schrödinger equation for noncovalent molecular interactions: worldwide distributed quantum Monte Carlo calculations. *J. Phys. Chem. A*, 2008, 112, 2104-2109.

[369] Tkatchenko, A.; Alfè, D.; Kim, K. S. First-principles modeling of non-covalent interactions in supramolecular systems: the role of many-body effects. *J. Chem. Theory Comput*., 2012, 8, 4317-4322.

[370] Hongo, K.; Cuong, N. T.; Maezono, R. The importance of electron correlation on stacking interaction of adenine-thymine base-pair step in βDNA: a quantum monte carlo study. *J. Chem. Theory Comput*., 2013, 9, 1081-1086.

[371] Benali, A.; Shulenburger, L.; Romero, N. A.; Kim, J.; von Lilienfeld, O. A. Application of diffusion monte carlo to materials dominated by van der Waals interactions. *J. Chem. Theory Comput.*, 2014, 10, 3417-3422.

[372] Dubecký, M.; Jurečka, P.; Derian, R.; Hobza, P.; Otyepka, M.; Mitas, L. Quantum Monte Carlo methods describe noncovalent interactions with subchemical accuracy. *J. Chem. Theory Comput*., 2013, 9, 4287-4292.

[373] Al-Hamdani, Y. S.; Alfè, D.; von Lilienfeld, O. A.; Michaelides, A. Water on BN doped benzene: a hard test for exchange-correlation functionals and the impact of exact exchange on weak binding. *J. Chem. Phys*., 2014, 141, 18C530.

[374] Gillan, M. J.; Manby, F. R.; Towler, M. D.; Alfè, D. Assessing the accuracy of quantum Monte Carlo and density functional theory for energetics of small water clusters. *J. Chem. Phys*., 2012, 136, 244105.

[375] Gurtubay, I. G.; Needs, R. J. Dissociation energy of the water dimer from quantum Monte Carlo calculations. *J. Chem. Phys*., 2007, 127, 124306.

[376] Santra, B.; Michaelides, A.; Fuchs, M.; Tkatchenko, A.; Filippi, C.; Scheffler, M. On the accuracy of density-functional theory exchange correlation functionals for H bonds in small water clusters. II. The water hexamer and van der Waals interactions. *J. Chem. Phys.*, 2008, 129, 194111.

[377] Ma, J.; Alfè, D.; Michaelides, A.; Wang, E. The water-benzene interaction: insight from electronic structure theories. *J. Chem. Phys*., 2009, 130, 154303.

[378] Korth, M.; Grimme, S.; Towler, M. D. The lithium-thiophene riddle revisited. *J. Phys. Chem. A*, 2011, 115, 11734-11739.

[379] Rezáč, J.; Dubecký, M.; Jurečka, P.; Hobza, P. Extensions and applications of the A24 data set of accurate interaction energies. *Phys. Chem. Chem. Phys*., 2015, 17, 19268-19277.

[380] Xu, J.; Deible, M. J.; Peterson, K. A.; Jordan, K. D. Correlation consistent Gaussian basis sets for H, B-Ne with Dirac-Fock AREP pseudopotentials: applications in Quantum Monte Carlo calculations. J. *Chem. Theory Comput*., 2013, 9, 2170-2178.

[381] Petruzielo, F. R.; Toulouse, J.; Umrigar, C. J. Approaching chemical accuracy with quantum Monte Carlo. *J. Chem. Phys*., 2012, 136.

[382] Manten, S.; Lüchow, A. On the accuracy of the fixed-node diffusion quantum Monte Carlo method. *J. Chem. Phys*., 2001, 115, 5362-5366.

[383] Grossman, J. C. Benchmark quantum Monte Carlo calculations. *J. Chem. Phys*., 2002, 117, 1434-1440.

[384] Nemec, N.; Towler, M. D.; Needs, R. J. Benchmark all-electron ab initio quantum Monte Carlo calculations for small molecules. *J. Chem. Phys*., 2010, 132, 034111.

[385] Morales, M. A.; McMinis, J.; Clark, B. K.; Kim, J.; Scuseria, G. E. Multideterminant wave functions in quantum Monte Carlo. J. *Chem. Theory Comput*., 2012, 8, 2181-2188.

[386] Mazziotti, D. A. Parametrization of the two-electron reduced density matrix for its direct calculation without the many-electron wave function. *Phys. Rev. Lett*., 2008, 101, No. 253002.

[387] DePrince, A. E.; Kamarchik, E.; Mazziotti, D. A. Parametric two electron reduced-density-matrix method applied to computing molecular energies and properties at nonequilibrium geometries. *J. Chem. Phys*., 2008, 128, 234103.

[388] Mazziotti, D. A. Two-electron reduced density matrix as the basic variable in many-electron quantum chemistry and physics. *Chem. Rev*., 2012, 112, 244-262.

[389] Chan, G. K.-L.; Head-Gordon, M. Highly correlated calculations with a polynomial cost algorithm: a study of the density matrix renormalization group. *J. Chem. Phys*., 2002, 116, 4462.

[390] Chan, G. K.-L. An algorithm for large scale density matrix renormalization group calculations. *J. Chem. Phys*., 2004, 120, 3172.

[391] Sharma, S.; Chan, G. K.-L. Spin-adapted density matrix renormalization group algorithms for quantum chemistry. *J. Chem. Phys.,* 2012, 136, 124121.

[392] Rasch, K. M.; Hu, S.; Mitas, L. M. Fixed-node errors in quantum Monte Carlo: interplay of electron density and node nonlinearities. *J. Chem. Phys*., 2014, 140, 041102.

[393] Yang, W. T. Direct calculation of electronic density in density functional theory. *Phys. Rev. Lett.*, 1991, 66, 1438-1441.

[394] Goedecker, S. Linear scaling electronic structure methods. Rev. Mod. Phys., 1999, 71, 1085-1123.

[395] Bowler, D. R.; Miyazaki, T. O(N) Methods in electronic structure calculations. *Rep. Prog. Phys.*, 2012, 75, 036503.

[396] Karplus, M.; Kuriyan, J. Molecular dynamics and protein function. *Proc. Natl. Acad. Sci. USA*., 2005, 102, 6679-6685.

[397] Dror, R. O.; Dirks, R. M.; Grossman, J. P.; Xu, H. F.; Shaw, D. E. Biomolecular simulation: a computational microscope for molecular biology. *Annu. Rev. Biophys*., 2012, 41, 429-452.

[398] Mikulskis, P.; Genheden, S.; Wichmann, K.; Ryde, U. A semiempirical approach to ligand-binding affinities: dependence on the hamiltonian and corrections. J. Comput. Chem., 2012, 33, 1179-1189.

[399] Söderhjelm, P.; Kongsted, J.; Ryde, U. Ligand affinities estimated by quantum chemical calculations. *J. Chem. Theory Comput*., 2010, 6, 1726-1737.

[400] Smith, R. Peer review: a flawed process at the heart of science and journals. *J. R. Soc. Med*., 2006, 99, 178-182.

[401] Emerson, G. B.; Warme, W. J.; Wolf, F. M.; Heckman, J. D.; Brand, R. A.; Leopold, S. S. Testing for the presence of positive-outcome bias in peer review. a randomized controlled trial. *Arch. Intern. Med*., 2010, 170, 1934-1939.

[402] Pople, J. A.; Beveridge, D. L. Approximate Molecular Orbital Theory. McGraw-Hill, New York, 1970.

[403] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model. *J. Am. Chem. Soc*., 1985, 107, 3902-3909.

[404] Stewart, J. J. P. Optimization of parameters for semiempirical methods II. Applications. *J. Comput. Chem*., 1989, 10, 221-264.

[405] Thiel, W.; Voityuk, A. A. Extension of MNDO to d-orbitals parameters and results for the second-row elements and for the zinc group. *J. Phys. Chem*., 1996, 100, 616-626.

[406] Weber, W.; Thiel, W. Orthogonalization corrections for semiempirical methods. *Theor. Chem. Acc*., 2000, 103, 495-506.

[407] Elstner, M.; Frauenheim, T.; Kaxiras, E.; Seifert, G.; Suhai, S. A self-consistent charge density-functional based tight-binding scheme for large biomolecules. *Phys. Status Solidi B*, 2000, 217, 357-376.

[408] Elstner, M.; Seifert, G. Density functional tight binding. *Philos. Trans. R. Soc. A*, 2014, 372, 20120483.

[409] Gaus, M.; Cui, Q.; Elstner, M. Density functional tight binding: application to organic and biological molecules. *WIREs Comput. Mol. Sci*., 2014, 4, 49-61.

[410] Cui, Q.; Elstner, M. Density functional tight binding: values of semi-empirical methods in an ab initio era. *Phys. Chem. Chem. Phys*., 2014, 16, 14368-14377.

[411] Maia, J. D. C.; Urquiza Carvalho, G. A.; Mangueira, C. P., Jr.; Santana, S. R.; Cabral, L. A. F.; Rocha, G. B. GPU linear algebra libraries and GPGPU programming for accelerating MOPAC semiempirical quantum chemistry calculations. *J. Chem. Theory Comput*., 2012, 8, 3072-3081.

[412] Wu, X.; Koslowski, A.; Thiel, W. Semiempirical quantum chemical calculations accelerated on a hybrid multicore CPU-GPU computing platform. *J. Chem. Theory Comput*., 2012, 8, 2272-2281.

[413] Gao, J. L.; Truhlar, D. G.; Wang, Y. J.; Mazack, M. J. M.; Loffler, P.; Provorse, M. R.; Rehak, P. Explicit polarization: a quantum mechanical framework for developing next generation force fields. *Acc. Chem. Res*., 2014, 47, 2837-2845.

[414] Giese, T. J.; Huang, M.; Chen, H. Y.; York, D. M. Recent advances toward a general purpose linear-scaling quantum force field. *Acc. Chem. Res*., 2014, 47, 2812-2820.

[415] Niklasson, A. M. N.; Cawkwell, M. J. Generalized extended lagrangian Born-Oppenheimer molecular dynamics. *J. Chem. Phys*., 2014, 141, 164123.

[416] Thiel, W. Perspectives on semiempirical molecular orbital theory. *Adv. Chem. Phys*., 1996, 93, 703-757.

[417] Sattelmeyer, K. W.; Tirado-Rives, J.; Jorgensen, W. Comparison of SCC-DFTB and NDDO-based semiempirical molecular orbital methods for organic molecules. J. Phys. Chem. A, 2006, 110, 13551-13559.

[418] Otte, N.; Scholten, M.; Thiel, W. Looking at self-consistent-charge density functional tight binding from a semiempirical perspective. J. Phys. Chem. A, 2007, 111, 5751-5755.

[419] Kruger, T.; Elstner, M.; Schiffels, P.; Frauenheim, T. Validation of the density functional based tight-binding approximation method for the calculation of reaction energies and other data. *J. Chem. Phys*., 2005, 122, 114110.

[420] Riley, K. E.; Hobza, P. Investigations into the nature of halogen bonding including symmetry adapted perturbation theory analyses. *J. Chem. Theory Comput*., 2008, 4, 232-242.

[421] Řezáč, J.; Hobza, P. A Halogen-bonding correction for the semiempirical PM6 method. *Chem. Phys. Lett*., 2011, 506, 286-289.

[422] Császár, A. G.; Allen, W. D.; Schaefer, H. F. In pursuit of the ab initio limit for conformational energy prototypes. *J. Chem. Phys*., 1998, 108, 9751-9764.

[423] Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Koch, H.; Olsen, J.; Wilson, A. K. Basis-set convergence in correlated calculations on Ne, N_2, and H_2O. *Chem. Phys. Lett*., 1998, 286, 243-252.

[424] Boys, S.; Bernardi, F. The calculation of small molecular interactions by the differences of separate total energies. Some procedures with reduced errors. *Mol. Phys*., 1970, 19, 553-566.

[425] Řezáč, J.; Riley, K. E.; Hobza, P. Extensions of the S66 data set: more accurate interaction energies and angular-displaced nonequilibrium geometries. J. Chem. Theory Comput., 2011, 7, 3466-3470.

[426] Řezáč, J.; Hobza, P. Describing noncovalent interactions beyond the common approximations: how accurate is the "gold standard," CCSD(T) at the complete basis set limit? *J. Chem. Theory Comput*., 2013, 9, 2151-2155.

[427] Christensen, A. S.; Elstner, M.; Cui, Q. Improving intermolecular interactions in DFTB3 using extended polarization from chemical-potential equalization. J. Chem. Phys., 2015, 143, 084123.

[428] Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the accuracy limits of local pair natural orbital coupled-cluster theory. *J. Chem. Theory Comput.*, 2015, 11, 1525-1539.

[429] Bissantz, C.; Kuhn, B.; Stahl, M. A medicinal chemist's guide to molecular interactions. *J. Med. Chem*., 2010, 53, 5061-5084.

[430] Guerra, C. F.; Bickelhaupt, F. M.; Snijders, J. G.; Baerends, E. J. The nature of the hydrogen bond in DNA base pairs: the role of charge transfer and resonance assistance. Chem. Eur. J., 1999, 5, 3581-3594.

[431] Wolters, L. P.; Bickelhaupt, F. M. Halogen bonding versus hydrogen bonding: a molecular orbital perspective. Chemistry Open, 2012, 1, 96-105.

[432] Dixon, S. L.; Merz, K. M., Jr. semiempirical molecular orbital calculations with linear system size scaling. J. Chem. Phys., 1996, 104, 6643-6649.

[433] Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. The semiempirical quantum mechanical scoring function for in silico drug design. *ChemPlusChem*, 2013, 78, 921-931.

[434] Mucs, D.; Bryce, R. A. The application of quantum mechanics in structure-based drug design. *Expert Opin. Drug Discovery*, 2013, 8, 263-276.

[435] Yilmazer, N. D.; Korth, M. Comparison of molecular mechanics, semi-empirical quantum mechanical, and density functional theory methods for scoring protein-ligand interactions. J. *Phys. Chem. B*, 2013, 117, 8075-8084

[436] Nikitina, E.; Sulimov, V.; Zayets, V.; Zaitseva, N. Semiempirical calculations of binding enthalpy for protein-ligand complexes. *Int. J. Quantum Chem*., 2004, 97, 747-763.

[437] Bikadi, Z.; Hazai, E. Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock. *J. Cheminf*., 2009, 1, 15.

[438] Fanfrlík, J.; Bronowska, A. K.; Řezáč, J.; Přenosil, O.; Konvalinka, J.; Hobza, P. A Reliable docking/scoring scheme based on the semiempirical quantum mechanical PM6-DH2 method accurately covering dispersion and H-bonding: HIV-1 Protease with 22 ligands. J. *Phys. Chem. B*, 2010, 114, 12666-12678.

[439] Peters, M. B.; Raha, K.; Merz, K. M., Jr. Quantum mechanics in structure-based drug design. *Curr. Opin. Drug Discovery Dev*., 2006, 9, 370-379.

[440] Raha, K.; Peters, M. B.; Wang, B.; Yu, N.; Wollacott, A. M.; Westerhoff, L. M.; Merz, K. M., Jr. The role of quantum mechanics in structure-based drug design. *Drug Discovery Today*, 2007, 12, 725-731.

[441] Cattoni, D. I.; Chara, O.; Kaufman, S. B.; González Flecha, F. L. Cooperativity in binding processes: new insights from phenomenological modeling. *PLoS one*, 2013, 10, e0146043.

[442] Garbett, N. C.; Chaires, J. B. Binding: a polemic and rough guide. *Methods Cell Biol*., 2008, 84, 3-23.

[443] Gutfreund, H. Kinetics for the life sciences: Receptors, transmitters and catalysts. Cambridge University Press, Cambridge, 1995.

[444] Hager, G. L.; McNally, J. G.; Misteli, T. Transcription dynamics. *Mol. Cell*., 2009, 35, 741-753.

[445] Freiburger, L.; Miletti, T.; Zhu, S.; Baettig, O.; Berghuis, A.; Auclair, K.; Mittermaier, A. Substrate-dependent switching of the allosteric binding mechanism of a dimeric enzyme. *Nat Chem Biol*., 2014; 10: 937-942.

[446] Dodes Traian, M. M.; Cattoni, D. I.; Levi, V.; González Flecha, F. L. A two-stage model for lipid modulation of the activity of integral membrane proteins. *PLoS one*, 2010, 7, e39255.

[447] Hillier, B. J.; Rodriguez, H. M.; Gregoret, L. M. Coupling protein stability and protein function in Escherichia coli CspA. *Fold Des*., 1998, 3, 87-93.

[448] Levi, V.; Rossi, J, P. F. C.; Castello, P. R.; González Flecha, F. L. Structural significance of the plasma membrane calcium pump oligomerization. *Biophys. J*., 2002, 82, 437-446.

[449] Shcheynikov, N.; Son, A.; Hong, J. H.; Yamazaki, O.; Ohana, E.; Kurtz, I.; Shin, D. M.; Muallem, S. Intracellular Cl^- as a signaling ion that potently regulates Na^+/HCO_3^- transporters. *Proc. Natl. Acad. Sci. USA*, 2015A, 112, E329-337.

[450] Ruscio, J. Z.; Kumar, D.; Shukla, M.; Prisant, M. G.; Murali, T. M.; Onufriev, A. V. Atomic level computational identification of ligand migration pathways between solvent and binding site in myoglobin. *Proc. Natl. Acad. Sci. USA*., 2008, 105, 9204-9209.

[451] Sakaguchi, M.; Shingo, T.; Shimazaki, T.; Okano, H. J.; Shiwa, M.; Ishibashi, S.; Oguro, H.; Ninomiya, M.; Kadoya, T.; Horie, H.; Shibuya, A.; Mizusawa, H.; Poirier, F.; Nakauchi, H.; Sawamoto, K.; Okano, H. A carbohydrate-binding protein, Galectin-1, promotes proliferation of adult neural stem cells. *Proc. Natl. Acad. Sci. USA*., 2006, 103, 7112-7117.

[452] Edelstein, S. J.; Le Novere, N. Cooperativity of allosteric receptors. *J. Mol. Biol*., 2013, 425, 1424-1432.

[453] Barcroft, J.; Hill, A.V. The nature of oxyhæmoglobin, with a note on its molecular weight. *J. Physiol*., 1910, 39, 411-428.

[454] Hill, A.V. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. *J. Physiol*., 1910, 40, iv-vii

[455] Holt, J. M.; Ackers, G. K. The Hill coefficient. Inadequate resolution of cooperativity in human hemoglobin. *Methods Enzymol*., 2009, 455, 193-212.

[456] Weber, G. Protein interactions. Chapman and Hall, New York, 1992.

[457] Ruzicka, F. J.; Frey, P. A. Kinetic and spectroscopic evidence of negative cooperativity in the action of lysine 2,3-aminomutase. *J. Phys. Chem. B*, 2010, 114, 16118-16124.

[458] Abeliovich, H. An empirical extremum principle for the Hill coefficient in ligand-protein interactions showing negative cooperativity. *Biophys. J*., 2005, 89, 76-79.

[459] Koshland, D. E.; Hamadani, K. Proteomics and models for enzyme cooperativity. *J. Biol. Chem.*, 2002, 277, 46841-46844.

[460] Thorsteinsdottir, H. B. Computational analysis of protein-ligand binding: from single continuous trajectories to multiple parallel simulations. Ph. D. Disertation, Basel University, 2010.

[461] Deng, Y.; Roux, B. Computations of standard binding free energies with molecular dynamics simulations. *J. Phys. Chem.*, 2009, 113, 2234-2246.

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

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

[464] Kirkwood, J. G. Statistical mechanics of fluid mixtures. *J. Chem. Phys.*, 1935, 3, 300-313.

[465] Zwanzig, R. W. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *J. Chem. Phys.*, 1954, 22, 1420-1426.

[466] 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 second generation forcefield for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.*, 1995,117, 5179-5197.

[467] MacKerell, A.D.; Bashford, D.; Bellott M.; Dunbrack, Jr., R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.T.K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhorn, B.; Reiher, III, W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe M.; Wiórkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*, 1998, 102, 3586-3616.

[468] Van Gunsteren, W. F.; Billeter, S. R.; Eising, A. A.; Hünenberger, P. H.; Krueger, P.; Mark, A. E.; Scott, W. R. P.; Tironi, I. G. Biomolecular Simulation: The GROMOS96 manual and user guide, 1996, vdf Hochschulverlag AG an der ETH Zürich, and BIOMOS b.v. Zürich, Groningen.

[469] Bowers, K. J.; Chow, E.; Xu, H.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Moraes, M. A.; Sacerdoti, F. D.; Salmon, J. K.; Shan, Y.; Shaw, D. E. Scalable algorithms for molecular dynamics simulations on commodity clusters, *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, 2006, Tampa, Florida, USA.

[470] Lee, M. S.; Olson, M. A. Calculation of absolute protein-ligand binding affinity using path and endpoint approaches. *Biophys. J.*, 2006, 90, 864-877.

[471] Mobley, D. L.; Graves, A. P.; Chodera, J. D.; McReynolds, A. C.; Shoichet, B. K.; Dill, K. A. Predicting absolute binding free energies to a simple model site. *J. Mol.Biol.*, 2007, 371, 1118-1134.

[472] Tembe, B.; McCammon, J. A. Ligand-receptor interactions. *Comput. Chem*., 1984, 8, 281-283.

[473] Rao, B. G.; Kim, E. E.; Murcko, M. A. Calculation of solvation and binding free energy differences between VX-478 and its analogs by free energy perturbation and Amsol methods. *J. Comput. Aided. Mol. Des*., 1996, 10, 23-30.

[474] Oostenbrink, C.; van Gunsteren, W. F. Free energies of binding of polychlorinated biphenyls to the estrogen receptor from a single simulation. *Proteins*, 2004, 54, 237-46.

[475] 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*, 2007, Chipot, C.; Pohorille, A. Eds. Springer-Verlag, Berlin-Heidelberg.

[476] 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.

[477] Singh, N.; Warshel, A. Absolute binding free energy calculations: on the accuracy of computational scoring of protein-ligand interactions. *Prot. Struct., Funct. Bioinf.,* 2010, 78, 1705-1723.

[478] Hayes, J. M.; Skamnaki, V. T.; Archontis, G.; Lamprakis, C.; Sarrou, 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, KT5720, and staurosporine as phosphorylase kinase ATP-binding site inhibitors: The role of water molecules examined. *Prot. Struct. Funct. Bioinf.*, 2011, 79, 703-719.

[479] Archontis, G.; Simonson, T. Proton binding to proteins: a free-energy component analysis using a dielectric continuum model. *Biophys. J.*, 2005, 88, 3888-3904.

[480] Schutz, C. N.; Warshel, A. What are the dielectric constants of proteins and how to validate electrostatic models? *Prot. Struct. Funct. Genet.*, 2001, 44, 400-417.

[481] Aleksandrov, A.; Thompson, D.; Simonson, T. Alchemical free energy simulations for biological complexes: powerful but temperamental. *J. Mol. Recog.*, 2010, 23, 117-127.

[482] 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.

[483] Thompson, D.; Plateau, P.; Simonson, T. Free-energy simulations and experiments reveal long-range electrostatic interactions and substrate-assisted specificity in an aminoacyl-tRNAsynthetase. *Chembiochem.,* 2006, 7, 337-344.

[484] 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.

[485] Brooks, B. R.; Janezic, D.; Karplus, M. Harmonic-analysis of large systems. 1. Methodology. *J. Comput. Chem*., 1995, 16, 1522-1542.

[486] Karplus, M.; Kushick, J. N. Method for estimating the configurational entropy of macromolecules. *Macromol.,* 1981, 14, 325-332.

[487] Tidor, B.; Karplus, M. The contribution of vibrational entropy to molecular association - the dimerization of insulin. *J. Mol. Biol.*, 1994, 238, 405-414.

[488] Kongsted, J.; Ryde, U. An improved method to predict the entropy term with the MM/PBSA approach. *J. Comp. Aid. Mol. Des.*, 2009, 23, 63-71.

[489] 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.

[490] Page, C. S.; Bates, P. A. Can MM-PBSA calculations predict the specificities of protein kinase inhibitors? *J. Comput. Chem.,* 2006, 27, 1990-2007.

[491] Simonson, T. Electrostatics and dynamics of proteins. *Rep. Prog. Phys.*, 2003, 66, 737-787.

[492] Tamamis, P.; Morikis, D.; Floudas, C. A.; Archontis, G. Species specificity of the complement inhibitor compstatin investigated by all-atom molecular dynamics simulations. *Prot. Struct., Funct. Bioinf.*, 2010, 78, 2655-2667.

[493] 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.

[494] Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. Interaction models of water in relation to protein hydration. In: *Intermolecular forces: Proceedings of the Fourteenth Jerusalem Symposium on Quantum Chemistry and Biochemistry*, B. 1981, Pullman, Editor, pp. 331-342, Reidel Publ. Company, Dordrecht, Holland.

[495] Berendsen, H. J. C.; Grigera, J. A.; Straatsma, T. P. The missing term in effective pair potentials. *J. Phys. Chem., 1*987, 91, 6269-5271.

[496] Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.*, 1983, 79, 926-935.

[497] Mahoney, M. W.; Jorgensen, W. L. A five-site model for liquid water and the reproduction of the density anomaly by rigid, nonpolarizable potential functions. *J. Chem. Phys.,* 2000. 112, 8910-8922.

[498] Wang, J.; Wang, W.; Huo, S.; Lee, M.; Kollman, P. A. Solvation model based on weighted solvent accessible surface area. *J. Phys. Chem. B*, 2001, 105, 5055-5067.

[499] 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

[500] Roux, B. Implicit solvent models. In: *Computational Biochemistry and Biophysics*, Becker, 2001, O.M., Mackerell, A. D. Jr, Roux, B., Watanabe, M. Editors, pp. 133-151, Marcel Dekker Inc., New York

[501] Roux, B.; Simonson, T. Implicit solvent models. *Biophys. Chem.*, 1999, 78, 1-20.

[502] Schwab, F.; van Gunsteren, W. F.; Zagrovic, B. Computational study of the mechanism and the relative free energies of binding of

anticholesteremic inhibitors to squalene-hopene cyclase. *Biochemistry,* 2008, 47, 2945-2951.

[503] Harvey, M. J.; Giupponi, G.; De Fabritiis, G. ACEMD: Accelerating biomolecular simulations in the microsecond time scale. *J. Chem. Theor.Comput.,* 2009, 5, 1632-1639.

[504] Stone, J. E.; Hardy, D. J.; Isralewitz, B.; Schulten, K. GPU algorithms for molecular modeling. Chapter 16in: *Scientific Computing with Multicore & Accelerators*, Dongarra J.; Bader J.A; Kurzak J., 2011,351-371, Chapman & Hall/CRC Press.

[505] Boresch, S.; Bruckner, S. Avoiding the van der Waals endpoint problem using serial atomic insertion. *J. Comput. Chem.*, 2011, 32, 2449-2458.

[506] 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. *J. Mol. Biol,.* 2003, 330, 891-913.

[507] van Gunsteren, W. F.; Berendsen, H. J. C. Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew. Chem. Int. Ed. Engl.,* 1990, 29, 992-1023.

[508] Mark, A. E.; van Helden, S. P.; Smith, P. E.; Janssen, L. H. M.; van Gunsteren. W. F. Convergence properties of free energy calculations: α-cyclodextrin complexes as a case study. *J. Am. Chem. Soc*., 1994, 116, 6293-6302.

[509] Mark, A. E. Free energy perturbation calculations. In Schleyer, P. V. R., Allinger, N. L.; Clark, T.; Gasteiger, J.; Kollman, P. A.; Schaefer, H. F. III; Schreiner, P.R. editors, *Encyclopedia of computational chemistry*, 1998, 1070-1083, Wiley and Sons, Chichester.

[510] Dixit, S. B.; Chipot, C. Can absolute free energies of association be estimated from molecular mechanical simulations? The biotin-streptavidin system revisited. *J. Phys. Chem. A*, 2001, 105, 9795-9799.

[511] van Gunsteren, W. F.; Daura, X.; Mark, A. E. Computation of free energy. *Helv. Chim. Acta*, 2002, 85, 3113-3129.

[512] Chipot, C.; Pearlman, D. A. Free energy calculations. The long and winding gilded road. *Mol. Sim.,* 2002, 28, 1-12.

[513] Piana, S.; Klepeis, J. L.; Shaw, D. E. Assessing the accuracy of physical models used in protein-folding simulations: quantitative evidence from long molecular dynamics simulations. Current Opin. Struct. Biol., 2014, 24, 98-105.

[514] Marelius, J.; Hansson, T.; Aqvist, J. Calculation of ligand binding free energies from molecular dynamics simulations. Int. J. Quantum. Chem., 1988, 69, 77-88.

[515] Zhao, C.; Caplan, D. A.; Noskov, S. Y. Evaluations of the absolute and relative free energies for antidepressant binding to the aminoacid membrane transporter leut with free energy simulations. *J. Chem. Theory Comput*., 2010, 6, 1900-1914.

[516] Hulten, J.; Bonham, N. M.; Nillroth, U.; Hansson, T.; Zuccarello, G.; Bouzide, A.; Aqvist, J.; Classon, B.; Danielson, U. H.; Karlen, A.; Kvarnstrom, I.; Samuelsson, B.; Hallberg, A. Cyclic HIV-1 protease inhibitors derived from mannitol: synthesis, inhibitory potencies, and computational predictions of binding affinities. *J. Med. Chem*., 1997, 40, 885-897.

[517] Homeyer, N.; Gohlke, H. FEW: a workflow tool for free energy calculations of ligand binding. *J. Comput. Chem*., 2013, 34, 965-973.

[518] Marelius, J.; Kolmodin, K.; Feierberg, I.; Aqvist, J. Q: an md program for free energy calculations and empirical valence bond simulations in biomolecular systems. *J. Mol. Graphics Modell*., 1998, 16, 213-225.

[519] van Gunsteren, W. F.; Berendsen, H. J. C. Groningen molecular simulation (GROMOS) library manual, Biomos; Nijenborgh: Groningen, The Netherlands, 1987.

[520] Aqvist, J.; Mowbray, S. L. Sugar recognition by a glucose/galactose receptor: evaluation of binding energetics from molecular dynamics simulations. *J. Biol. Chem*., 1995, 270, 9978-9981.

[521] Ersmark, K.; Feierberg, I.; Bjelic, S.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. C2-symmetric inhibitors of plasmodium falciparum Plasmepsin II: synthesis and theoretical predictions. *Bioorg. Med. Chem*., 2003, 11, 3723-3733.

[522] Ersmark, K.; Feierberg, I.; Bjelic, S.; Hamelink, E.; Hackett, F.; Blackman, M. J.; Hulten, J.; Samuelsson, B.; Aqvist, J.; Hallberg, A. Potent inhibitors of the plasmodium falciparum enzymes Plasmepsin I and II devoid of Cathepsin D inhibitory activity. *J. Med. Chem*., 2004, 47, 110-122.

[523] Marelius, J.; Graffner-Nordberg, M.; Hansson, T.; Hallberg, A.; Aqvist, 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.

[524] Ersmark, K.; Samuelsson, B.; Hallberg, A. Plasmepsins as potential targets for new antimalarial therapy. *Med. Res. Rev*., 2006, 26, 626-666.

[525] Ljungberg, K. B.; Marelius, J.; Musil, D.; Svensson, P.; Norden, B.; Aqvist, J. Computational Modelling of Inhibitor Binding to Human Thrombin. *Eur. J. Pharm. Sci*., 2001, 12, 441-446.

[526] Valiente, P. A.; Gil, A.; Batista, P. R.; Caffarena, E. R.; Pons, T.; Pascutti, P. G. New parameterization approaches of the LIE method to improve free energy calculations of plmii-inhibitors complexes. *J. Comput. Chem*., 2010, 31, 2723-2734.

[527] Miranda, W. E.; Noskov, S. Y.; Valiente, P.A. Improving the LIE method for binding free energy calculations of protein-ligand complexes. *J. Chem. Inf. Model*., 2015, 55, 1867-1877.

[528] 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.,* 2010, 31, 797-810.

[529] Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free-energies using macroscopic solvent models. *J. Phys. Chem*., 1994, 98, 1978-1988.

[530] Janezic, D.; Brooks, B. R. Harmonic-analysis of large systems. 2. Comparison of different protein models. *J. Comput. Chem*., 1995, 16, 1543-1553.

[531] Janezic, D.; Venable, R. M.; Brooks, B. R. Harmonic-analysis of large systems. 3. Comparison with molecular dynamics. *J. Comput. Chem*., 1995, 16, 1554-1566.

[532] Chong, L. T.; Duan, Y.; Wang, L.; Massova, I.; Kollman, P. A. Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proc. Natl. Acad. Sci. USA*., 1999, 96, 14330-14335.

[533] Massova, I.; Kollman, P. A. Computational alanine scanning to probe protein-protein interactions: a novel approach to evaluate binding free energies. *J. Am. Chem. Soc*., 1999, 121, 8133-8143.

[534] Reyes, C. M.; Kollman, P. A. Investigating the binding specificity of U1A-RNA by computational mutagenesis. *J. Mol. Biol*., 2000, 295, 1-6.

[535] Prime, version 3.0, Schrödinger, LLC, New York, 2011.

[536] Jacobson, M. P.; Pincus, D. L.; Rapp, C, S.; Day, T. J. F.; Honig, B.; Shaw, D. E.; Friesner, R. A. A hierarchical approach to all-atom protein loop prediction. *Proteins Struct. Funct. Bioin*f., 2004, 55, 351-367.

[537] Miller B. R. III; McGee, T. D; Swails, J. M.; Homeyer, N.; Gohlke, H.; Roitberg, A. E. MMPBSA.py: An efficient program for end-state free energy calculations. *J. Chem. Theor. Comput*., 2012, 8, 3314-3321.

[538] Case, D. A.; Darden, T. A.; Cheatham, T. E. III; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R.C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman P.A. AMBER 12, 2012, University of California, San Francisco.

[539] Liu, X.; Liu, J.; Zhu, T.; Zhang, L.; He, X.; Zhang, J. Z. H. PBSA_E: A PBSA-based free energy estimator for protein-ligand binding affinity. *J. Chem. Inf. Model*., 2016, 56, 854-861.

[540] Polydoridis, S.; Leonidas, D. D.; Oikonomakos, N. G.; Archontis G. Recognition of ribonuclease A by 3'-5'-pyrophosphate-linked dinucleotide inhibitors: a molecular dynamics/continuum electrostatics analysis. *Biophys. J.*, 2007, 92, 1659-1672.

[541] Lyubartsev, A. P.; Forrisdahl, O. K.; Laaksonen, A. Solvation free energies of methane and alkali halide ion pairs: an expanded ensemble molecular dynamics simulation study. *J. Chem. Phys*., 1998, 108, 227-233.

[542] Lyubartsev, A. P.; Laaksonen, A.; Vorontsovvelyaminov, P. N. Free energy calculations for Lennard-Jones systems and water using the expanded ensemble method. A Monte Carlo and molecular dynamics simulation study. *Mol. Phys*., 1994, 82, 455-471.

[543] Aberg, K. M.; Lyubartsev, A. P.; Jacobsson, S. P.; Laaksonen, A. Determination of solvation free energies by adaptive expanded

ensemble molecular dynamics. *J. Chem. Phys*., 2004, 120, 3770-3776.

[544] Fajer, M.; Hamelberg, D.; McCammon, J. A. Replica-exchange accelerated molecular dynamics (REXAMD) applied to thermodynamic integration. J. Chem. Theory. Comput., 2008, 4, 1565-1569.

[545] Straatsma, T. P.; Berendsen, H. J. C. Free energy of ionic hydration: Analysis of a thermodynamic integration technique to evaluate free energy differences by molecular dynamics simulations. J. Chem. Phys., 1988, 89, 5876-5886.

[546] Chaitanya, A. K.; Koppisetty, M. F.; Lyubartsev, A. P.; Nyholm, P.-G. Binding energy calculations for hevein-carbohydrate interactions using expanded ensemble molecular dynamics simulations. J. Comput. Aided Mol. Des., 2015, 29, 13-21

[547] Huang, D.; Caflisch, A. Small molecule binding to proteins: affinity and binding/unbinding dynamics from atomistic simulations. *ChemMedChem*, 2011, 6, 1578-1580.

[548] Huang, D.; Caflisch, A. The free energy landscape of small molecule unbinding. *PLoS Comput. Biol*., 2011, 7, e1002002.

[549] Grubmüller, H.; Heymann, B.; Tavan, P. Ligand binding: molecular mechanics calculation of the streptavidin-biotin rupture force. *Science*, 1996, 271, 997-999.

[550] Isralewitz, B.; Gao, M.; Schulten, K. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol*., 2001, 11, 224-230.

[551] Isralewitz, B.; Baudry, J.; Gullingsrud, J.; Kosztin, D.; Schulten, K. Steered molecular dynamics investigations of protein function. *J. Mol. Graphics Modell*., 2001, 19, 13-25.

[552] Lu, H.; Isralewitz, B.; Krammer, A.; Vogel, V.; Schulten, K. Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys. J*., 1998, 75, 662-671.

[553] Giorgino, T.; Fabritiis, G. D. A High-throughput steered molecular dynamics study on the free energy profile of ion permeation through Gramicidin A. *J. Chem. Theory Comput*., 2011, 7, 1943-1950.

[554] Gwan, J. F.; Baumgaertner, A. Cooperative transport in a potassium ion channel. *J. Chem. Phys*., 2007, 127, 045103.

[555] Hénin, J.; Schulten, E. T. K.; Chipot, C. Diffusion of glycerol through escherichia coli aquaglyceroporin GlpF. *Biophys. J*., 2008, 94, 832-839.

[556] Jensen, M. Ø.; Park, S.; Tajkhorshid, E.; Schulten, K. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci.,* 2002, 99, 6731-6736.

[557] Xu, Y.; Shen, J.; Luo, X.; Silman, I.; Sussman, J. L.; Chen, K.; Jiang, H. How does huperzine A enter and leave the binding gorge of acetylcholinesterase? Steered molecular dynamics simulations. *J. Am. Chem. Soc*., 2003, 125, 11340-11349.

[558] Shen, L.; Shen, J.; Luo, X.; Cheng, F.; Xu, Y.; Chen, K.; Arnold, E.; Ding, J.; Jiang, H. Steered molecular dynamics simulation on the binding of NNRTI to HIV-1RT. *Biophys, J*., 2003, 84, 3547-3563.

[559] Binnig, G.; Quate, C. F.; Berger, C. H. Atomic force microscope. *Phys. Rev. Lett*., 1986, 56, 930-933.

[560] Simmons, R.; Finer, J.; Chu, S.; Spudich, J. Quantitative measurements of force and displacement using an optical trap. *Biophys. J*., 1996, 70, 1813-1822.

[561] Amblard, F.; Yurke, B.; Pargellis, A.; Leibler, S. Magnetic manipulation for studying local rheology and micromechanical properties of biological systems. *Rev. Sci. Instrum*., 1996, 67, 1-10.

[562] Florin, E. L.; Moy, V. T.; Gaub, H. E. Adhesion forces between individual ligand-receptor pairs. *Science*, 1994, 264, 415-417.

[563] Kumar, S.; Li, M. S. Biomolecules under mechanical force. *Phys. Reports*, 2010, 486, 1-74.

[564] Li, M. S. Secondary structure, mechanical stability and location of transition state of proteins. *Biophys. J*., 2007, 93, 2644-2654.

[565] Rief, M.; Gautel, M.; Oesterhelt, F.; Fernandez, J. M.; Gaub, H. E. Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science*, 1997, 276, 1109-1112.

[566] Li, M. S.; Mai, B. K. Steered molecular dynamics-a promising tool for drug design. *Current Bioinformatics*, 2012, *7,* 342-351

[567] Mai, B. K.; Viet, M. H.; Li, M. S. Top-leads for swine influenza A/H1N1 virus revealed by steered molecular dynamics approach. *J. Chem. Inf. Model*., 2010, 50, 2236-2247.

[568] Petrek, M.; Otyepka, M.; Banas, P.; Kosinova, P.; Koca, J.; Damborsky, J. CAVER: a new tool to explore routes from protein clefts, pockets and cavities. *BMC Bioinformatics*, 2006, 7, 316.

[569] 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, 10388-10397.

[570] Vuong, Q. V.; Nguyen, T. T.; Li, M. S. A new method for navigating optimal direction for pulling ligand from binding pocket: application to ranking binding affinity by steered molecular dynamics. *J. Chem. Inf. Model*., 2015, 55, 2731-2738

[571] Andér, M.; Luzhkov, V. B.; Aqvist, J. Ligand binding to the voltage-gated Kv1.5 potassium channel in the open state—docking and computer simulations of a homology model. *Biophys. J*., 2008, 94, 820-831.

[572] Carlsson, J.; Boukharta, L.; Aqvist, J. Combining docking, molecular dynamics and the linear interaction energy method to predict binding modes and affinities for non-nucleoside inhibitors to HIV-1 reverse transcriptase. *J. Med. Chem*., 2008, 51, 2648-2656.

[573] Nervall, M.; Hanspers, P.; Aqvist, J. Predicting binding modes from free energy calculations. J. Med. Chem., 2008, 51, 2657-2667.

[574] Waszkowycz, B. Towards improving compound selection in structure-based virtual screening. *Drug Discov. Today*, 2008, 13, 219-226.

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

[576] Ahmed, M.; Sadek, M. M.; Abouzid, K. A.; Wang, F. In silico design, extended molecular dynamic simulations and binding energy calculations for a new series of dually acting inhibitors against EGFR and HER2. J. Mol. Graph. Model., 2013, 44, 220-231.

[577] Zhao, H.; Caflisch, A. Molecular dynamics in drug design. Eur. J. Med. Chem., 2015, 91, 4-14.

[578] Spiliotopoulos, D.; Caflisch, A. Molecular dynamics simulations of bromodomains reveal binding-site flexibility and multiple binding modes of the natural ligand acetyl-lysine. *Isr. J. Chem*., 2014, 54, 1084-1092.

[579] Huang, D.; Caflisch, A. Library screening by fragment-based docking. *J. Mol. Recognit*., 2010, 23, 183-193.

[580] Pieffet, G. P. The application of molecular dynamics simulation techniques and free energy calculations to predict protein-protein and protein-ligand interactions, Ph. D. Dissertation, Groningen University, 2005

[581] Villa, A.: Mark, A. E. Calculation of free energy of solvation for neutral analogs of amino acid side chains. J. *Comp. Chem*., 2002, 23, 548-553.

[582] Shirts, M. R.; Pitera, J. W.; Swope, W. C.; Pande, V, S. Extremely precise free energy calculations of amino acid side chain analogs: comparison of common molecular mechanics force fields for proteins. J. Chem. Phys., 2003, 119, 5740-5761.