Part I
Binding Thermodynamics

Protein-Ligand Interactions, First Edition. Edited by Holger Gohlke.

1
Statistical Thermodynamics of Binding and Molecular Recognition Models

Kim A. Sharp

1.1
Introductory Remarks

Equilibrium binding or association of two molecules to form a bimolecular complex, $A + B \Leftrightarrow AB$, is a thermodynamic event. This chapter will cover some of the fundamental thermodynamics and statistical mechanics aspects of this event. The aim is to introduce general principles and broad theoretical approaches to the calculation of binding constants, while later chapters will provide examples. Only the noncovalent, bimolecular association under ambient pressure conditions will be considered. However, extension to higher order association involves no additional principles, and extension to high pressure by inclusion of the appropriate pressure–volume work term is straightforward. In terms of the binding reaction above, the association and dissociation constants are defined as $K = [AB]/[A][B]$ and $K_D = [A][B]/[AB]$ respectively, where [] indicates concentration. Either K or K_D is the primary experimental observable measured in binding reactions. K_D is sometimes obtained indirectly by inhibition of binding of a different ligand as a K_i. From a thermodynamic perspective, the information content from K, K_D, and K_i is the same.

1.2
The Binding Constant and Free Energy

To connect the experimental observable K to thermodynamics, one often finds in the literature the relationship

$$\Delta G_{bind} = -kT \ln K, \tag{1.1}$$

where k is the Boltzmann constant, T is the absolute temperature, and ΔG_{bind} is the "absolute" or "standard" binding free energy. Several comments are given to avoid misuse of this expression. First, one cannot properly take the logarithm of a quantity with units such as K, so Eq. (1.1) is implicitly

$$\Delta G_{bind} = -kT \ln \frac{K}{V_{ref}}, \tag{1.2}$$

Protein-Ligand Interactions, First Edition. Edited by Holger Gohlke.

where V_{ref} is the reference volume in units consistent with the units of concentration in K, that is, 1 l/mol or about 1660 Å^3/molecule for molarity units. The choice of V_{ref} is often referred to as the "standard state" problem. Equivalently, one says that ΔG_{bind} is the free energy change when reactants A and B and the product AB are all at the reference concentration. Second, although the units of concentration used in K are almost always moles/liter, this is entirely a *convention*, so the actual numerical value for ΔG_{bind} obtained from Eq. (1.2) is arbitrary. Put another way, any method for calculating the free energy of binding must explicitly account for a particular choice of V_{ref} before it can meaningfully be compared with experimental values of ΔG_{bind} obtained using Eq. (1.2). Furthermore, ligand efficiency-type measures, such as $\Delta G_{bind}/n$ where n is the number of heavy atoms in a ligand or the molecular weight of a ligand [1], can change radically with (arbitrary) choice of concentration units. Of course, differences in ΔG_{bind} can be sensibly compared provided the same reference state concentration is used. Finally, in Eq. (1.2), the free energy actually depends on the ratio of activities of reactants and products, not on concentrations. For neutral ligands and molecules of low charge density at less than micromolar concentrations, the activity and concentration are nearly equal and little error is introduced. However, this is not true for high charge density molecules such as nucleic acids and many of the ligands and proteins that bind to nucleic acids. Here, the activity coefficient can be substantially different from unity even at infinitely low concentration. Indeed, much of the salt dependence of ligand–DNA binding can be treated as an activity coefficient effect [2–4]. The issue of standard state concentrations, the formal relationship between the binding constant and the free energy, and the effect of activity coefficients are all treatable by a consistent statistical mechanical treatment of binding, as described in Section 1.3.

1.3
A Statistical Mechanical Treatment of Binding

Derivation of a general expression for the binding constant follows closely the approach of Luo and Sharp [5], although somewhat different treatments using chemical potentials, which provide the same final result, are given elsewhere [6–8]. It is a statistical mechanical principle that any equilibrium observable can be obtained as an ensemble, or Boltzmann weighted average, of the appropriate quantity. Here, the binding constant $K = [AB]/[A][B]$ is the required observable. Consider a single molecule each of A and B in some volume V (Figure 1.1) and for convenience define a coordinate system centered on B (the target) in a fixed orientation. Over time, the ligand (A) will explore different positions and orientations (poses) relative to B, where $\mathbf{r}$ and Ω represent the three position and three orientation coordinates of A with respect to B. Now A and B interact with each other with an energy that depends not only on their relative position ($\mathbf{r}$, Ω) but in general also on the conformations of A, B, and the surrounding solvent. If n_a, n_b, n_s are the number of atoms in A, B, and solvent, then the energy is a function of $3n_a + 3n_b + 3n_s - 6$ coordinates. In principle, one could keep all these degrees of freedom explicit. From a

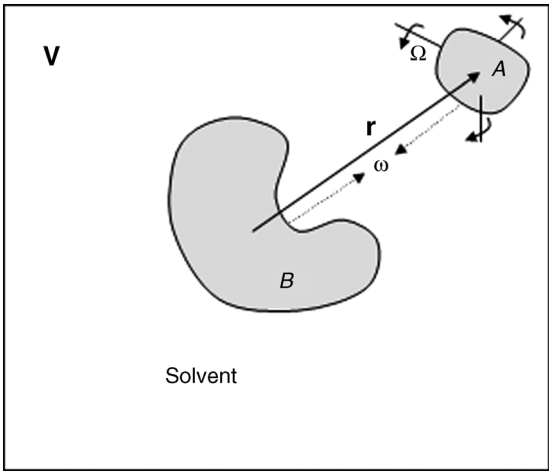

Figure 1.1 Schematic illustration of two molecules A and B interacting through solvent with a potential of mean force ω as a function of their relative position **r** and orientation Ω.

practical standpoint, this would be a complicated and expensive function to evaluate. However, one may integrate over the solvent coordinates and the $(3n_a - 6) + (3n_b - 6)$ internal coordinates so that the interaction between A and B for a given (**r**, Ω) is described by an interaction potential of mean force (pmf) $\omega(\mathbf{r}, \Omega)$. If one defines the pmf between A and B at infinite separation in their equilibrium conformations to be 0, then $\omega(\mathbf{r}, \Omega)$ is the thermodynamic work of bringing A and B from far apart to some mutual pose (**r**, Ω), accounting for both solvent effects and internal degrees of freedom of A and B. A will sample each pose (**r**, Ω) with a probability given by the Boltzmann factor of the pmf:

$$p(\mathbf{r}, \Omega) \propto e^{-\beta\omega(\mathbf{r},\Omega)}, \tag{1.3}$$

where $\beta = 1/kT$. Indeed, one may consider the pmf to be defined by this equation. The binding constant will then be given by the fraction of time A is in the bound state, f_{ab}, relative to that in the free state, f_f:

$$K = \frac{[\mathrm{AB}]}{[\mathrm{A}][\mathrm{B}]} = \frac{f_{ab}/V}{(f_f/V)(f_f/V)} \underset{V\to\infty}{\longrightarrow} f_{ab}V, \tag{1.4}$$

where in the dilute limit $f_f \to 1$. It is convenient to introduce a function $H(\mathbf{r}, \Omega)$ that takes a value of 1 for poses where A is bound and a value of 0 when it is free. Then, the fraction of the time A is bound is given by the ensemble average of H:

$$f_{ab} = \int d\mathbf{r}\, d\Omega H(\mathbf{r}, \Omega) e^{-\beta\omega(\mathbf{r},\Omega)} / \int d\mathbf{r}\, d\Omega e^{-\beta\omega(\mathbf{r},\Omega)}. \tag{1.5}$$

The integrals are taken over all orientations and over the entire volume of the solution, so the denominator gives $8\pi^2 V$. Substituting into Eq. (1.4), the final expression for the association constant is

$$K = \frac{1}{8\pi^2}\int \mathrm{d}\mathbf{r}\,\mathrm{d}\Omega H(\mathbf{r},\Omega)e^{-\beta\omega(\mathbf{r},\Omega)}. \tag{1.6}$$

One may then convert this to an "absolute" binding free energy using Eq. (1.2):

$$\Delta G_{\mathrm{bind}} = kT\ln(8\pi^2 V_{\mathrm{ref}}) - kT\ln\int \mathrm{d}\mathbf{r}\,\mathrm{d}\Omega H(\mathbf{r},\Omega)e^{-\beta\omega(\mathbf{r},\Omega)}. \tag{1.7}$$

- Equation 1.6 is a general and exact expression for the association constant. The integral depends explicitly on just six variables describing the pose of A with respect to B. The other degrees of freedom are included implicitly, but exactly through the thermodynamic quantity $\omega(\mathbf{r}, \Omega)$, the potential of mean force.
- The different treatment of coordinates for translation/orientation versus the others is a formal one: Any subset of coordinates may in principle be kept explicit, with the appropriate pmf being used for the rest. For example, one may keep the internal coordinates of A and B explicit, making the solvent coordinates implicit. The choice here is designed to highlight the translation/rotation contribution to binding that has been widely discussed, with some disagreement, in the literature [5, 6, 9–13]. It also reflects the practical fact that in many docking and screening applications, a particular pose is generated explicitly, that is, $(\mathbf{r}, \Omega)$ is specified, and then the pose is "scored" in some way. The pmf also provides a natural way to introduce approximations necessary for any practical calculation of K in biological systems, for example, in the treatment of solvent.
- The integral has the correct units of volume, with the length scale for the translation coordinates being determined by the units of concentration used in K. The first term in Eq. (1.7) is the contribution of the rotation/translation (R/T) entropy in the unbound state, which depends on the reference concentration. The integral term in Eq. (1.6) is the Boltzmann phase volume of the bound state.
- Through $H(\mathbf{r}, \Omega)$, there is explicit consideration of what constitutes the bound complex, in terms of the relative position and orientation of A with respect to B. For example, if B has more than one binding site for A, this would be taken into account in the specification of where $H = 1$.
- Either Cartesian coordinates or the bond length, bond angle, and dihedral angle coordinates may be used. The trend now is toward the latter, as they lend themselves more naturally to the analysis of different internal motions of the molecules and their contribution to binding.

The meaning of Eq. (1.6) is illustrated by two simple examples.

1.3.1 Binding in a Square Well Potential

Let the pmf be approximated by a simple, three-dimensional square well potential of depth ε and width b in each of the x, y, z directions and the bound complex be the region in the well only. From Eq. (1.6), the association constant is

$$K = b^3 e^{\beta\varepsilon} \tag{1.8}$$

and Eq. (1.2) yields

$$\Delta G_{\text{bind}} = -\varepsilon + kT \ln(V_{\text{ref}}/b^3). \tag{1.9}$$

The first term, the well depth, makes a direct, linear contribution to the binding free energy. The second term is positive and comes from the restriction of the ligand to the square well. It is the translation entropy penalty for binding, and it depends on the ratio of the volumes available to the ligand in the free state at say 1 M (the entire volume V_{ref}) versus that in the bound state. In this simple example, there is no rotational entropy penalty because in the bound state the ligand can rotate freely in $8\pi^2$ of orientation phase volume, just as in the free state. However, restriction in rotation in the bound state will add another positive term to ΔG_{bind}, the rotation entropy penalty, with a similar form: $kT\ln(8\pi^2/V_\Omega)$, where $V_\Omega < 8\pi^2$ is the orientation phase volume in the bound state. We can see even from this simple example that for any meaningful degree of binding, the translational and rotational phase volumes available to a ligand in the bound state must be less than V_{ref} and $8\pi^2$, respectively, so there is *always* a R/T entropy penalty to be overcome for binding to occur. The question is how much is it in specific cases. A related point is that even though the depth of the well may be known, for example, from some calculation (in the parlance of the field, from a single point energy determination), this cannot be directly compared with ΔG_{bind} because the second term is not included. The numerical value of the binding free energy depends on the reference concentration, which is nowhere in the single point calculation. One way or another, the residual R/T entropy of A in the bound state must be accounted for.

1.3.2 Binding in a Harmonic Potential

If one is starting from a known complex structure derived from, for example, X-ray, NMR, or molecular mechanics minimization, one is presumably close to the minimum energy (pmf) configuration. The pmf in this region may be close to harmonic or at least expandable in a Taylor expansion, which to second order is harmonic. It is, therefore, instructive to consider binding in a harmonic potential, although this is a simplified model of the real situation. Let the potential well be a three-dimensional harmonic potential of the form

$$\omega(r) = \varepsilon\left(\left(\frac{r}{b}\right)^2 - 1\right)(r < b), \quad \omega(r) = 0(r \geq b), \tag{1.10}$$

where ε is the depth of the well at the minimum, r is the radial distance from the minimum, and b defines the width so that for $r \geq b$, $\omega = 0$ (Figure 1.2). Again, the bound complex is defined to be the region in the well only. Substituting Eq. (1.10) into Eq. (1.6) and integrating, the association constant for this truncated harmonic potential is

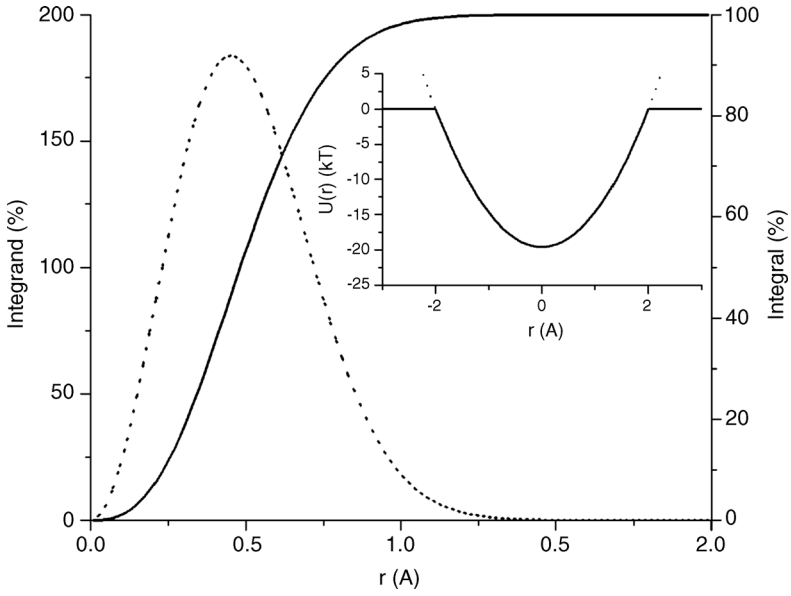

Figure 1.2 Contributions to the binding-phase integral. *Dotted line*: Value of the integrand of Eq. (1.6) at *r*. *Solid line*: Value of the resulting integral from 0 to *r*. Both are expressed as a percentage of the total association constant. Contributions were calculated for a truncated three-dimensional harmonic well potential, half-width 2 Å, and depth 19.6*kT* (inset) that has $K_d = 10\,\mu M$.

$$K = b^3 e^{\beta\varepsilon}\left(4\pi\sum_{n=0}^{\infty}\frac{-1^n(\beta\varepsilon)^n}{(2n+3)n!}\right) \approx b^3 e^{\beta\varepsilon}\left(\frac{\pi}{\beta\varepsilon}\right)^{3/2}. \tag{1.11}$$

The approximate equality comes from using an untruncated harmonic potential (i.e., the potential goes to infinity as the complex is dissociated), which for this case gives a binding free energy of

$$\Delta G_{\text{bind}} = -\varepsilon + kT\ln(V_{\text{ref}}/b^3) - 3/2kT\ln(\pi/\beta\varepsilon). \tag{1.12}$$

Comparing the square well and harmonic potential models, one sees that the "depth" and "volume" factors, $e^{\beta\varepsilon}$ and b^3, contribute in the same way to the binding constant, the difference being a "well shape" factor. We see from the form of the expression for the association constant that the lower the pmf, the more the contribution to the integral by that region, so most of the contribution to binding should come from the near minimum energy configuration. This is illustrated in Figure 1.2, using a well half-width of 2 Å and a depth of 19.6*kT*. These parameters are chosen to give a moderate affinity of 10 μM – typical of the compounds studied by virtual screening and docking calculations in early lead identification – with a reasonable degree of motion in the binding pocket. It can be seen that almost all the contribution to the binding constant comes significantly before reaching the well boundary. Thus, the problem of giving the exact definition of the complex in

($\mathbf{r}$, Ω) space goes away if the binding well is more than a few kT deep, as it usually is. Under these conditions, the untruncated harmonic well approximation in Eq. (1.11), with its simpler form, is nearly exact. Note that most of the contribution to the binding comes from conformations significantly *away* from the minimum, here at distances of 0.25–0.75 Å. This is property of the three-dimensional nature of the binding well, and it occurs for the same reason that the Maxwell–Boltzmann distribution of velocities is peaked at $3/2\ kT$, not at 0: The amount of phase volume right at the minimum is small, but it increases as r^2 as we move away. When we add in the three degrees of rotational freedom in multiatom ligand binding, we have a six-dimensional well and this effect will be even greater. The relative contribution from the minimum will drop even more. The point to be drawn from these simple models is that accurate calculation of a binding affinity cannot just rely on the estimation of the well depth or use the minimum energy configuration alone. Some sampling of configurations around the minimum, either explicit or implicit, is needed. In docking studies, typically many poses are generated, but the scoring is based only on the best pose. There is no reason why all the poses could not be used to build up some estimation of the Boltzmann phase volume or at least rank equally good $E_{\min}$ candidates accounting for the different number of nearby poses.

1.4 Strategies for Calculating Binding Free Energies

1.4.1 Direct Association Simulations

Given the fundamental expression for the association constant, Eq. (1.7), the most direct approach is to calculate the thermodynamic work of bringing the ligand into the binding site. Starting from the unbound state, one applies a series of harmonic positional and orientational restraining potentials that gradually maneuver the ligand into the binding pocket [8, 14]. The probability distributions of the ligand position/orientation are obtained from molecular dynamics (MD) for each restraining potential, corrected for the effect of the restraint, and spliced together to provide the complete probability distribution, from which the change in pmf is obtained as $-kT\ln\left(p(\mathbf{r},\ \Omega)/p^{\text{unbound}}(\mathbf{r},\ \Omega)\right)$. In this type of simulation, the ligand in the bound state feels the harmonic restraint plus the potential from its binding partner. In the unbound state, the ligand feels just the harmonic restraining potential, for which one knows the reference state free energy (see Eq. (1.12)). Adding this to the calculated change in pmf, one obtains the absolute binding free energy. This approach is computationally intensive as one needs to simulate not only the bound and free states but also many intermediate states. If an indirect route into the binding site requires additional incremental restraining potential steps, this will increase the computation. For this reason, atoms far from the binding site are usually frozen in the simulation.

Other approaches to the calculation of absolute binding free energies rely on splitting up the contributions to ΔG_{bind} in Eq. (1.7), combined with one or more approximations. Several examples are presented here.

1.4.2 The Quasi-Harmonic Approximation

If one assumes that fluctuations around the minimum energy configuration are Gaussian in distribution, which is equivalent to the assumption that the pmf is harmonic, one may analytically integrate the Boltzmann probability factor in Eq. (1.7) [15, 16]. Applied to the position and orientation degrees of freedom in Eq. (1.6), this gives

$$\Delta G_{\text{bind}} = kT \ln(8\pi^2 V_{\text{ref}}) + \omega_{\text{min}} - \frac{kT}{2} \ln((2\pi)^6 \det(C_{\mathbf{r},\Omega})), \tag{1.13}$$

where $C_{\mathbf{r},\Omega}$ is the 6×6 fluctuation covariance matrix of the three positional and three orientation coordinates, namely, x, y, z, and α, $\sin\beta$, γ, when using a Cartesian orientation angle coordinate set. The determinant is symbolized by det(...). The matrix elements are given by $C_{x,y} = \langle (x - \overline{x})(y - \overline{y}) \rangle$ and so on, where $\langle\rangle$ and the overbar indicate an ensemble average. The first term of Eq. (1.13) is known. Calculation of the remaining terms of Eq. (1.13) requires sampling $\mathbf{r}$, Ω, for example, by molecular dynamics, to build up $C_{\mathbf{r},\Omega}$, in addition to the calculation of ω_{min}. Here, ω_{min} is the potential of mean force at the minimum in $\mathbf{r}$, Ω space: It contains solvation terms, contributions from changes in the internal and intermolecular interaction energies of A and B, and changes in fluctuations of A and B upon binding. Note that because of this last contribution, ω_{min} here is not just the difference in A and B free energies evaluated at their conformational minima. In implementation of Eq. (1.13) and in any of the methods discussed in this section where angular variable correlations are accumulated, it is necessary to deal with the modulo 2π issue, either by using complex variable representation [17] or by explicit handling of 0–2π crossing events. Both the quasi-harmonic approximation and Schlitter's quantum mechanical version of it [18] have been applied to calculate the R/Tentropy contribution of ligand binding [19]. Both methods gave almost identical results.

One may apply the quasi-harmonic approximation still further to the internal degrees of freedom of A and B to obtain

$$\Delta G_{\text{bind}} = kT \ln(8\pi^2 V_{\text{ref}}) + \omega^*_{\text{min}} - \frac{kT}{2} \ln\left((2\pi)^6 \frac{\det(C_{\mathbf{r},\Omega,\text{A,B}})}{\det(C_{\text{A}})\det(C_{\text{B}})} \right), \tag{1.14}$$

where $C_{\mathbf{r},\Omega,\text{A,B}}$, C_{A}, and C_{B} are the coordinate fluctuation covariance matrices for the complex and A and B alone, accounting for fluctuations in $\mathbf{r}$, Ω in the complex and internal degrees of freedom in A and B in the bound and free states. The factorization in the denominator reflects the absence of correlation between A and B motions in the unbound state. If $\mathbf{r}$, Ω fluctuations are uncorrelated with internal motions, Eq. (1.14) becomes

$$\Delta G_{\text{bind}} = kT \ln\left(\frac{8\pi^2 V_{\text{ref}}}{\sqrt{(2\pi)^6 \det(C_{\text{r},\Omega})}}\right) + \omega^*_{\text{min}} - \frac{kT}{2} \ln\left(\frac{\det(C_{A,B})}{\det(C_{\text{A}})\det(C_{\text{B}})}\right), \tag{1.15}$$

where the first term now represents the contribution from restriction in R/T motion of the ligand upon binding. Implementation of Eqs. (1.14) and (1.15) requires sampling of conformations of A and B in the bound state, for example, by molecular dynamics, to build up the correlation matrices. In this form of the quasi-harmonic model, ω^*_{min} is the pmf at the minimum of position, orientation, and internal coordinates. Thus, it includes only the internal energy changes of A and B upon binding, the direct interaction energy, and solvation. Application to protein–ligand and protein–protein binding is relatively feasible. First, the method is in principle less computationally intensive than the direct simulation of the full association pmf in that only two states, the bound and free, are simulated. Second, the pmf can be obtained using a combination of molecular mechanics minimization energy and some implicit solvent model such as the Poisson–Boltzmann surface area (PBSA) model [5, 20] or the generalized Born (GB) model [21–23]. The full quasi-harmonic model has been applied with some success to binding [5, 19, 24], but in many cases the assumption of Gaussian fluctuations in R/T and internal coordinates is unrealistic [25].

1.4.3 Estimation of Entropy Contributions to Binding

The main limitation with quasi-harmonic models is in their estimation of R/T and internal conformational entropy changes. A less restrictive approach can be developed by using better approximations for conformational entropy. We start by regrouping the terms arising from fluctuations in Eq. (1.14) and breaking ω^*_{min} into components:

$$\begin{aligned}
\Delta G_{\text{bind}} &= (U_{\text{AB}} - U_{\text{A}} - U_{\text{B}}) + \frac{6kT}{2} + \left(G^{\text{solv}}_{\text{AB}} - G^{\text{solv}}_{\text{A}} - G^{\text{solv}}_{\text{B}}\right) \\
&+ kT \ln(8\pi^2 V_{\text{ref}}) \\
&- \frac{kT}{2} \ln\left((2\pi e)^{3n_{\text{A}}+3n_{\text{B}}-6} \det(C_{\text{r},\Omega,\text{A},B})\right) \\
&- \left(-\frac{kT}{2} \ln\left((2\pi e)^{3n_{\text{A}}-6} \det(C_{\text{A}})\right) - \frac{kT}{2} \ln\left((2\pi e)^{3n_{\text{B}}-6} \det(C_{\text{B}})\right)\right).
\end{aligned} \tag{1.16}$$

The first line contains the contribution of internal energy and the direct interaction energy of A and B. In addition, there is $1/2kT$ of potential energy for each degree of freedom (by equipartition) acquired by the R/T motions upon complex formation. These are all contributions to the enthalpy of binding. Then, there are solvation free energy terms. The remaining terms are all entropic. The third line is the

conformational entropy of the complex arising from internal motions of A and B and R/T motion of A with respect to B. The last line is the conformational entropy contributions from internal motions of free A and B. Thus,

$$\Delta G_{\text{bind}} = \Delta U + 3kT + \Delta G_{\text{solv}} + TS^{\text{free}}_{\text{R}/T} - T\left(S^{\text{conf}}_{\text{AB}} - S^{\text{conf}}_{A} - S^{\text{conf}}_{B}\right). \quad (1.17)$$

Now we are free to use different approximations for the entropy. The exact expression for the conformational entropy is

$$S^{\text{conf}} = -k \int \mathrm{d}\mathbf{q} p(\mathbf{q}) \ln(p(\mathbf{q})), \quad (1.18)$$

where the multidimensional integral is taken with respect to all the coordinates $\mathbf{q}$ of A, B, or AB as appropriate. In practice, the integral is of such high dimension that adequate sampling is a challenge and some approximations must be introduced. The simplest approximation is neglect of all correlations between different degrees of freedom. Then,

$$S^{\text{conf}} \approx -k \sum_{i}^{n} \int \mathrm{d}q_i p(q_i) \ln(p(q_i)), \quad (1.19)$$

where only one-dimensional probability density functions (pdfs) of each of the n coordinates q_i are needed. This would require the minimal amount of sampling. Investigation with small ligands shows that correlations contribute significantly to the entropy of binding [17, 24, 26]. The next step would be to include pairwise correlations. This can be done within the quasi-harmonic model by factoring out the leading entropy contribution, replacing it with the exact entropy expression with no correlation – Eq. (1.19) – and treating only the correlations harmonically [27]. Thus,

$$\begin{aligned} S^{\text{conf}}_{qh} &= \frac{k}{2} \ln((2\pi e)^n \det(C_{ij})) = \frac{k}{2} \ln((2\pi e)^n \prod_{i}^{n} C_{ii}) + \frac{k}{2} \ln(\det(R_{ij})) \\ &\rightarrow -k \sum_{i}^{n} \int \mathrm{d}q_i p(q_i) \ln(p(q_i)) + \frac{k}{2} \ln(\det(R_{ij})), \end{aligned} \quad (1.20)$$

where R_{ij} is the correlation coefficient matrix, whose elements are the correlation coefficients $\langle (q_i - \bar{q}_i)(q_j - \bar{q}_j) \rangle / \sqrt{\langle (q_i - \bar{q}_i)^2 \rangle \langle (q_j - \bar{q}_j)^2 \rangle}$. Another approach is to accumulate two-dimensional and higher pdfs, either directly or as part of some expansion [28–30]. For example, for a two-dimensional case, with coordinates q and s,

$$\begin{aligned} S^{\text{conf}}(q,s) &= -k \int \mathrm{d}\, q\, \mathrm{d}s p(q,s) \ln(p(q,s)) \\ &= -k \int \mathrm{d}\, q\, p(q) \ln(p(q)) - k \int \mathrm{d}\, q\, p(q) \int \mathrm{d}\, s\, p(s|q) \ln(p(s|q)) \\ &= S(q) + S(s) - I_{q,s}, \end{aligned} \quad (1.21)$$

where $p(q, s)$ is the full two-dimensional pdf, $p(s|q)$ is the marginal probability of s given q, $S(q)$ and $S(s)$ are the uncorrelated entropies from q and s obtained from one-dimensional pdfs as in Eq. (1.19), and $I_{q,s}$ is the correction or mutual information arising from the fact that fluctuations in q and s are not independent [31]. This approach can be carried to higher order, although estimating three-dimensional and higher pdfs would be extremely challenging for a protein [30]. Note that the true entropy is always less than that from uncorrelated motions given by Eq. (1.19), thus $\det(R_{ij}) < 1$, and $I_{q,s} > 0$. A variety of other methods based on approximating Eq. (1.18) have been used to calculate molecular conformational entropies. These include the hypothetical scanning approach [32], the nearest neighbor method [33], mining minima [24, 34, 35], mode scanning [36], superposition approximations, [28] minimum information expansion and nearest neighbor methods [26, 37], and adaptive density kernels [38, 39].

1.4.4
The Molecule Mechanics Poisson–Boltzmann Surface Area Method

A very practical hybrid method for calculating protein–ligand interactions is the molecule mechanics Poisson–Boltzmann surface area (MMPBSA) method [40, 41]. One runs a molecular dynamics trajectory on the complex, protein and ligand, postprocesses the snapshots, and computes the free energy of A, B, or AB as the average of

$$G = G_{\mathrm{PB}} + G_{\mathrm{np}} + E_{\mathrm{MM}} - TS_{\mathrm{solute}} \tag{1.22}$$

over the snapshots. Since the snapshots are generated from an ensemble, they are arithmetically averaged. In Eq. (1.22), G_{PB} is the electrostatic solvation free energy obtained from the Poisson–Boltzmann (PB) two-dielectric continuum electrostatics model, using 78.6 for the water. Usually the internal dielectric is set to 1 since orientational polarization effects are accounted for by atomic motions during the MD simulation [40], although an internal dielectric of 2 has been used to account for electronic polarization [42]. G_{np} is the nonpolar solvation term obtained from the molecular surface area times the hydrophobic coefficient (usually $\approx$5 cal/(mol Å^{-2})). E_{MM} is the molecular mechanics term, equivalent to the U term in Eq. (1.16). The solute entropy term can be obtained by minimizing the snapshots and running a normal mode calculation [43]. From the normal mode analysis, the harmonic model provides the entropy contributions from the R/T mode frequencies and internal motion mode frequencies. Then, $\Delta G_{\mathrm{bind}} = G_{\mathrm{AB}} - G_{\mathrm{A}} - G_{\mathrm{B}}$. To cut down the amount of computation, the MMPBSA computation is often run only on the complex trajectory. Uncomplexed structures are generated by omitting atoms of each binding partner in turn [41]. This effectively omits any contribution of a change in average conformation of protein and ligand to the binding. Contributions of changes in internal and R/T entropy to binding would also not be included unless an entropy calculation from, for example, normal modes is run on the complexed and uncomplexed structures.

1.4.5
Thermodynamic Work Methods

The relative binding free energy of A′ with respect to A is defined as

$$\Delta\Delta G_{\text{bind}} = kT \ln(K'/K), \tag{1.23}$$

where K' and K are the two association constants. Here, there is no issue with a reference state concentration. In addition, A and A′ are often closely related ligands. Both factors permit a different set of methods to be applied than that in absolute binding free energy calculations. Use is often made of the thermodynamic cycle:

$$\begin{array}{ccccc} & & & \xrightarrow{\Delta G_1} & \\ A & + & B & & AB \\ \Delta G_3 \downarrow & & & & \downarrow \Delta G_4 \\ A' & + & B & \xrightarrow[\Delta G_2]{} & A'B \end{array}$$

What is required is the difference $\Delta G_2 - \Delta G_1$. Instead, one may calculate the thermodynamic equivalent $\Delta G_4 - \Delta G_3$. This requires only the computation of the free energy difference due to changing the ligand in the bound and free states rather than calculation of two binding events. The most exact methods compute the thermodynamic work necessary to alchemically change A into A′, using either free energy perturbation (FEP) or thermodynamic integration (TI) [44, 45]. Even for relatively small differences, both methods usually require breaking up the change into small steps, in which the Hamiltonian representing the ligand is changed stepwise from A to A′ by means of a perturbation parameter λ:

$$\Delta G = \sum_{\lambda=0}^{\lambda=1} \Delta G_\lambda, \tag{1.24}$$

where at $\lambda = 0$, the ligand is A and at $\lambda = 1$, the ligand is fully A′, and λ is changed in increments of $\delta\lambda$. The free energy for each step is obtained using

$$\Delta G_\lambda = -kT \ln \left\langle e^{\beta(H(\lambda+\delta\lambda)-H(\lambda))} \right\rangle_\lambda \tag{1.25}$$

for the free energy perturbation method, where $H(\lambda)$ is the Hamiltonian for a value of λ and $\langle\rangle_\lambda$ indicates an ensemble average over configurations generated using $H(\lambda)$. In practice, one obtains two ΔG contributions, forward (using $H(\lambda + \delta\lambda)$) and backward (using $H(\lambda - \delta\lambda)$) from a single sampling. For thermodynamic integration, one computes

$$\Delta G_\lambda = \delta\lambda \left\langle \frac{\mathrm{d}H(\lambda)}{\mathrm{d}\lambda} \right\rangle_\lambda. \tag{1.26}$$

The free energy obtained from FEP and TI used in this way contains all the terms involved in binding implicitly, including changes in the R/T term. They are very

general methods amenable to various levels of treatment, including explicit or implicit solvent models, all atom models, and coarse-grained models. Since the accuracy of FEP and TI equations requires rather small changes in λ, they are very computationally intensive, even for quite similar ligand pairs.

1.4.6 Ligand Decoupling

The binding–uncoupling method is also based on computation of thermodynamic work [46, 47]. Here the parameter λ operates on the part of the Hamiltonian that describes the interaction between A and B. As it is decreased to zero, the binding partners cease to see each other, and A becomes unbound from B. Again, either FEP or TI may be used to compute the thermodynamic work of uncoupling. In order for this method to provide a true free energy change, the unbound ligand must end in a well-defined state *vis-a-vis* its translational freedom. A convenient state is a harmonic translational body restraint. In this case, adding the decoupling work to the binding free energy for a harmonic well via Eq. (1.12) provides the absolute binding free energy for the given V_{ref}. A known rotational body constraint in the unbound state may also be added to enhance sampling and convergence [48].

1.4.7 Linear Interaction Methods

Another method that uses the properties of a harmonic model is the linear interaction energy (LIE) method. If a system is fluctuating in an effective harmonic potential, then the response to some perturbation is linear. This enables one to approximate the free energy change due to this perturbation or change in the system as [49]

$$\Delta G \approx \frac{1}{2}(\langle \Delta V_{\text{a}} \rangle - \langle \Delta V_{\text{b}} \rangle), \tag{1.27}$$

where ΔV is the fluctuation in the potential energy in states a or b. Generalized somewhat to binding, LIE expression is of the following form [50, 51]:

$$\Delta G_{\text{bind}} = \sum_i \alpha_i \left(\left\langle V_i^{\text{bound}} \right\rangle - \left\langle V_i^{\text{free}} \right\rangle \right), \tag{1.28}$$

where V_i is a component of the potential energy function and the ensemble average $\langle\rangle$ is taken over both the bound state (AB) and the free state for A and B, using, for example, molecular dynamics. The components in Eq. (1.28) are those involved in the interaction between A and B, namely, the nonbonded terms. In the original formulation, these were the van der Waals (vdw) and electrostatic terms. Later an apolar solvation or hydrophobic term was added [52], although further parameterization shows some redundancy between this and the vdw term, so it is not clear if both are needed. In a truly linear response model, all the coefficients α_i would be 1/2. In practice, these are taken as adjustable parameters obtained by fitting the method to experimental binding free energies. An inconsistency in LIE model is that it has no

explicit R/T term or explicit dependence on V_{ref}, yet the numbers are compared with absolute experimental binding free energies using the usual 1 M reference. However, the model has been parameterized on a wide range of ligand binding reactions, so it is possible that the R/T term is implicitly included through the α_i terms, which often vary substantially from 1/2. Examination of Eq. (1.28) shows that the model only uses interaction potential terms in the calculated ΔG_{bind}, although all the usual potential energy terms are used in the simulations to generate the molecular and solvent conditions from which the averages are formed. Contributions from conformational changes and conformational entropy changes upon binding are effectively subsumed in the linear response by contributing to the change in average value of the different V_i terms.

1.4.8 Salt Effects on Binding

Binding of charged ligands and proteins to DNA or binding between any kind of highly charged molecules is strongly affected by salt concentration. As already discussed, highly charged molecules have activity coefficients that differ significantly from unity at any concentration. This is because of the electrostatic energy of the ionic double layer that forms around them. The salt dependence of binding is usually expressed as

$$s(K) = \frac{\mathrm{d}\ln K}{\mathrm{d}\ln[\mathrm{Salt}]}, \tag{1.29}$$

where $s(K)$ is defined as the slope of a log–log plot of the binding constant versus salt concentration. In many cases, this is linear over one–two decades of salt concentration. From the dependence of the electrostatic free energy of the molecule on the salt concentration, $s(K)$ is given by [3, 4]

$$s(K) = \Pi_{AB} - \Pi_A - \Pi_B, \tag{1.30}$$

where Π_x is the sum of the integrated excess/deficit of ions around molecule x:

$$\Pi_x = \sum_i \int \mathrm{d}\mathbf{r} c_i \left(e^{-\beta z_i e \phi(\mathbf{r})} - 1 \right), \tag{1.31}$$

where e is the unit proton charge. The sum is over all ion types i, of valence z_i and bulk concentration c_i. The electrostatic potential at position $\mathbf{r}$ is $\phi(\mathbf{r})$. The integration is over the entire solvent volume. The ion integral components of Π_x are closely related with the Donnan coefficients and preferential interaction coefficients of that ion [2–4, 53] and with the salt dependence of the activity coefficient of the molecule. These may be viewed as alternative descriptions of the same physical effect: enrichment of counterions and depletion of coions near the molecule. $s(K)$ for a particular system may be obtained by calculating Π for the bound and free states, using the nonlinear Poisson–Boltzmann model [54, 55] or Monte Carlo simulation of the preferential interaction coefficients [56, 57]. Simpler models such as the counterion–condensation model

predict that $s(K)$ is simply the net charge of the ligand [58, 59]. This model requires no computation, but it fails when applied to a complex ligand such as a protein, where charges are distributed over distances comparable to the molecular size away from the binding interface.

Besides calculating $s(K)$, PB model can also be used to calculate the net contribution of salt to binding at a given salt concentration by evaluating the salt contribution to ΔG_{solv} in Eq. (1.17). One simply recalculates the solvation energy contributions with Salt = 0 and takes the difference.

1.4.9 Statistical Potentials

From the general expression for the association constant, Eqs. (1.6) and (1.3), one finds that the pmf of interaction between the protein and the ligand determines the relative probability of a pose **(r, Ω)**. A pose will occur with a high probability if the pmf is favorable. Since the pmf is a function of the type and arrangement of atoms brought together in the complex, it is reasonable to assume that a favorable pose will have an arrangement of atoms that is seen with higher than average probability in other protein–ligand complexes. This is the motivation behind the use of statistical or knowledge-based potentials in binding [60–64]. Specifically, one analyzes the arrangement of atoms brought together in terms of distances between pairs of atoms, $r(i, j)$, where one atom is in each of the binding partners. Then, in analogy with the definition of a pmf, one can define a statistical potential of the form

$$\phi_{ij}(\mathrm{r}) = -kT \ln\left(p_{ij}(r)/p_{ij}^{0}\right), \tag{1.32}$$

where $p_{ij}(r)$ is the database-derived probability that atoms of type i and j are found at a distance r. p_{ij}^{0} is the important normalization factor for that pair type derived from the distribution that one would expect if i and j were randomly distributed. The binding free energy is then estimated as the sum of statistical potential terms over all ligand–protein atom pairs.

The normalization constant in Eq. (1.32) must take into account the fact that near the ligand one would expect less protein atoms on average, since the ligand is occupying space, and vice versa [61]. It must also account for the fact that proteins and ligands are of finite size, so at longer distances the probability of finding atom pairs of any type decreases [63]. A related aspect is that atoms near the surface of either molecule will have less pairwise interactions, but more interaction with the solvent. This latter interaction will greatly affect the binding free energy, but since it is not a pairwise atomic interaction, it must be added in separately. This can also be done through a solvent-accessible area-derived statistical potential [63] or implicit solvent models such as PBSA and GB models. As the developers of statistical potentials have made clear, they are not rigorous thermodynamic pmfs, since they are obtained from many different static structures, not from a single equilibrium ensemble. Nevertheless, they do encode energetic information about complexes. One may think of them as statistical similarity measures: A complex with a favorable statistical

potential-derived free energy will have a strong statistical similarity with other high affinity complexes.

1.4.10 Empirical Potentials

Equation 1.17 provides a conceptual starting point for more empirical methods for computing ΔG_{bind}. We assume that R/T, A, and B contributions to the conformational entropy of the complex can be separated into three components: $S_{\text{R/T,A,B}} = S_{\text{R/T}}^{\text{bound}} + S_{\text{A}}^{\text{bound}} + S_{\text{B}}^{\text{bound}}$. We also separate the internal and interaction energy components of the complex AB as $U_{\text{AB}} = U_{\text{A}}^{\text{bound}} + U_{\text{B}}^{\text{bound}} + U_{\text{A-B}}$. This gives

$$\Delta G_{\text{bind}} = \Delta U_{\text{A-B}} + \Delta U_{\text{A}} + \Delta U_{\text{B}} + 3kT + \Delta G^{\text{solv}} - T\Delta S_{\text{R/T}} - T\Delta S_{A}^{\text{conf}} - T\Delta S_{B}^{\text{conf}}, \tag{1.33}$$

where ΔU_{A} and ΔU_{B} are the changes in internal energy of A and B due to changes in their conformation, and $\Delta S_{\text{A}}{}^{\text{conf}}$ and $\Delta S_{\text{B}}{}^{\text{conf}}$ are changes in the entropy of A and B due to changes in their fluctuations upon binding. $U_{\text{A-B}}$ is the direct (*in vacuo*) interaction energy between A and B in the complex. ΔG^{solv} is the change in solvation free energy of A and B as they mutually desolvate each other upon binding. $\Delta S_{\text{R/T}} = S_{R/T}^{\text{bound}} - S_{R/T}^{\text{free}}$ is the change in ligand R/T entropy upon binding, relative to the free state at the reference concentration. Typically, empirical binding potentials assume a fixed value for the $3kT + -T\Delta S_{\text{R/T}}$ term (to be determined by fitting), they neglect the internal energy changes, and they use some kind of inventory of interactions to estimate the net effect of the remaining terms, $U_{\text{A-B}} + \Delta G_{\text{solv}} - T(\Delta S_{A}^{\text{conf}} + \Delta S_{B}^{\text{conf}})$, by using a binding potential of the form

$$\Delta G_{\text{bind}} \approx \sum c_{\text{i}} \Delta G_{\text{i}} + C, \tag{1.34}$$

where C is a constant accounting for the R/T contribution, ΔG_{i} is a free energy contribution per interaction, and c_{i} quantifies the number or extent of that interaction, depending on how it is defined. So, H-bonding, for example, would be defined in terms of the number of H-bonds and the strength of a single H-bond, whereas a hydrophobic interaction may be defined in terms of the strength per unit area and the solvent-accessible area. Conformational entropy terms, if included, may be represented by rotamer counting or inventorying the number of rotatable bonds immobilized by binding. The degree of resolution varies between potentials, so some may define different classes of H-bonds depending on the groups involved, each with different strength, or different surface free energy coefficients for different atoms or groups. Interactions may be defined at the atomic, group, or residue level. Many variants of empirical binding potentials exist. A seminal example is the SCORE potential [66]. Despite the manifold forms of empirical potentials, the general principle behind them is the same: separation of the free energy into a sum of linear terms and determination of the strength of each interaction type by extensive parameterization against experimental binding free energies. Because of this,

empirical potentials are usually most successful when they are parameterized for a specific subset of ligand–protein complexes and used within that set.

References

1 Hopkins, A.L., Groom, C.R., and Alex, A. (2004) Ligand efficiency: A useful metric for lead selection. *Drug Discovery Today*, **9**, 430–431.

2 Record, T., Lohman, T., and de Haseth, P. (1976) Ion effects on ligand–nucleic acid interactions. *Journal of Molecular Biology*, **107**, 145–158.

3 Sharp, K.A. (1995) Polyelectrolyte electrostatics: salt dependence, entropic and enthalpic contributions to free energy in the nonlinear Poisson–Boltzmann model. *Biopolymers*, **36**, 227–243.

4 Sharp, K.A., Friedman, R., Misra, V., Hecht, J., and Honig, B. (1995) Salt effects on polyelectrolyte–ligand binding: comparison of Poisson–Boltzmann and limiting law counterion binding models. *Biopolymers*, **36**, 245–262.

5 Luo, H. and Sharp, K.A. (2002) On the calculation of absolute binding free energies. *Proceedings of the National Academy of Sciences of the United States of America*, **99**, 10399–10404.

6 Gilson, M.K., Given, J.A., Bush, B.L., and McCammon, J.A. (1997) The statistical–thermodynamic basis for computation of binding affinities: a critical review. *Biophysical Journal*, **72**, 1047–1069.

7 Swanson, J.M., Henchman, R., and McCammon, J.A. (2004) Revisiting free energy calculations: A theoretical connection to MM/PBSA and direct calculation of the association free energy. *Biophysical Journal*, **86**, 67–74.

8 Woo, H.-J. and Roux, B. (2005) Calculation of absolute protein–ligand binding free energy from computer simulations. *Proceedings of the National Academy of Sciences of the United States of America*, **102**, 6825–6830.

9 Janin, J. (1995) For Guldberg and Waage, with love and cratic entropy. *Proteins: Structure, Function, and Genetics*, **24**, R1–R2.

10 Tamura, A. and Privalov, P.L. (1997) The entropy cost of protein association. *Journal of Molecular Biology*, **273**, 1048–1060.

11 Karplus, M. and Janin, J. (1999) Comment on: "The entropy cost of protein association". *Protein Engineering*, **12**, 185–186.

12 Murphy, K.P., Xie, D., Thompson, K., Amzel, M., and Freire, E. (1994) Entropy loss in biological processes: estimate of translational entropy loss. *Proteins*, **18**, 63–67.

13 Amzel, L.M. (2000) Calculation of entropy changes in biological processes: folding, binding, and oligomerization. *Methods in Enzymology*, **323**, 167–177.

14 Deng, Y. and Roux, B. (2006) Calculation of standard binding free energies: aromatic molecules in the T4 lysozyme L99A mutant. *Journal of Chemical Theory and Computation*, **2**, 1255–1273.

15 Karplus, M. and Kushick, J. (1981) Method for estimating the configurational entropy of macromolecules. *Macromolecules*, **14**, 325–332.

16 Levy, R., Karplus, M., Kushick, J., and Perahia, D. (1984) Evaluation of the configurational entropy for proteins: application to molecular dynamics simulations of an alpha helix. *Macromolecules*, **17**, 1370–1374.

17 Wang, J. and Bruschweiler, R. (2006) 2D entropy of discrete molecular ensembles. *Journal of Chemical Theory and Computation*, **2**, 18–24.

18 Schlitter, J. (1993) Estimation of absolute and relative entropies of macromolecules using the covariance matrix. *Chemical Physics Letters*, **215**, 617–621.

19 Carlsson, J. and Aqvist, J. (2005) Absolute and relative entropies from computer simulation with application to ligand binding. *The Journal of Physical Chemistry B*, **109**, 6448–6456.

20 Prabhu, N.V., Zhu, P.-J., and Sharp, K.A. (2004) Implementation and testing of

stable, fast implicit solvation in molecular dynamics using the smooth-permittivity finite difference Poisson–Boltzmann method. *Journal of Computational Chemistry*, **25**, 2049–2064.

21 Tsui, V. and Case, D.A. (2001) Calculations of the absolute free energies of binding between RNA and metal ions using molecular dynamics simulations and continuum electrostatics. *The Journal of Physical Chemistry B*, **105**, 11314–11325.

22 Still, C., Tempczyk, A., Hawley, R., and Hendrickson, T. (1990) Semianalytical treatment of solvation for molecular mechanics and dynamics. *Journal of the American Chemical Society*, **112**, 6127–6129.

23 Dominy, B. and Brooks, C. (1999) Development of a generalized Born model parameterization for proteins and nucleic acids. *The Journal of Physical Chemistry B*, **103**, 3765–3773.

24 Chang, C.A., Chen, W., and Gilson, M.K. (2007) Ligand configurational entropy and protein binding. *Proceedings of the National Academy of Sciences of the United States of America*, **104**, 1534–1539.

25 Chang, C.-E., Chen, W., and Gilson, M.K. (2005) Evaluating the accuracy of the quasiharmonic approximation. *Journal of Chemical Theory and Computation*, **1**, 1017–1028.

26 Killian, B.J., Yudenfreund Kravitz, J., Somani, S., Dasgupta, P., Pang, Y.-P., and Gilson, M.K. (2009) Configurational entropy in protein–peptide binding. *Journal of Molecular Biology*, **389**, 315–335.

27 Di Nola, A., Berendsen, H., and Edholm, O. (1984) Free energy determination of polypeptide conformations generated by molecular dynamics. *Molecular Physics*, **17**, 2044–2050.

28 Killian, B.J., Kravitz, J., and Gilson, M.K. (2007) Extraction of configurational entropy from molecular simulations via an expansion approximation. *Journal of Chemical Physics*, **127**, 024107.

29 Baron, R., Hunenberger, P.H., and McCammon, J.A. (2009) Absolute single-molecule entropies from quasi-harmonic analysis of microsecond molecular dynamics: correction terms and convergence properties. *Journal of Chemical Theory and Computation*, **5**, 3150–3160.

30 Li, D.-W., Showalter, S., and Bruschweiler, R. (2010) Entropy localization in proteins. *The Journal of Physical Chemistry B*, **114**, 16036–16044.

31 Zhou, H.-X. and Gilson, M. (2009) Theory of free energy and entropy in noncovalent binding. *Chemical Reviews*, **109**, 4092–4107.

32 Cheluvaraja, S. and Meirovitch, H. (2006) Calculation of the entropy and free energy of peptides by molecular dynamics simulations using the hypothetical scanning molecular dynamics method. *Journal of Chemical Physics*, **125**, 024905–024913.

33 Hnizdo, V., Darian, E., Fedorowicz, A., Demchuk, E., Li, S., and Singh, H. (2007) Nearest-neighbor nonparametric method for estimating the configurational entropy of complex molecules. *Journal of Computational Chemistry*, **28**, 655–668.

34 Head, M.S., Given, J.A., and Gilson, M.K. (1997) Mining minima: direct computation of conformational free energy. *The Journal of Physical Chemistry A*, **101**, 1609–1618.

35 David, L., Luo, R., and Gilson, M.K. (2001) Ligand–receptor docking with the mining minima optimizer. *Journal of Computer-Aided Molecular Design*, **15**, 157–171.

36 Chang, C.-E., Potter, M.J., and Gilson, M.K. (2003) Calculation of molecular configuration integrals. *Journal of Physical Chemistry B*, **107**, 1048–1055.

37 Hnizdo, V., Tan, J., Killian, B.J., and Gilson, M.K. (2008) Efficient calculation of configurational entropy from molecular simulations by combining the mutual-information expansion and nearest neighbor methods. *Journal of Computational Chemistry*, **29**, 1605–1614.

38 Hensen, U., Lange, O.F., and Grubmuller, H. (2010) Estimating absolute configurational entropies of macromolecules: the minimally coupled subspace approach. *PLoS One*, **5**, e9179.

39 Hensen, U., Grubmuller, H., and Lange, O.F. (2009) Adaptive anisotropic kernels for nonparametric estimation of absolute configurational entropies in

high-dimensional configuration spaces. *Physical Review E*, **80**, 011913.

40 Srinivasan, J., Cheatham Thomas, E., III, Cieplak, P., Kollman, P.A., and Case, D. (1998) Continuum solvent studies of the stability of DNA, RNA and phosphoramidate. *Journal of the American Chemical Society*, **120**, 9401–9409.

41 Kuhn, B. and Kollman, P.A. (2000) Binding of a diverse set of ligands to avidin and streptavidin. *Journal of Medicinal Chemistry*, **43**, 3786–3791.

42 Eriksson, M., Pitera, J., and Kollman, P. (1999) Prediction of the binding free energies of new TIBO-like HIV-1 reverse transcriptase inhibitors using a combination of PROFEC, PB/SA, CMC/MD, and free energy calculations. *Journal of Medicinal Chemistry*, **42**, 868–881.

43 Chong, L., Duan, Y., Massova, I., and Kollman, P. (1999) Molecular dynamics and free-energy calculations applied to affinity maturation in antibody 48G7. *Proceedings of the National Academy of Sciences*, **96**, 14330–14335.

44 Lybrand, T., McCammon, J.A., and Wipf, G. (1986) Theoretical calculation of relative binding affinity in host–guest system. *Proceedings of the National Academy of Sciences*, **83**, 833.

45 Beveridge, D. and DiCapua, F.M. (1989) Free energy via molecular simulations: applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry*, **18**, 431.

46 Hermans, J. and Shankar, S. (1986) The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Israel Journal of Chemistry*, **27**, 225–227.

47 Roux, B., Nina, M., Pomes, R., and Smith, J.C. (1996) Thermodynamic stability of water molecules in the bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophysical Journal*, **71**, 670–681.

48 Hermans, J. and Wang, L. (1997) Inclusion of loss of translational and rotational freedom in theoretical estimates of free energies of binding. *Journal of the American Chemical Society*, **119**, 2707–2714.

49 Aqvist, J. and Hansson, T. (1996) On the validity of electrostatic linear response in polar solvents. *The Journal of Physical Chemistry*, **100**, 9512–9521.

50 Hansson, T., Marelius, J., and Aqvist, J. (1998) Ligand binding affinity prediction by linear interaction energy methods. *Journal of Computer-Aided Molecular Design*, **12**, 27–35.

51 Zhou, R., Friesner, R.A., Ghosh, A., Rizzo, R.C., Jorgensen, W.L., and Levy, R.M. (2001) New linear interaction method for binding affinity calculations using a continuum solvent model. *The Journal of Physical Chemistry B*, **105**, 10388–10397.

52 Jones-Herzog, D. and Jorgensen, W.L. (1997) Binding affinities for sulfonamide inhibitors with human thrombin using Monte Carlo simulations with a linear response method. *Journal of Medicinal Chemistry*, **40**, 1539.

53 Stigter, D. (1960) Interactions in aqueous solutions: III. On statistical thermodynamics of colloidal electrolytes. *The Journal of Physical Chemistry*, **64**, 838–842.

54 Misra, V., Hecht, J., Sharp, K., Friedman, R., and Honig, B. (1994) Salt effects on protein–DNA interactions: the lambda cI repressor and EcoR1 endonuclease. *Journal of Molecular Biology*, **238**, 264–280.

55 Misra, V., Sharp, K., Friedman, R., and Honig, B. (1994) Salt effects on ligand–DNA binding: minor groove antibiotics. *Journal of Molecular Biology*, **238**, 245–263.

56 Olmsted, M.C., Anderson, C.F., and Record, M.T. (1989) Monte Carlo description of oligoelectrolyte properties of DNA oligomers. *Proceedings of the National Academy of Sciences of the United States of America*, **86**, 7766–7770.

57 Record, T., Olmsted, M., and Anderson, C. (1990) Theoretical studies of the thermodynamics of ion interaction with DNA, in *Theoretical Biochemistry and Molecular Biophysics* (eds. D. Beveridge and R. Lavery), Adenine Press.

58 Manning, G. (1969) Limiting laws and counterion condensation in polyelectrolyte solutions 1: colligative

properties. *The Journal of Physical Chemistry*, **51**, 924.

59 Friedman, R.A. and Manning, G.S. (1984) Polyelectrolyte effects on site-binding equilibria with application to the intercalation of drugs into DNA. *Biopolymers*, **23**, 2671–2714.

60 Muegge, I. and Martin, Y.C. (1999) A general and fast scoring function for protein–ligand interactions: a simplified potential approach. *Journal of Medicinal Chemistry*, **42**, 791–804.

61 Muegge, I. (2006) PMF scoring revisited. *Journal of Medicinal Chemistry*, **49**, 5895–5902.

62 Gohlke, H., Hendlich, M., and Klebe, G. (2000) Knowledge-based scoring function to predict protein–ligand interactions. *Journal of Molecular Biology*, **295**, 337–356.

63 Gohlke, H. and Klebe, G. (2001) Statistical potentials and scoring functions applied to protein–ligand binding. *Current Opinion in Structural Biology*, **11**, 231–235.

64 Velec, H., Gohlke, H., and Klebe, G. (2005) DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction. *Journal of Medicinal Chemistry*, **48**, 6296–6303.

65 Bohm, H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *Journal of Computer-Aided Molecular Design*, **8**, 243–256.