Computational Approaches to the Chemical Equilibrium Constant in Protein-ligand Binding

Joel José Montalvo-Acosta and Marco Cecchini

DOI: 10.1002/minf.201600052
Abstract: The physiological role played by protein-ligand recognition has motivated the development of several computational approaches to the ligand binding affinity. Some of them, termed rigorous, have a strong theoretical foundation but involve too much computation to be generally useful. Some others alleviate the computational burden by introducing strong approximations and/or empirical calibrations, which also limit their general use. Most importantly, there is no straightforward correlation between the predictive power and the level of approximation introduced. Here, we present a general framework for the quantitative interpretation of protein-ligand binding based on statistical mechanics. Within this framework, we derive self-consistently the fundamental equations of some popular approaches to the binding constant and pinpoint the inherent approximations. Our analysis represents a first step towards the development of variants with optimum accuracy/efficiency ratio for each stage of the drug discovery pipeline.

Keywords: molecular recognition · protein-ligand binding · binding affinity · free energy calculations · rational drug design

1 Introduction

Protein-ligand binding is central to many biological processes that take place in living systems. Some ligands inhibit protein function, some others promote it by stabilizing a large isomerization of their receptor, thereby controlling key cell-signaling pathways. Understanding how ligands bind to biomolecules is of fundamental importance not only for the basic fields of biophysics and biochemistry, but also for applied disciplines such as medicinal chemistry and pharmacology. Although kinetics may strongly affect the yield of binding in cellular and other non-equilibrium environments, the primary factors that govern molecular recognition are of thermodynamic nature. In particular, the value of the binding equilibrium constant, which is dictated by the standard free energy change on complexation, is the quantity of interest. Being able to access the binding constant accurately and from first principles would provide a chemical understanding of protein-ligand recognition, thus unraveling guidelines for drug design.

Over the last decades, innovation costs in the Pharma industry have exceedingly increased and have recently approached 4 billion U.S. dollars per FDA-approved drug. Lead optimization alone is estimated to involve about 150 million U.S. dollars per hit compound. Reliable predictions of the protein-ligand binding affinity by computation would greatly reduce these costs and boost a more efficient development of new pharmaceuticals. However, the calculation of the binding constant in protein-ligand binding poses an outstanding theoretical and computational challenge. For instance, there exists no method to solve this problem when binding of the ligand involves a global structural change of the receptor, which is crucially important in ligand-modulated allosteric equilibria. When the protein response is more local, the calculation of the standard free energy of binding is possible and several computational approaches at various levels of sophistication have been developed.

Among the available methods, the so-called rigorous approaches evaluate the free energy of binding based on simplified descriptions of the reaction path, which typically involves a series of non-physical intermediates. In the chemical route first introduced by Jorgensen and later improved by others, the ligand is decoupled reversibly from its environment with the free energy of binding accessed by perturbation theory. Alternatively, the ligand can be physically separated from the receptor by forcing the unbinding along a one-dimensional reaction coordinate, and the free energy of binding measured by umbrella sampling. In both cases, to ensure configurational overlap between consecutive steps of the microscopic transformation, these approaches involve a large number of intermediates between the end-points, which results in a large computational effort per free energy determination; see Figure 1 (on top). Moreover, to improve the efficiency of sampling, the transformation is typically performed in the presence of appropriately chosen restraints whose contribution to the binding affinity must be evaluated by additional computation. As a rule of thumb, rigorous free energy approaches may grant an approximate output rate of one determination per week, which is clearly not suited for screening purposes.

To increase the efficiency of the calculation, more simplified computational approaches have been developed to focus only on the end-points of the binding reaction; see Figure 1 (middle). These methods reduce the computational burden by using approximated expressions for the solvent free energy, which can be efficiently evaluated by a continuum treatment of the solvent (i.e. an implicit solvent model), or in the limit of the linear response approximation. Prominent examples of end-points methods are MM/PBSA (molecular mechanics [MM] with Poisson-Boltzmann [PB] and surface area [SA]), which was originally developed by Kollman et al., and the linear interaction energy (LIE) approach by Åqvist. In MM/PBSA, the free energy of binding is estimated from the total change in the gas-phase internal energy, the solvation free energy and the configurational entropy upon protein-ligand associa-

[a] J.J. Montalvo-Acosta, M. Cecchini
Laboratoire d’Ingéniérie des Fonctions Moléculaires ISIS, UMR 7006 CNRS
Université de Strasbourg, F-67083 Strasbourg Cedex, France
*e-mail: mcecchini@unistra.fr
tion, with the solvation free energy by Poisson-Boltzmann calculations plus a term accounting for the nonpolar contribution. In contrast, the LIE approach considers the solute/solvent interactions explicitly and estimates the binding free energy from changes in the electrostatic and van der Waals components of the ligand-surroundings interaction energy when the ligand is transferred from the solution bulk to the binding site of the receptor.\textsuperscript{[20]} Both approaches rely on a detailed description of the bound and the unbound states and explicitly include conformational effects by averaging over structural ensembles generated e.g. by Molecular Dynamics. These semi-rigorous approaches are clearly more efficient and grant an approximate output rate of one free-energy determination per day. For this reason, they have been fairly popular in the Pharma industry in both the hit-to-lead and lead-optimization phases.

Screening libraries or databases of small-molecule compounds require significantly more simplified schemes to allow for an efficient evaluation of thousands of millions of binding modes. Because of the strong computational restraints, empirical approaches based e.g. on molecular docking, which focus exclusively on the bound state (Figure 1, bottom) and generally neglect the internal flexi-

\[\text{Figure 1. Different schemes to access the absolute free energy of binding in protein-ligand association.}\]

\[\text{Empirical}\]

\[\text{Rigorous}\]

\[\text{End-point}\]

\[\text{Joel José Montalvo-Acosta received a BSc degree in Pharmaceutical Chemistry and a MSc degree in Pharmaceutical Sciences with distinctions from the University of Cartagena (Colombia) in 2013. He is now pursuing a PhD in theoretical chemistry at the University of Strasbourg under supervision of Dr. Marco Cecchini. His research interests focus on the development of theoretical and computational approaches to study molecular recognition and their application in computer-aided drug design.}\]

\[\text{Marco Cecchini is head of the laboratory of “Molecular Function and Design” at the Institut de Science et d’Ingénierie Supramoléculaires (ISIS) of the University of Strasbourg. He received his BSc and MSc degrees in Chemistry from the University of Bologna (Italy) and obtained a Ph.D degree in Natural Sciences from the University of Zurich (Switzerland). After a postdoctoral training with Martin Karplus between the University of Strasbourg and Harvard University, he was appointed junior group leader at ISIS (France). In 2011, he was awarded a “Chaire d’Excellence” CNRS prize and appointed assistant professor at the University of Strasbourg. His research interests span the domains of life science and material science with emphasis on the elucidation of the principles of chemical design by theoretical and computational approaches.}\]
More thorough reviews on computational approaches for protein-ligand binding can be found elsewhere.\cite{20,21,25,26}

2 Theory

Let us consider the spontaneous association of a protein molecule (P) with a ligand (L) to form a non-covalent complex (PL) in aqueous solution

\[ P + L \rightleftharpoons PL. \]  

(1)

At chemical equilibrium, the chemical potentials of the product and the reactants equalize so that

\[ \Delta \mu_\text{b} = \mu_\text{f} - (\mu_\text{r} + \mu_\text{i}) = 0. \]  

(2)

By separating out the volume dependence of the chemical potentials in Eq. 2, which is customary done by introducing an arbitrary state of reference or standard state, as

\[ \mu_i(V, T) = \mu_i^0(T) + kT \ln \left( \frac{C_i}{C_i^\text{eq}} \right) \]  

(3)

with \( T \) being the absolute temperature, \( C_i^\text{eq} \) the standard concentration, and \( \mu_i^0 \) and \( C_i \) the standard chemical potential and the molar concentration of the \( i \)-th solute, respectively, and rearranging, it yields

\[ \exp \left( -\frac{\Delta \mu_i^0}{kT} \right) = \frac{C_i^\text{eq} (C_i)}{C_i^0 C_i^\text{eq}} = K_{\mu_i} C_i^\text{eq} \]  

(4)

which shows that the ratio between the equilibrium concentrations of the product over the reactants is volume independent (i.e. it is independent of the initial solute concentrations) and is therefore a chemical equilibrium constant. Importantly, the value of \( K_{\mu_i} \) or customarily its inverse \( K_{\mu_i}^{-1} \) which corresponds to the initial concentration of ligand for which the probability of binding at equilibrium is one-half (i.e. \( C_i = C_i^\text{eq} \)), sets an absolute scale of ligand-binding affinities. Hence, Eq. 4 quantifies the strength of protein-ligand binding through the evaluation of \( \Delta \mu_i^0 \).

The grand-canonical approach. Straightforward access to the chemical potential difference in Eq. 4 is provided by a statistical mechanics treatment of the binding reaction in the canonical ensemble \( (N, V, T) \). In the limit of idealized solution behavior (i.e. the particle independent ansatz) and at constant temperature \( T \) and volume \( V \), the chemical potential of the solute is

\[ \mu_i(V, T) = -kT \ln \frac{q_i(V, T)}{N_i} \]  

(5)

with \( q_i \) and \( N_i \) being the molecular partition function and the number of solute molecules. Introducing the rigid-rotor harmonic-oscillator (RRHO) approximation in vacuum and in the limit of Born-Oppenheimer (BO) hypothesis, the molecular partition function in Eq. 5 can be further separated into translational, rotational, vibrational and electronic contributions, all having a closed form.\cite{27} Also, by incorporating the net effect of the solvent in the electronic contribution through the evaluation of the potential of mean force corresponding to the solvation free energy of the solute in its configuration at the minimum of the potential energy (\( X_0 \)), Eq. 5 yields

\[ \mu_i(V, T) = \mu_i^0(V, T) + W_{\text{solv}}(X_0) \]  

(6)

which provides the chemical potential of the solute as a correction to the harmonic result in vacuum (\( \mu_i^0 \)). By evaluating Eq. 6 for each component of the binding reaction at the standard 1M concentration (Figure 2) and introducing the results into Eq. 2, a rigorous estimate of \( \Delta \mu_i^0 \) is obtained, which provides numerical access to the binding constant in the limit of the RRHO approximation.

Figure 2. Schematic representation of the initial (free) and final (bound) states for the protein-ligand association used to compute \( \Delta \mu_i^0 \) in the canonical approach; see Main Text.

The grand-canonical approach. A conceptually different approach to the binding constant in solution goes through a statistical mechanics treatment in the grand canonical ensemble at constant pressure (\( \mu, p, T \)). This formalism\cite{28} provides an alternative expression for the standard chemical potential of the solute

\[ \mu_i^0(T) = -kT \ln \left( \frac{Q(p, T)}{V} \right) \]  

(7)

with \( Q \) being an effective partition function of the solute in the solvent at the constant pressure \( p \). Introducing this result into Eq. 4 yields

\[ \exp \left( -\frac{\Delta \mu_i^0}{kT} \right) = \frac{\left( \frac{Q(p, T)}{V} \right)}{\left( \frac{Q_i^0}{V_i} \right)} \left( \frac{C_i^0}{C_i} \right) = K_{\mu_i} \]  

(8)

which shows that the binding constant can be expressed by an effective partition function ratio. Because in the limit of infinite dilution each effective partition function can be approximated as
Minireview

$$Q_i(p, T) \approx \frac{Q_{i,1}}{Q_{i,0}}$$

(9)

with $Q_{i,1}$ and $Q_{i,0}$ being the canonical partition functions of a binary solution with $N$ solvent molecules and one or no solute molecule, and the solution volume can be assumed as unchanged upon ligand binding (i.e. $V_p = V_{ri}$), Eq. 8 yields

$$K_{eq}^C = \frac{Q_{i,1}/Q_{i,0}}{V_L}$$

(10)

which shows that the value of the binding constant is related to the reversible work (or the free energy) to make the ligand disappear from the protein binding site (at the numerator) minus the work to make the ligand disappear from a box of solvent per unit of volume (at the denominator). Interestingly, Eq. 10 indicates that protein-ligand binding (Eq. 1) can be viewed as a transfer of ligand from the solution bulk to the binding site of the receptor (Figure 3), or as a partition equilibrium of the ligand

$$L_{free} \leftrightarrow L_{bound}$$

(11)

Finally, if the ligand is small relative to the receptor, the protein contributions in Eq. 10 almost cancel out and the numerator can be considered as an effective partition function of the ligand in the bound state, such that

$$K_{eq}^C = \frac{Q_i(p)}{Q_i/V_i}$$

(12)

with $Q_i(p)$ being the volume-independent effective partition function of one ligand bound to the protein. As we shall see, this result sets the ground to most statistical mechanics approaches to the binding constant.

2.1 Rigorous Statistical Mechanics Approaches

In the limit of highly dilute solutions, the results of Eq. 6 and Eq. 12 provide access to the binding equilibrium constant without too much approximation. In practice though, the use of Eq. 6 is strongly limited in protein-ligand problems by the evaluation of the solvation free energy of a large and flexible solute as the protein alone or the complex, which is computationally challenging, and so far this approach has been successfully applied only to small peptide systems. On the other hand, by factorizing out the kinetic energy contribution from the numerator and the denominator of Eq. 12, which cancel out as the total number of degrees of freedom is conserved, the binding constant can be expressed in terms of configurational integrals over the relevant portions of the configurational space accessible to the ligand in the bound and the unbound states yielding

$$K_{eq}^C = \frac{\int_{bulk} dL \int dX \exp(-\beta U)}{\int_{bulk} dL \delta(r_l - r^*) \int dX \exp(-\beta U)}$$

(13)

with $U$ being the total potential energy of the system, $\beta = 1/kT$, and $L$ and $X$ the coordinates of the ligand and the remaining (solvent and protein) atoms; note that the $\delta$ function at the denominator has been introduced to make the bulk configurational integral volume independent. Importantly, Eq. 13 provides numerical access to the binding constant by computer simulations and can be considered as the master equation for most rigorous approaches to protein-ligand binding. Two of them are briefly reviewed below.

2.1.1 Alchemical Free Energy Perturbation

An effective strategy to compute protein-ligand binding free energies based on Eq. 13 was originally introduced by Jorgensen and later improved by Gilson. This approach, which is usually referred to as “double annihilation” or “double decoupling”, solves Eq. 13 by making use of a thermodynamic cycle in which the ligand is transformed into a fictitious non-interacting body both in the bound and the unbound states. Such an alchemical transformation, termed annihilation, is achieved through the use a hybrid Hamiltonian with a coupling parameter $\lambda$ of the form $U(\lambda) = (1-\lambda) U_i + \lambda U_p$ with $U_p$ and $U_i$ being the total potential energy of the system with a non-interacting (decoupled) and a full-interacting (coupled) ligand. By introducing an intermediate state in which the ligand is transferred to the gas phase, Eq. 13 yields

$$K_{eq}^C = \frac{\int_{bulk} dL \int dX \exp(-\beta U_i)}{\int_{bulk} dL \delta(r_l - r^*) \int dX \exp(-\beta U_b)}$$

(14)

where the first factor of the right-hand side is related to the reversible work for decoupling the ligand in the bound
state and the second factor to the work for decoupling it in
the solution bulk, which are both numerically accessible by
multi-stage free energy perturbation molecular dynamics
(FEP/MD) simulations.[31] Despite the elegance of the strat-
ey, straightforward applications of Eq. 14 are often imprac-
tical as e.g. the ligand in the highly decoupled states be-
comes free to “wander” in the volume of the simulation
box, which seriously hinders statistical convergence.[31] To
improve sampling efficiency, more recent implementations
of the double decoupling method make use of external re-
straints which reduce the configurational space accessible
to the ligand. The sequential activation/deactivation of re-
straints on the position, the orientation, and the internal
configuration of the ligand significantly improves statisti-
cal convergence but introduces a series of additional inter-
mediates along the reaction path, which involve extra com-
pilation. Following one of the most recent implementa-
tions,[32] which denotes the harmonic restraints on the
translation, rotation, and the conformation of the ligand as
$u_o, u_r, u_c$. Eq. 13 becomes

$$
K_b C = \frac{\int_{\text{site}} dL \int dX \exp(-\beta U_i)}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o))}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_r))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_c))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))}
$$

(15)

Eq. 15 shows that numerical access to the binding con-
stant is provided by the sequential confinement of the con-
formation, the orientation and the position of the ligand in
the binding site; the decoupling of the ligand from the
protein environment; the release of the positional and ori-
etentational restraints in the gas phase; the re-coupling
of the ligand in the bulk; and the release of the conformational
restraint in solution. In this approach, the standard free
energy of binding is determined by summing up the revers-
ible work associated with each of the eight steps of a com-
plex microscopic transformation.

2.1.2 Potential of Mean Force

An alternative approach, which is also based on Eq. 13 con-
ists on measuring the free-energy of binding/unbinding
along a simplified representation of the reaction path by
a potential of mean force (PMF) calculation. In this case,
the ligand is physically separated from the receptor and
the free energy of binding is obtained by umbrella sam-
ping over one or more geometric reaction coordinates,
typically the Euclidean distance between its initial position
in the binding site and an arbitrary point in the bulk.[32] Be-
cause of the high dimensionality of the true reaction coor-
dinate for binding, the conformational freedom of the
ligand that may differ substantially in the bound and un-
bound states, and the intrinsic difficulty of sampling the
translational degrees of freedom in the unbound state, this
approach is used in combination with restraints.[33] In
analogy to the alchemical route, a series of intermediates
corresponding to both orientational and conformational
confinement/release of the ligand in the bound and the
unbound states are introduced, which provides the follow-
ing expression for the binding constant

$$
K_b C = \frac{\int_{\text{site}} dL \int dX \exp(-\beta U_i)}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o))} \times
\frac{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_r))}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_c))}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{site}} dL \int dX \exp(-\beta(U_i))}{\int_{\text{site}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_r))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_c))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_r))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_c))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_r))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_c))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))} \times
\frac{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i))}{\int_{\text{bulk}} dL \int dX \exp(-\beta(U_i + u_o + u_r + u_c))}
$$

(16)

Compared to Eq. 15, the fourth factor on the right-hand
side of the equation above is unique to the PMF approach
and corresponds to the physical separation of the ligand
from the protein in the presence of configurational and ori-
etentational restraints.

In the absence of large conformational changes of the
protein, Eqs. 15 and 16 provide rigorous, first-principle
access to the binding constant and have been successfully
used in a number of cases.[32,34,35] However, these ap-
proaches require thorough statistical sampling over a large
number of intermediates and are computationally very de-
manding. Thus, despite their accuracy, they are not suited
to handle large databases of ligands and have found, so
far, little room in the Pharma industry, although this trend
may change in the near future.[36]
2.2 Simplified End-Points Approaches

To facilitate the computational task without compromising too much the quality of the results, simplified approaches based on Eq. 6 and Eq. 12 have been also developed. A prominent group of them explicitly considers the conformational dynamics of the protein-ligand complex and the ligand alone in solution and accesses the binding constant by focusing on the relevant initial and final states (end points) of the reaction (Figure 1). In this section, two amongst the most popular end-points strategies for protein-ligand binding, i.e. the MM/PBSA and the linear interaction energy (LIE) methods, are shortly reviewed. Based on the statistical mechanics framework introduced above, their fundamental equations are re-derived with emphasis on the approximations introduced.

2.2.1 The MM/PBSA Method

The MM/PBSA approach aims at the binding constant from the numerical evaluation of the absolute chemical potentials of the ligand, the protein and the complex in isolation (Figure 2). Starting from Eq. 6 and separating out the enthalpy versus entropy contributions to the chemical potential in vacuum, it yields

$$
\mu_i(V, T) = (3n - 3) kT + U(X_i) - TS_i(V) + W_{bulk}(X_i)
$$

with $S_i$ being the configurational entropy in vacuum that includes contributions from the translational, rotational and vibrational degrees of freedom and $W_{bulk}$ the solvation free energy of the solute, which accounts for all enthalpic and entropic contributions of the solvent. Assuming that $W_{bulk}$ can be accessed by a continuum model of water such PB/SA, which evaluates the polar contribution by solving the Poisson-Boltzmann equation and the non-polar contribution as a linear function of the solvent accessible surface area (SASA), Eq. 17 yields

$$
\mu_i(V, T) = (3n - 3) kT + U(X_i) + G_{BSA}(X_i) - TS_i(V)
$$

where the configurational entropy can be evaluated in the harmonic limit by classical statistical mechanics in combination with normal-mode analysis including quantum corrections. To account for part of the anharmonicity, which can be significant in proteins, the electronic energy of the solute as well as its solvation free energy are often accessed by ensemble averages based on sampling from room-temperature Molecular Dynamics in a box of explicit water; note that in this case the electronic energy contribution is obtained by subtracting the temperature-dependent vibrational energy from the average potential energy of the solute in vacuum, as $-D_e = \langle U \rangle - \sum 1/2kT$, with $\kappa$ being the total number of internal degree of freedom of the solute. Introducing this result into Eq. 18, it yields

$$
\mu_i(V, T) = \frac{3}{2} n kT + \langle U \rangle + \langle G_{BSA} \rangle - TS_i(V)
$$

with the configurational entropy corrected for anharmonicity by a quasi-harmonic vibrational analysis. Using the notation in the original paper by Kollman, i.e. decomposing the force field energy into bonded ($E_{bond}$), electrostatic ($E_{elec}$), and van der Waals ($E_{vdw}$) contributions and splitting the solvation free energy into polar ($G_{pol}$) and nonpolar ($G_{np}$) terms, Eq. 19 gives

$$
\mu_i(V, T) = \frac{3}{2} n kT + E_{bond} + E_{elec} + E_{vdw} + G_{pol} + G_{np} - TS_i(V).
$$

Eq. 20 is the master equation for the MM/PBSA approach; note that the first term on the right-hand side corresponds to the total kinetic energy of the solute, which was missing in the original MM/PBSA formulation as pointed out by Gohlke and Case. Evaluating Eq. 20 at the standard 1M concentration for the ligand, the protein and the complex separately (Figure 2), the standard chemical potential difference and therefore the binding constant are straightforwardly accessed.

In practice, to solve Eq. 20 explicit-solvent MD simulations of the ligand, the protein and the complex are carried out and ensemble averages of the force field energy of the three solutes in the gas phase and the configurational-dependent solvation free energy are evaluated as arithmetic averages over a large series of MD snapshots. Alternatively, configurational ensembles for the uncomplexed reactants can be generated from the trajectory of the complex only, by removing selectively the atoms of the protein or the ligand. Surprisingly, this simplified version of MM/PBSA, named the one-average variant, was shown to yield more accurate results than the original strategy, perhaps due to the exact cancellation of the bonded energy in Eq. 20. Great effort was put in recent years to improve both the accuracy and the efficiency of MM/PBSA by benchmarking and optimizing the calculation of the individual contributions in Eq. 20. To this aim, various continuum-electrostatics models including e.g. the generalized Born (GB) model in the MM/GBSA variant have been tested and benchmarked searching for improved estimates of the polar solvation term ($G_{pol}$). Similarly, alternative charge schemes for computing the electrostatic contribution ($E_{elec}$) or different approaches to evaluate the configurational entropy of the solute e.g. by quasi-harmonic analyses have been reported. In general, the accuracy of the predictions was found to be fairly system-dependent, which makes it difficult to draw conclusions on the performance of the individual variants. Finally, by replacing the MM terms by a quantum mechanical model, a significantly improved correlation with the experimental binding affinity results was obtained. In this case, a hybrid QM/MM approach with the atoms of the ligand assigned to the QM
region and the rest treated as MM, was used to sample the configurational ensembles for the evaluation of Eq. 20, which significantly deteriorates the performance of the calculation.

2.2.2 The Linear Interaction Energy (LIE) Method

An alternative simplified approach to the binding constant, which treats the protein-ligand binding reaction as a ligand partition equilibrium (Eq. 11), can be derived from Eq. 12. Introducing the RRHO approximation and incorporating the partition function contributions from the protein and the solvent in the electronic energy of the ligand in the bound and the unbound states, Eq. 12 yields

$$K_q = \frac{q_{CM}q_{rock}e^{-\beta W_{cm}}}{q_{tr}/V q_{rot}e^{-\beta W_{rot}}}$$

(21)

with $q_{CM}$ and $q_{rock}$ corresponding to oscillations and rocking of the ligand in the bound state, $q_{tr}/V$ and $q_{rot}$ to translations (per unit of volume) and rotations of the free ligand, and $W_{cm}$ and $W_{rot}$ to the reversible work for transferring the ligand from the gas phase to the protein binding site and the solution bulk, respectively. Note that Eq. 21 includes no contribution from the internal vibrations or the potential energy of the ligand in vacuum, which effectively cancel out in the limit of rigid ligands. By extracting the logarithm and multiplying by $-kT$, Eq. 21 yields

$$\Delta \mu_q^c = -kT \log \zeta + W_{cm} - W_{rot}$$

(22)

with $\zeta$ corresponding to the fraction of translational and rotational motion left to the ligand in the bound state relative to its free rotation in the unbound state in solution. Importantly, Eq. 22 indicates that the standard free energy of binding can be accessed from the difference in the “solvation” free energy of the ligand in the protein and the solvent environments plus a contribution corresponding to the entropic confinement. Provided that accurate estimates of $W_{cm}$ and $W_{rot}$ can be determined e.g. by FEP, Eq. 21 is essentially equivalent to Eq. 13 and provides a quantitative estimate of the binding constant in the harmonic limit. However, this result would come with no computational advantage. To simplify the calculation of the binding constant, the idea developed in the linear interaction energy (LIE) approach, which follows from linear response theory,[50,51] is that both polar and non-polar contributions to the solvation free energies in Eq. 22 can be approximated by linear functions of the mean electrostatic and van der Waals interaction energies of the ligand with the surroundings. In fact, in the limit of a linear response of the solution to changes in the local electric field, e.g. the appearance of a charged solute, it can be formally shown[19] that the polar contribution to the solvation free energy equals one half of the mean solute-solvent electrostatic contribution as

$$W_{pol} = \frac{1}{2} \langle U_{elec}^{pol} \rangle$$

(23)

where the brackets $\langle \rangle$ indicate a thermodynamic average of the ligand-surroundings (l/s) interaction energy. Furthermore, based on the observations that the experimental free energy of solvation for various hydrocarbons in water is approximately linear with the length of the carbon chain,[52] and that the corresponding solute-solvent van der Waals energies from computer simulations are also linear with the number of carbon atoms, Åqvist et al. assumed that the non-polar contribution to the solvation free energy could be approximated as a linear function of the mean van der Waals interaction energy as

$$W_{non} \approx \alpha \langle U_{vdw}^{pol} \rangle$$

(24)

with $\alpha$ being an adjustable parameter subject to empirical calibration. If so, by expressing the solvation free energy of the ligand in the bound and the unbound states as a sum of polar (Eq. 23) and non-polar (Eq. 24) contributions, Eq. 22 yields

$$\Delta \mu_q^c = \frac{1}{2} \left[ \langle U_{elec}^{pol} \rangle_{site} - \langle U_{elec}^{pol} \rangle_{bulk} \right] + \alpha \left[ \langle U_{vdw}^{pol} \rangle_{site} - \langle U_{vdw}^{pol} \rangle_{bulk} \right] + \gamma$$

(25)

which provides the master equation for the LIE approach.[19] In this expression, $\gamma$ includes the entropic confinement contribution, which can be evaluated numerically in the limit of the RRHO model (Eq. 21) or determined empirically by fitting on experimental binding data. To account for (minor) deviations of the polar term from the exact (linear response) scaling factor of 1/2, a more general expression of LIE includes an additional fitting parameter $\beta$, which was shown to improve the accuracy of the computational predictions.[13] Note that unlike the original derivation of LIE, Eq. 25 was obtained in the limit of the harmonic approximation, which would restrain the validity of the approach to rigid ligands. Although this assumption is not strictly required in LIE, the treatment above provides an explicit expression for the entropic confinement contribution, which would be otherwise hidden in the empirical coefficient of the non-polar term.

In the limit of the linear response theory, Eq. 25 solves the protein-ligand problem by measuring the mean ligand/ surroundings interaction energy in the bound and the unbound states on a series of snapshots extracted from room-temperature MD simulations of the fully solvated complex and the free ligand in solution. In this approach, the strongest approximations regard both the validity of the linear response assumption and the rather simplistic
idea that the non-polar contribution to the binding free energy, which includes hydrophobic effects and both repulsive and dispersive solute/solvent interactions, can be extracted from the analysis of the non-electrostatic component of the ligand/surroundings interaction. Significant effort was made to validate the former hypothesis e.g. by comparing LIE results with rigorous FEP calculations, and the linear response approximation was found to be accurate for both the solvent and protein environments. More difficult is the validation of the second assumption, which involves the determination of the parameter $\alpha_i$, and possibly $\gamma$ when absolute binding free energies are of interest. Because the physical nature of these parameters is unclear and their value is force-field, ligand and even protein dependent, their existence introduces a significant degree of empiricism, which has hindered the development of a "universal" and fully transferable LIE parameterization. Nonetheless, the fact that the intermolecular energies from simulations of the end-points are in some cases sufficient to predict absolute binding free energies with an accuracy of $<1$ kcal/mol from experiments is absolutely remarkable and justifies the use of LIE in computer-aided drug discovery.

From a practical viewpoint, the implementation of LIE requires extensive configurational sampling of both the complex and the free ligand in solution typically by Molecular Dynamics or Monte Carlo simulations with an explicit treatment of the solvent, which makes this approach not suitable for high-throughput screening. To increase the computational performance, Huang and Caflisch developed the Linear Interaction Energy with Continuum Electrostatics (LIECE), where the MD sampling is replaced by energy minimization plus finite-difference Poisson calculations for a rigorous treatment of both the protein and the ligand desolvation energies. When applied to $\beta$-secretase (BACE) and HIV-1 protease, a two-parameter LIECE model was shown to reach a predictive power of $<1$ kcal/mol relative to experiments, while being about two orders of magnitude faster than previous LIE. As such, the LIECE method can be effectively used to screen large libraries of compounds docked by automatic computational tools and has been successfully applied in virtual screening campaigns against important drug targets although the parameters developed for one target are generally not transferable. The strong dependence of the binding affinities on the electrostatic component of the protein-ligand interaction energy motivated the development of a LIECE variant (QMLIECE) in which the ligand-surrounding interactions are evaluated by a semiempirical quantum mechanical model to include e.g. polarization effects. Interestingly, the QM variant was shown to be superior when dealing with formally charged compounds as the peptidic inhibitors of the West Nile serine protease.

### 2.3 Empirical Approaches

Although the endpoint approaches significantly reduce the computing time, they are still too expensive to estimate binding affinities for a large library or a database of compounds. Therefore, even more approximated approaches are required for structure-based virtual high-throughput screening, where hundreds of thousands of compounds for a specific target must be evaluated and ranked. To speed up the calculations, the so-called empirical approaches simplify the description of the binding reaction one step further and focus exclusively on the bound state (Figure 1, bottom).

Among them, molecular docking is with no doubt the most popular approach. In this case, the binding affinity is usually estimated in the context of a rigid conformation of the receptor with no sampling procedure to account for its dynamics. The flexibility of the ligand is included through a systematic or stochastic search typically Monte Carlo or a genetic algorithm, and its fitness is quantified by a crude scoring function, which is used both for ranking the binding modes and prioritizing compounds extracted from the library. In most docking protocols, the atomistic details of the protein are replaced by a grid representation centered on the binding site, where each point stores the interaction energy of an atomic "probe" with the rest of the receptor. Also, solvent effects are usually neglected or efficiently accounted for by continuum solvation models. In general, three main classes of scoring functions are used in protein-ligand binding: empirical, force field-based and knowledge-based.

In the following, we will focus on the first two, which despite their crude nature can still be interpreted in the framework of statistical mechanics.

#### 2.3.1 Empirical Scoring Function

A first approach to efficiently score a large number of docking poses is based on the evaluation of the binding free energy by a weighted sum of empirical descriptors as

$$
\Delta \mu_k = \sum_i W_i \Delta \mu_i
$$

with $\Delta \mu_i$ corresponding to independent contributions selected based on chemical intuition, e.g. the electrostatic and van der Waals components of the protein-ligand interaction energy, the number of H-bonding donors and acceptors, the number of rotatable bonds, etc., each accounting for a critical interaction in protein-ligand binding. The weight of the contributions, i.e. the coefficients in Eq. 26 ($W_i$), are empirically determined by fitting on a training set of experimental binding affinities typically using multivariate linear regressions. The simplicity and flexibility of Eq. 26 grants for the required computational efficiency at the hit-identification stage. However, the accuracy of the predic-
Minireview

functions is strongly dependent on the quality of the training set (both in size and composition), which makes the transferability of the model challenging. One of the earliest model for docking was introduced by Böhm
\[ \Delta G_0 = \Delta G_0 + \Delta G_{hb} \sum_{i \neq j} f(\Delta R, \Delta \alpha) + \]
\[ \Delta G_{io} \sum_{i \neq j} f(\Delta R, \Delta \alpha) + \]
\[ \Delta G_{lip} \sum_{lip cont.} A_{lip} + \]
\[ \Delta G_{aro} \sum_{aro cont.} f(\Delta R) + \Delta G_{rot} \times N_{rot}. \]

In this equation, the polar contribution is accounted for by explicit H-bonding (\( \Delta G_{hb} \)) and ionic interaction (\( \Delta G_{ih} \)) terms, which are both distance and angle dependent to penalize deviations from optimum geometries. The apolar contribution is accounted for by lipophilic contacts (\( \Delta G_{lip} \)) and aromatic interactions (\( \Delta G_{aro} \)), with the former being dependent on the contact surface (\( A_{lip} \)) and the latter on the distance. Finally, the flexibility of the ligand is indirectly included by counting the number of rotatable bonds (\( N_{rot} \)), which approximates the entropy cost to confine the ligand in the bound state.

Another example of fast and simple empirical scoring is provided by Fresno, a model introduced by Rognan et al.

\[ \Delta \mu_f = K + \alpha(\text{HB}) + \beta(\text{LIPO}) + \gamma(\text{ROT}) + \delta(\text{BP}) + \epsilon(\text{DESOLV}). \]

This empirical scoring function includes five terms corresponding to hydrogen bonds, lipophilic, rotational, buried-polar and ligand desolvation contributions. The first three terms are common to the Böhm model and are similarly evaluated. The buried-polar term (BP) accounts for unfavorable interactions or “bumps”, which result from contacts between polar and nonpolar groups in the binding site. The last term accounts for the desolvation energy of the ligand in the unbound state (DESOLV), which is efficiently evaluated by a Poisson-Boltzmann calculation. Despite the empirical nature of Eq. 28, this scoring function can be connected with the theoretical framework above, particularly with the LIE approach. In fact, the hydrogen bond, lipophilic and buried-polar terms are clearly related to the electrostatic and van der Waals interaction energies of the ligand in the bound state, whereas the rotational and the desolvation contributions are related to the unbound state of the ligand and account for the rotational entropy loss and the ligand desolvation upon binding. However, the arbitrary selection of the energy components in such empirical schemes makes it difficult to rationalize the connection with first-principle statistical mechanics.

2.3.2 Force Field Scoring Function

Another class of scoring functions aims at the evaluation of the binding affinity through the quantification of the physical protein-ligand interactions as given by molecular mechanics, where the nonbonding energy is typically evaluated as a sum of pairwise electrostatic and van der Waals atomic contributions. This two ingredients provide the minimalistic score used by the program Dock (v. 4.0)\[^{[64]}\]

\[ \Delta \mu_f = \sum_{i} \sum_{j} \left( \frac{A_{i}}{r_{ij}} - \frac{B_{i}}{r_{ij}^{2}} + 332.0 \frac{q_{i} q_{j}}{e(r_{ij})r_{ij}^{3}} \right) \]

with \( r_{ij} \) being the distance between the \( j \)-th atom of the ligand and the \( i \)-th atom of the protein, \( A_{i} \) and \( B_{i} \) the atomic van der Waals parameters, and \( q_{i} \) and \( q_{j} \) their partial charges. These empirical parameters are usually obtained from popular biomolecular force fields such as AMBER or CHARMM. Calculations of the nonbonding interactions in the bound state are typically performed in vacuum, sometimes using a distance-dependent dielectric constant \( \varepsilon(r_{ij}) \) to account for solvation effects. Alternatively, more realistic descriptions of the solvent are achieved through the use of more rigorous implicit solvent models as PB/SA or GB/SA, which require extra computation. Overall, the great advantage of these methods is that the binding affinity can be evaluated, in principle, for any non-covalent protein-ligand association as the force-field parameters are developed to be transferable. Interestingly, Eq. 29 can be derived from the fundamental equation of LIE (Eq. 25) introducing two additional assumptions: i. the binding affinity can be accessed from a single structure of the protein-ligand complex, which turns the ensemble averages in Eq. 25 into single-point energy evaluations; and ii. the desolvation of the ligand upon binding is negligible, i.e. the ligand/surrounding interaction energy in the unbound state can be set to zero. Thus, Eq. 25 yields

\[ \Delta \mu_f = \left[ (U_{f}^{\text{HSE}})_{\text{site}} - 0 \right] + \left[ (U_{f}^{\text{des}})_{\text{site}} - 0 \right] \]

\[ = \sum_{i} \sum_{j} \left( \frac{A_{i}}{r_{ij}} - \frac{B_{i}}{r_{ij}^{2}} + 332.0 \frac{q_{i} q_{j}}{e(r_{ij})r_{ij}^{3}} \right) \]

Finally, to effectively account for contributions from the unbound state, i.e. the strain energy of the ligand or the total entropy change upon binding, hybrid implementations of Eq. 29 have been developed. A prototypical example is implemented in AutoDock (v. 4.2)\[^{[65]}\].
The situation is different for the semi-rigorous or end-points approaches where MM/PBSA (Eq. 20) belongs to the first class and LIE (Eq. 25) to the second class; see Figure 4. In both cases, the binding constant is accessed by solving a thermodynamic cycle that involves molecular transfer to the gas phase. This strategy effectively transforms the calculation of the standard free energy of binding into a difference between (approximate) solvation free energies, which can be evaluated with much less computation. By replacing the explicit representation of the binding path with approximate solvation free energy estimates based on continuum models or the linear response theory, these methods alleviate the computational burden quite significantly, extending their scope to the hit-to-lead stage, where thousands of compounds must be evaluated and ranked. Of course, the quality of the predictions critically depends on the accuracy of the solvation free energy calculations, which motivates further effort on the development of more accurate implicit solvent models. Also, the striking similarity with the strategy implemented in the (rigorous) alchemical route suggests that the use of restraints to control the configurational freedom of the ligand, particularly in the unbound state, could be beneficial to accelerate numerical convergence.

Analysis of the fast empirical approaches to protein-ligand (including some of the most popular scoring functions for docking) shows that these methods break down the computational cost by focusing exclusively on the bound state, i.e. forgetting about the protein or the ligand in solution, and therefore belong to the second class; see Figure 4. Because the average output rate is of one free-energy determination per second, these simplified approaches are suitable for screening millions of compounds and find widespread use at the hit identification stage. Nonetheless, the significant speed-up is achieved by introducing a series of theoretically unjustified approximations, which result in sizeable systematic errors that make the predictions often unreliable and/or highly system-dependent. Comparison of the force-field (FF) scoring functions with less-approximated approaches in the same class, i.e. LIE, demonstrates that the strongest approximations in the former are related to both neglecting of entropic effects, which results from a rigid-body treatment of the receptor, and the deliberate exclusion of contributions from the unbound state, i.e. the strain energy upon binding and the ligand desolvation. In light of this, the straightforward implementation of a statistical mechanics treatment of the vibrational entropy e.g. by normal-mode analysis and/or the explicit inclusion of ligand desolvation by fast implicit solvent models are expected to improve the quality of the docking predictions. Interestingly, these contributions are already included in some empirical scoring (ES) functions.63

Finally, our classification of methods (Figure 4) highlights different sources of systematic error in the evaluation of the protein-ligand binding affinity. Sampling of the configurational space accessible to the system in the bound and the unbound state is one of them, which explains, for in-

\[ \Delta \mu _B = W_{\text{vdew}} \sum_{\text{prot}} \sum_{\text{lig}} \left( \frac{A_{ij}}{T_i} - \frac{B_{ij}}{T_j} \right) + \]

\[ W_{\text{elec}} \sum_{\text{prot}} \sum_{\text{lig}} \left( \frac{q_i q_j}{r_{ij}} \right) + \]

\[ W_{\text{bond}} \sum_{\text{prot}} \sum_{\text{lig}} \left( E(t) \frac{C_{ij}}{T_i} - \frac{D_{ij}}{T_j} \right) + \]

\[ W_{\text{desolv}} \sum_{\text{prot}} \sum_{\text{lig}} \left( S_{i \text{lig}} V_j + S_{j \text{prot}} V_i \right) \exp \left( -r_{ij}^2 / 2 \sigma^2 \right) \]

where \( W_i \) are weighting constants calibrated on experimental binding data. In this case, the van der Waals and the electrostatic contributions, which are typical of a force field, are complemented by a specialized 10/12 Lennard-Jones potential for directional hydrogen bonding and a desolvation term that is related to the excluded volume for the ligand in the bound state. In general, the goal of a hybrid scheme is to introduce critical free energy contributions to binding (e.g. ligand desolvation) which are missing in the force field representation without compromising the computational efficiency.

3 Discussion

The calculation of the protein-ligand binding affinity is a fundamental problem that poses an outstanding theoretical and computational challenge. To this aim, several computational approaches have been developed over years, which tackle the problem at various degrees of approximation. The statistical mechanics interpretation presented in this mini-review suggests that there are two general approaches to the standard free energy of binding. One approach goes through the (direct) evaluation of the absolute chemical potentials for all components of the binding reaction (i.e. the ligand, the protein and the complex). The other one treats the binding reaction as a partition equilibrium of the ligand between the bound and the unbound states, which assumes that most of the protein contributions to the chemical potential difference effectively cancel out. To the best of our knowledge, all rigorous approaches to the binding constant such as FEP (Eq. 15) or PMF (Eq. 16) fall in the second class; see Figure 4. Surprisingly and perhaps due to the intrinsic challenge posed by the accurate evaluation of the solvation free energy of the protein with and without the ligand, there exists no rigorous approach belonging to the former class. Because the evaluation of the free energy of binding involves the detailed analysis of a large number of intermediate states along the reaction path, these methods are computationally very intensive and may be useful to rank only a small number of compounds, typically less than a hundred, at the lead-optimization stage.
stance, the observed increase in accuracy on moving from the fast empirical methods to the end-points strategies. An accurate treatment of the solvent is another important aspect, which is well exemplified by the comparison of the end-points strategies with the rigorous methods. In this case, when the computationally intensive evaluation of the solvation free energy in the latter is replaced by continuum models (MM/PBSA) or an ensemble average of the ligand/surroundings interactions (LIE) sizable inaccuracies may be introduced. Last but not least, a force-field representation of interactions, which neglects polarization effects, is another source of systematic error that affects the quality of the predictions independently of sampling. Simplified quantum-mechanical treatments as those introduced in QM-MM/PBSA\(^{69}\) and QMLIECE\(^{59}\) represent pioneering attempts to quantify this type of errors. The development of strategies in which errors due to undersampling or a continuum treatment of the solvent are roughly equal in size to those introduced by the force-field parameterization are key for the development of optimal computational approaches to protein-ligand binding.

In conclusion, our interpretation of apparently unrelated computational approaches to protein-ligand binding in the common framework of statistical mechanics allows to pinpoint the approximations that are introduced to speed up the calculations, which is useful to rationalize their impact on the accuracy of the binding affinity predictions. Our comparative analysis already highlights possible improvements to well established semi-rigorous and empirical approaches and will help in the development of variants with an optimum balance between accuracy and efficiency at each stage of the drug-discovery pipeline.

**Conflict of Interest**

None declared.

**Acknowledgments**

Nicolas Muzet (Sanofi) and Michael Schaefer (Novartis Pharma) are gratefully acknowledged for useful discussions and a critical reading of the manuscript. JMA received financial support from the Fondation pour la Recherche Médicale. This work was funded in part by the Agence National de la Recherche through the LabEx project Chemistry of Complex Systems (ANR-10-LABX-0026).

**References**

[1] J.-P. Changeux, S. J. Edelstein, *Science* 2005, 308, 1424–1428.
[2] F. Feixas, S. Lindert, W. Sinko, J. A. McCammon, *Biophysical chemistry* 2014, 186, 31–45.
[3] C. Chipot, *Wiley Interdisciplinary Reviews: Computational Molecular Science* 2014, 4, 71–89.
[4] P. Kollman, *Chemical reviews* 1993, 93, 2395–2417.
[5] S. M. Paul, D. S. Mytelka, C. T. Dunwiddie, C. C. Persinger, B. H. Munos, S. R. Lindborg, A. L. Schacht, *Nature reviews Drug discovery* 2010, 9, 203–214.
[6] S. Morgan, P. Grootendorst, J. Lexchin, C. Cunningham, D. Greyson, *Health Policy* 2011, 100, 4–17.
[7] N. Homeyer, F. Stoll, A. Hillisch, H. Gohlke, Journal of chemical theory and computation 2014, 10, 3331 – 3344.
[8] W. L. Jorgensen, J. K. Buckner, S. Boudon, J. Tirado-Rives, The Journal of chemical physics 1988, 89, 3742 – 3746.
[9] J. Hermans, L. Wang, Journal of the American Chemical Society 1999, 121, 12970 – 12975.
[10] S. Boresch, F. Tettiger, M. Leitgeb, M. Karplus, The Journal of Physical Chemistry B 2003, 107, 9535 – 9551.
[11] M. K. Gilson, J. A. Given, B. L. Bush, J. A. McCammon, Biophysical journal 1997, 72, 1047.
[12] W. L. Jorgensen, Accounts of Chemical Research 1989, 22, 184 – 189.
[13] H.-J. Woo, B. Roux, Proceedings of the National Academy of Sciences of the United States of America 2005, 102, 6825 – 6830.
[14] M. S. Lee, M. A. Olson, Biophysical journal 2006, 90, 864 – 877.
[15] G. Rastelli, A. D. Rio, G. Degliesposti, M. Sgobba, Journal of computational chemistry 2010, 31, 797 – 810.
[16] M. Ikeguchi, J. Ueno, M. Sato, A. Kidera, Physical review letters 2005, 94, 078102.
[17] P. A. Kollman, I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, et al., Accounts of chemical research 2000, 33, 889 – 897.
[18] J. Srivivasan, T. E. Cheatham, P. Cieplak, P. A. Kollman, D. A. Case, Journal of the American Chemical Society 1998, 120, 9401 – 9409.
[19] J. Åqvist, C. Medina, J.-E. Samuelsson, Protein engineering 1994, 7, 385 – 391.
[20] H. Gutiérrez-de Terán, J. Åqvist, Computational Drug Discovery and Design 2012, 305 – 323.
[21] D. B. Kitchen, H. Decornez, J. R. Furr, J. Bajorath, Nature reviews Drug discovery 2004, 3, 935 – 949.
[22] H.-Y. Liu, I. D. Kuntz, X. Zou, The Journal of Physical Chemistry B 2004, 108, 5453 – 5462.
[23] H. Gohlke, M. Hendlich, G. Klebe, Journal of molecular biology 2000, 295, 337 – 356.
[24] Q. U. Ain, A. Aleksandrova, F. D. Roessler, P. J. Ballester, Wiley Interdisciplinary Reviews: Computational Molecular Science 2015, 5, 405 – 424.
[25] Y. Deng, B. Roux, The Journal of Physical Chemistry B 2009, 113, 2234 – 2246.
[26] S. Genheden, U. Ryde, Proteins: Structure, Function, and Bioinformatics 2012, 80, 1326 – 1342.
[27] D. McQuarrie, Statistical Mechanics, New York: Harper and Row, 1976.
[28] T. L. Hill, Cooperativity theory in biochemistry: steady-state and equilibrium systems, Springer Verlag, 1985.
[29] J. Esque, M. Cecchini, The Journal of Physical Chemistry B 2015, 119, 5194 – 5207.
[30] T. Simonson, G. Archontis, M. Karplus, Accounts of chemical research 2002, 35, 430 – 437.
[31] A. Pohorille, C. Jarzynski, C. Chipot, The Journal of Physical Chemistry B 2010, 114, 10235 – 10253.
[32] J. C. Gumbart, B. Roux, C. Chipot, Journal of chemical theory and computation 2012, 9, 794 – 802.
[33] J. Wang, Y. Deng, B. Roux, Biophysical journal 2006, 91, 2798 – 2814.
[34] A. Y. Lau, B. Roux, Nature structural & molecular biology 2011, 18, 283 – 287.
[35] J. C. Gumbart, B. Roux, C. Chipot, Journal of chemical theory and computation 2013, 9, 3789 – 3798.
[36] L. Wang, Y. Wu, Y. Deng, B. Kim, L. Pierce, G. Krilov, D. Lupyian, S. Robinson, M. K. Dahlgren, J. Greenwood, et al., Journal of the American Chemical Society 2015, 137, 2695 – 2703.
[37] B. Brooks, M. Karplus, Proceedings of the National Academy of Sciences 1983, 80, 6571 – 6575.
[38] M. Cecchini, Journal of chemical theory and computation 2015, 11, 4011 – 4022.
[39] R. M. Levy, M. Karplus, J. Kushick, D. Perahia, Macromolecules 1984, 17, 1370 – 1374.
[40] M. Karplus, J. Kushick, Macromolecules 1981, 14, 325 – 332.
[41] H. Gohlke, D. A. Case, Journal of computational chemistry 2004, 25, 238 – 250.
[42] T. Hou, J. Wang, Y. Li, W. Wang, Journal of chemical information and modeling 2010, 50, 69 – 82.
[43] T. Hou, J. Wang, Y. Li, W. Wang, Journal of computational chemistry 2011, 32, 866.
[44] H. Sun, Y. Li, S. Tian, L. Xu, T. Hou, Physical Chemistry Chemical Physics 2014, 16, 16719 – 16729.
[45] S. Genheden, T. Luchko, S. Gusarov, A. Kovalenko, U. Ryde, The Journal of Physical Chemistry B 2010, 114, 8505 – 8516.
[46] L. Xu, H. Sun, Y. Li, J. Wang, T. Hou, The Journal of Physical Chemistry B 2013, 117, 8408 – 8421.
[47] C. Gao, M.-S. Park, H. A. Stern, Biophysical journal 2010, 98, 901 – 910.
[48] S. Genheden, O. Kuhn, P. Mikulsik, D. Hoffmann, U. Ryde, Journal of chemical information and modeling 2012, 52, 2079 – 2088.
[49] M. Kaukonen, P. Soderhjelm, J. Heimdal, U. Ryde, The Journal of Physical Chemistry B 2008, 112, 12537 – 12548.
[50] F. S. Lee, Z.-T. Chu, M. B. Bolger, A. Warshel, Protein Engineering 1992, 5, 215 – 228.
[51] R. Marcus, Annual Review of Physical Chemistry 1964, 15, 155 – 196.
[52] A. Ben-Naim, Y. Marcus, The Journal of chemical physics 1984, 81, 2016 – 2027.
[53] M. Almlöf, J. Carlsson, J. Åqvist, Journal of chemical theory and computation 2007, 3, 2162 – 2175.
[54] J. Åqvist, J. Marelius, Combinatorial chemistry & high through-put screening 2001, 4, 613 – 626.
[55] J. Åqvist, T. Hansson, The Journal of Physical Chemistry 1996, 100, 9512 – 9521.
[56] D. Huang, A. Caffisch, Journal of medicinal chemistry 2004, 47, 5791 – 5797.
[57] D. Huang, A. Caffisch, Journal of Molecular Recognition 2010, 23, 183 – 193.
[58] T. Zhou, D. Huang, A. Caffisch, Journal of medicinal chemistry 2008, 51, 4280 – 4288.
[59] T. Zhou, D. Huang, A. Caffisch, Current topics in medicinal chemistry 2010, 10, 33 – 45.
[60] A. Breda, L. A. Basso, D. S. Santos, J. De Azevedo, F. Walter, Current Computer-Aided Drug Design 2008, 4, 265 – 272.
[61] S.-Y. Huang, S. Z. Griner, X. Zou, Physical Chemistry Chemical Physics 2010, 12, 12899 – 12908.
[62] H. J. Böh, Journal of computer-aided molecular design 1998, 12, 309 – 309.
[63] D. Rognan, S. L. Laurermeier, A. Holm, S. Buss, V. Tschinke, Journal of medicinal chemistry 1999, 42, 4650 – 4658.
[64] E. C. Meng, B. K. Shoichet, I. D. Kuntz, Journal of computational chemistry 1992, 13, 505 – 524.
[65] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, Journal of computational chemistry 2009, 30, 2785 – 2791.