On the calculation of absolute macromolecular binding free energies

Hengbin Luo and Kim Sharp*

Department of Biochemistry and Biophysics, University of Pennsylvania, Philadelphia, PA 19104-6059

The standard framework for calculating the absolute binding free energy of a macromolecular association reaction A + B → AB with an association constant K_{AB} is to equate chemical potentials of the species on the left- and right-hand sides of this reaction and evaluate the chemical potentials from theory. This theory involves (usually hidden) assumptions about what constitutes the bound species, AB, and where the contribution of the solvent appears. We present here an alternative derivation that can be traced back to Bjerrum, in which the expectation value of K_{AB} is obtained directly through the statistical mechanical method of evaluating its ensemble (Boltzmann-weighted) average. The generalized Bjerrum approach more clearly delineates: (*i*) the different contributions to binding; (*ii*) the origin of the much-discussed and somewhat controversial association entropy term; and (*iii*) where the solvent contribution appears. This approach also allows approximations required for practical evaluation of the binding constant in complex macromolecular systems, to be introduced in a well defined way. We provide an example, with application to test cases that illustrate a range of binding behavior.

The importance to biochemistry of being able to calculate absolute and relative binding affinities from the structures of macromolecular complexes is clear. A large body of literature on the theory, practice, and application exists. See, for example, refs. 1–8 for recent papers that discuss or review the theory and methods. Nevertheless, there is still controversy over some basic theoretical issues. There is agreement that association is accompanied by a loss of translational and rotational (association) entropy but disagreement over the magnitude, whether it varies significantly from system to system or can be approximated by a single value, what theoretical model should be used to estimate it, and how to include solvent effects. This controversy has motivated recent experimental studies using tethered dimerization to estimate translational entropy loss (9, 10). Accurate accounting of the association term is essential for the calculation of absolute binding free energies. Alternatively, if it could be reliably established by experiment or theory that this term is constant for some set of binding reactions, this would greatly simplify the calculation of relative binding free energies for those systems.

The traditional theoretical framework for calculating the absolute binding free energy for the reaction A + B → AB is to equate the chemical potentials of the species on the left- and right-hand sides of this reaction and to evaluate the chemical potentials of the three species, A, B, and AB by using some theory, as described in Hill's classic statistical mechanics treatment of association (11). Recent work describes the extension of this approach to macromolecular association and to ligand–membrane association, along with a detailed discussion of other theoretical issues (1, 2). However, the chemical potential approach is not the only way to approach the problem. We present here an alternative derivation, which can be traced back to Bjerrum, in which the expectation value of the association constant K_{AB} is obtained directly from ensemble average quantities. Our approach also, like several previous treatments of binding (see, e.g., refs. 1, 12, and 13) emphasizes the advantage of treating the interaction between associating partners by using a potential of mean force. Our goals are: First, to provide a straightforward but rigorous theoretical treatment of binding free energy. Second, to clarify, in response to recent work, the origin of the association entropy term, how it may be calculated or measured, and what the role of the solvent is in this term. Third, to use the theoretical framework to guide the approximations one must inevitably introduce for a computational tractable method for macromolecular association in solvent.

Theory

The "Bjerrum" approach to calculating the absolute binding free energy of a reaction A + B → AB is simply to calculate the expectation value of the equilibrium constant K_{AB} = [AB]/[A][B] directly from the usual statistical mechanical expression for evaluating the required observables (in this case concentrations) from Boltzmann-weighted averages. Consider one molecule each of A and B in a large volume V of solvent. The same equilibrium constant will be obtained from this system averaged over a sufficiently long time as from a large number of A and B molecules at equilibrium. Without loss of generality, we define the coordinate system by placing A in a fixed orientation at the origin. A particular configuration of this system is then specified by the N solvent atom coordinates, $\mathbf{r}_V = (\mathbf{r}_{V1}, \mathbf{r}_{V2} \ldots \mathbf{r}_{VN})$, the position and orientation coordinates of B with respect to A, $\mathbf{r}$, and Ω, respectively, and the internal coordinates of A and B and $\mathbf{q}_A$ and $\mathbf{q}_B$, respectively. The orientation coordinate is $\Omega = \{\phi, \psi, \chi\}$, where ϕ and ψ define an axis in polar coordinates, and χ is the rotation about that axis. Some subset of all possible configurations corresponds to the associated state (AB is present); the rest correspond to the dissociated state (A and B present). The ratio of average concentrations is

$$K_{AB} = \frac{[AB]}{[A][B]} = f_{AB}/V \big/ (f_A/V)(f_B/V), \qquad [1]$$

where f_i is the fraction of the configurations in which the species i = A, B or AB is present. To evaluate f_i, we define a function $H(\mathbf{r}, \Omega)$, which is 1 for those positions/orientations of B where it forms a complex with A, and which is 0 otherwise. f_i may now be evaluated from ensemble averages of $H(\mathbf{r}, \Omega)$:

$$f_{AB} = \langle H(\mathbf{r}, \Omega) \rangle, f_A = f_B = \langle 1 - H(\mathbf{r}, \Omega) \rangle. \qquad [2]$$

Now A and B interact with some potential of mean force (pmf), which depends on their relative position and orientation and which is defined as

$$\omega(\mathbf{r}, \Omega) = -\int_{r=\infty}^{r,\Omega} \left\langle dU(\mathbf{r}_v, \mathbf{q}_a, \mathbf{q}_b, \mathbf{r}, \Omega)/d\mathbf{r}, \Omega \right\rangle_{r,\Omega\text{fixed}} d\mathbf{r}d\Omega, \qquad [3]$$

where U is potential energy of the whole system. The pmf $\omega(\mathbf{r}, \Omega)$ is the reversible thermodynamic work necessary to bring A and B from a large separation (where ω = 0 by definition) to some separation/orientation $(\mathbf{r}, \Omega)$. It is an exact quantity obtained

BIOPHYSICS

from the integration of the Boltzmann-weighted solvent and internal coordinate averaged force between A and B (14). Clearly, the region where ω is significantly negative (< -1 kT) provides one intuitively appealing criterion for defining the bound state, $H(\mathbf{r}, \mathbf{\Omega}) = 1$. This point is discussed below. In the dilute limit ($V \to \infty$), clearly $f_A, f_B \to 1, f_{AB} \to 0$, but the product $f_{AB}V$ tends to a limit, equal to the binding constant, of

$$
\langle H(\mathbf{r}, \mathbf{\Omega}) \rangle V \big|_{\lim V \to \infty} = \frac{V \int H(\mathbf{r}, \mathbf{\Omega}) e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega}}{\int e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega}} \Bigg|_{\lim V \to \infty}
$$

$$
= \frac{1}{8\pi^2} \int H(\mathbf{r}, \mathbf{\Omega}) e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega} = K_{AB},
$$

[4]

where over most of the integration volume $\omega = 0$, the integrand of the denominator is unity, and its integral is just B's rotation/translation (R/T) phase space volume $8\pi^2V$. The binding constant in Eq. 4 can be written $K_{AB} = V_{BPV}/8\pi^2$, which defines a Boltzmann-weighted phase volume (BPV) as $V_{BPV} = \int He^{-\beta\omega} d\mathbf{r} d\mathbf{\Omega}$ with units (length·radians)^3. K_{AB} has the units of volume in which concentrations are expressed (usually 1 liter/mol = 1,660 Å^3/molecule), setting the units for integration length of $d\mathbf{r}$ (11.84 Å in this example). It is customary to define the absolute binding free energy as

$$
\Delta G_{AB} = -kT \ln K_{AB}/V_{ref},
$$

[5]

where it is understood that this refers to the free energy of binding B from a reference concentration $1/V_{ref}$, although because V_{ref} is usually omitted in the literature, the choice of reference state must be inferred from the units used for K_{AB}. Differentiation of Eqs. 4 and 5 by T gives the enthalpy and entropy contributions,

$$
\Delta H = \langle\omega\rangle_{H=1} - T \left\langle \frac{d\omega}{dT} \right\rangle_{H=1}
$$

[6]

$$
\frac{\Delta S}{k} = \ln\left(\frac{1}{8\pi^2 V_{ref}} \int H(\mathbf{r}, \mathbf{\Omega}) e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega}\right)
$$

$$
+ \beta\langle\omega\rangle_{H=1} - \frac{1}{k}\left\langle\frac{d\omega}{dT}\right\rangle_{H=1},
$$

[7]

where

$$
\langle\ldots\rangle_{H=1} = \frac{\int (\ldots) H(\mathbf{r}, \mathbf{\Omega}) e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega}}{\int H(\mathbf{r}, \mathbf{\Omega}) e^{-\beta\omega(\mathbf{r}, \mathbf{\Omega})} d\mathbf{r} d\mathbf{\Omega}}
$$

[8]

indicates an ensemble average over the bound state phase volume. In the bound state, the probability of a particular position/orientation of B is $p(\mathbf{r}, \mathbf{\Omega}) = \exp(-\beta\omega(\mathbf{r}, \mathbf{\Omega}))/V_{BPV}$, whereas in the unbound state, the probability is $1/(8\pi^2 V_{ref})$, thus the loss of R/T entropy on binding may be formally defined, by using $S = -k \int p \ln p$, as

$$
\frac{\Delta S_{R/T}}{k} = \langle \ln p(\mathbf{r}, \mathbf{\Omega}) \rangle_{H=1} + \ln\left(\frac{1}{8\pi^2 V_{ref}}\right)
$$

$$
= \beta\langle\omega\rangle_{H=1} + \ln\left(\frac{V_{BPV}}{8\pi^2 V_{ref}}\right),
$$

[9]

which can also be obtained from Eq. 7 by setting the last (solvent/internal entropy) term to zero (see also equation 51 and following discussion of ref. 1).

Eqs. 4, 6, and 7 are exact for dilute solutions. However, direct evaluation of these expressions to sufficient accuracy is likely be impractical for most binding reactions of interest to biochemists, so it is necessary to introduce some approximations for computational tractability. We illustrate one approach, based on the quasiharmonic method (although we emphasize this involves significant approximations, and that other approaches are possible), and consider only cases where the pmf between A and B has a single well defined minimum ω_{min}. Methods for multiple minima are discussed in detail in recent papers (8, 15). The energy of the AB complex near the minimum can be expanded in a Taylor series and to first order is a harmonic function of displacement in the translational, rotational, and internal coordinates away from this minimum, $\Delta\mathbf{q} = (\Delta\mathbf{q}_a, \Delta\mathbf{q}_b, \Delta\mathbf{r}, \Delta\mathbf{\Omega})$:

$$
\omega = \omega_{min} + \frac{1}{2}\Delta\mathbf{q}\cdot\mathbf{F}\cdot\Delta\mathbf{q},
$$

[10]

where $\mathbf{F}$ is the force constant matrix. The quasiharmonic approximation of $\mathbf{F}$ is (16, 17)

$$
\mathbf{F} = kT[\sigma_{ij}^2]^{-1},
$$

[11]

where $[\sigma_{ij}^2]^{-1}$ is the inverse of the coordinate fluctuation covariance matrix $\sigma_{ij}^2 = \langle\Delta q_i \Delta q_j\rangle$. The configurational integral over these coordinates is then

$$
\int_0^\infty e^{-\beta\omega(\Delta\mathbf{q}_a, \Delta\mathbf{q}_b, \Delta\mathbf{r}, \Delta\mathbf{\Omega})} = e^{-\beta\omega_{min}} \sqrt{(2\pi)^n |\sigma_{ij}^2|},
$$

[12]

where n is the total number of internal, translational, and orientational degrees of freedom, and $|\sigma_{ij}^2|$ is the determinant of the coordinate fluctuation covariance matrix, σ_{ij}^2. The usefulness of the quasiharmonic method is that even if the fluctuations in $\mathbf{q}_a$, $\mathbf{q}_b$, $\mathbf{r}$, and $\mathbf{\Omega}$ about the minimum energy structure are not strictly Gaussian (i.e., the energy surface is anharmonic), the integral of the Boltzmann factor is often well approximated by mean squared magnitude of the fluctuations $|\sigma_{ij}^2|$.

Eq. 12 accounts for the correlation between internal, translational, and orientational motions in the off-diagonal terms of σ_{ij}^2. A further reasonable simplification is that different kinds of motions are uncorrelated, so that internal motions of A, B, the three translational motions of B, and the librational motion of B with respect to A are independent, i.e., the corresponding off-diagonal terms in σ_{ij}^2 are zero. The determinant may then be factored, and substitution of Eq. 12 into Eq. 6 yields

$$
K_{AB} = e^{-\beta\omega_{min}} \cdot \sqrt{8\pi^3} \sigma_x \sigma_y \sigma_z \cdot \frac{\sigma_\chi^3}{\sqrt{6^3\pi}} \cdot \sqrt{\frac{|\sigma_{A(bnd)}^2||\sigma_{B(bnd)}^2|}{|\sigma_{A(free)}^2||\sigma_{B(free)}^2|}},
$$

[13]

where $\sigma_k = x, y, z$ and σ_χ are the rms fluctuations in position and orientation of B with respect to A in the complex. The contribution from libration of B in the complex is obtained by assuming that it is isotropic (the axis ϕ/ψ is uniformly oriented in space) and quasiharmonic over small magnitudes of χ, which results in the factor of $\sigma_\chi^3/\sqrt{6^3\pi}$ (18). Fluctuations in internal coordinates are assumed to be correlated within each molecule but uncorrelated with fluctuations in the position and internal coordinates of the other molecule. Because ω is defined with respect to A and

B at large separation, the contributions of internal fluctuations of A and B in the bound (bnd) state are normalized by those in the dissociated (free) state, namely $|\sigma_{i(bnd)}^2|/|\sigma_{i(free)}^2|$, $i =$ A, B, and contribute only if they change on binding. Because we have used the quasiharmonic approximation to factor out and treat explicitly internal coordinate fluctuation contributions to the pmf, ω_{min} in Eq. **13** contains just the intermolecular interactions between A and B at the minimum energy conformation, the solvent contributions, plus any internal energy contributions due to the change in the average conformation of A and B on binding. The last three factors in Eq. **13** are the contributions from residual translational motion, residual librational motion, and changes in internal fluctuations, respectively. The Gaussian distribution of **r**, Ω about the pmf minimum means that the greatest contributions to the integral of $H(\mathbf{r}, \Omega)$ come from near the minimum, and that contributions further away rapidly decrease because of exponential weighting. Thus little error comes from dropping the restriction of the integral to the region where $H(\mathbf{r}, \Omega) = 1$ entailed in the upper integration limit of ∞ in Eq. **12**.

The quasiharmonic approximations for the binding enthalpy and entropy are obtained by differentiation of Eq. **13** with respect to T:

$$\Delta H = \omega_{min} - T\frac{d\omega_{min}}{dT} + \frac{6kT}{2} \qquad [14]$$

and

$$\frac{\Delta S}{k} = \ln\left(\frac{e^{3/2}\sqrt{2\pi^3}\,\sigma_x\sigma_y\sigma_z}{V_{ref}}\right) + \ln\left(\frac{e^{3/2}\sigma_\chi^3}{\sqrt{6^3\pi}}\right) + \ln\left(\sqrt{\frac{|\sigma_{a(bnd)}^2||\sigma_{b(bnd)}^2|}{|\sigma_{a(free)}^2||\sigma_{b(free)}^2|}}\right)$$
$$- \frac{1}{k}\frac{d\omega_{min}}{dT}, \qquad [15]$$

where the four terms in Eq. **15** can be identified as the translational, rotational, internal, and solvent contributions to entropy, respectively. The solvent contribution arises from changes in the water configurations, as distinct from the internal and intermolecular configurations of A and B.

Methods

The quasiharmonic approximation of the generalized Bjerrum approach is implemented in an extension of our previous approach to absolute binding free energy calculations (19). Starting from a known complex structure (derived from x-ray crystallography, NMR, or model building), Eq. **13** is evaluated by:

(*i*) Minimization of AB, A, and B by using molecular mechanics, to obtain the structures for the complex at the pmf minimum, ω_{min}, and the unbound species, respectively.
(*ii*) Determination of the solvation free energies of A, B, and AB by using the Finite Difference Poisson–Boltzmann (FDPB)/ surface area method (20) to yield:

$$\omega_{min} = \Delta U_{internal} + \Delta U_{intermolecular} + \gamma\Delta Area + \Delta G_{solvation}^{electrostatic}, \qquad [16]$$

where $\Delta U_{internal} + \Delta U_{intermolecular}$ is obtained from the difference in molecular mechanics potential energy of A and B in the bound and free states, ΔArea is the change in solvent accessible surface area on binding, γ is the apolar hydration surface free energy coefficient, and $\Delta G_{solvation}^{electrostatic}$ is the difference in PB electrostatic solvation free energy between the bound and free molecules. The solvent contribution to the entropy of binding (last term of Eq. **15**) arises from the temperature dependence of γ and $\Delta G_{solvation}^{electrostatic}$, i.e., from the temperature dependence of the hydrophobic strength and the dielectric constant of water, ε. Here we do not split the solvation free energy contribu-

tions into enthalpic and entropic terms by including the temperature dependence of γ and ε, but it would be important to do so when comparing total entropies of binding with experimental data, otherwise the estimation of the translational and rotational contributions would be in serious error.
(*iii*) Molecular dynamics simulations of A, B, and AB followed by analysis of coordinate snapshots to determine σ_X, σ_Y, σ_Z, σ_χ, $|\sigma_{i(bnd)}^2|$ and $|\sigma_{i(free)}^2|$.

Thus the quasiharmonic assumption allows us to replace the evaluation of the pmf over the entire $H(r, \Omega) = 1$ region with a less expensive (but still computationally intensive) evaluation at a single point (ω_{min}) combined with an estimate of the "width" of the binding well (the rotational and translational entropy terms). We reemphasize the point made by Gilson *et al.* (1) that without the latter contribution, ω_{min}, even though it is a free energy, cannot be meaningfully compared to a measured absolute binding free energy. Furthermore, estimation of ω_{min} by the difference in potential energy between bound and free states is usually a poor estimator of even relative free energies, because it also lacks the solvent and molecular internal entropies.

To illustrate the contribution of the different terms to the binding constant and free energy, we examined four test cases that exhibit a range of binding characteristics: ion–cryptate binding, *N*-methyl acetamide (NMA) dimerization, antibody FAB fragment– digoxigenin binding, and streptavidin–biotin binding.

Minimization and Molecular Dynamics. Generation of the NMA structure, model building, minimization, and molecular dynamics were carried out with the INSIGHTII and DISCOVER packages with the consistent valence forcefield (CVFF) (Accelrys, San Diego), except for biotin, for which partial atomic charges were taken from Miyamoto and Kollman (21). The ion–tricyclic cryptate complex Cl⁻SC24 structure was taken from Metz *et al.* (22). Br⁻SC24 was built by changing the identity of the ion and reminimizing the structure. The structure of the FAB fragment 21-10-digoxigenin complex (23) was taken from entry 1IGJ of the Protein Data Bank (PDB); the structure of the streptavidin–biotin complex was taken from PDB entry 1STP (24).

Hydrogens were added and the structures minimized by conjugate gradient with a nonbond cutoff of 12 Å and a dielectric of 1. Molecular dynamics simulations were performed with a time step of 1 fs, nonbonded cutoff of 12 Å, temperature of 298 K, and a distant dependent dielectric of 1R. The distant-dependent dielectric is a rather crude approximation for the energetic effects of solvent, but it is convenient, rapid, and produces reasonable dynamic behavior, which is appropriate because the dynamics is used in our protocol not for solvent energetics but for estimating the amount of internal and intermolecular motion of A and B. Coordinates were saved every 100 steps. To compute $|\sigma_{i(bnd/free)}^2|$, coordinate snapshots of either A or B in the complex or alone were overlaid by using a rigid body alignment. The mean coordinates and the coordinate fluctuation covariance matrix were then calculated. The matrix was diagonalized using the Housholder method (25) and the eigen values determined. The six lowest values (R/T modes) were generally close to zero, as expected, and were discarded. The determinant of the internal coordinate fluctuations was then computed as the product of the remaining eigen values. To compute σ_X, σ_Y, σ_Z, and σ_χ coordinate snapshots of the complex were overlaid by using rigid body alignment of the largest partner (A). The mean position of B was computed. Then for each aligned snapshot of the complex, the required translation (x, y, z) and rotation (θ, ψ, χ) required to realign B with its average position were computed and the variances calculated. Numerical precisions were estimated from the variance in the final quantities between batches of 0.1-ns simulations.

Table 1. Binding free energy contributions at 298 K, kcal/mol

Contribution	Sc24 (Cl-Br)*	NMA	Ab-DXN	BTN-STP
Polar solvation	4.8 ± 0.5[†]	7.9 ± 0.4	16.3 ± 0.5	84.4 ± 1
Nonpolar solvation	0	−0.6 ± 0.1	−4.1 ± 0.1	−3.3 ± 0.1
Total solvation	4.8 ± 0.5	7.3 ± 0.4	12.2 ± 0.5	81.1 ± 1
van der Waals + internal	−4.0	−2.0	−2.3	−44.8
Electrostatic	−6.5[‡]	−7.0	−31.5	−62.1
Translation (at 1 M)	−0.7 ± 0.1	2.4 ± 0.1[§]	4.3 ± 0.1	4.3 ± 0.1
Rotation	0	0.8 ± 0.1	5.0 ± 0.1	4.2 ± 0.1
Internal vibration	1.6 ± 0.5	1.0 ± 2	0.0 ± 3	3.5 ± 4
Total nonsolvation	−9.6 ± 0.6	−4.8 ± 2	−24.5 ± 3	−94.9 ± 4
Total ΔG at 1 M	−4.8 ± 0.8	2.5 ± 2.5[§]	−12.3 ± 3	−13.8 ± 5
Measured ΔG at 1 M[¶]	−4.2	3.1	−13.9	−18.3

Ab-DXN, antibody FAB fragment–digoxigenin; BTN-STP, biotin–streptavidin.
*Binding of Cl relative to Br.
[†]−5.7 kcal/mol of which comes from the difference in ion hydration free energies.
[‡]Includes −2.8 kcal/mol from polarization of Sc24 by the anion.
[§]Includes a factor of ln(2) for the exchange symmetry of the dimer.
[¶]Data taken from refs. 31, 34, 49, and 50.

FDPB/Surface Area. The procedure for calculating the polar and nonpolar contributions to the solvation energy was taken from previous work (20, 26, 27). Atomic radii were taken from the PARSE set (20), except for Cl⁻ and Br⁻ radii of 1.94 and 2.09 Å, respectively (28), and digoxigenin, for which consistent valence forcefield (CVFF) radii were used. The nonpolar contribution to solvation was obtained by using $\gamma = 6$ cal/mol/Å². Solvent-accessible surface areas were calculated by using the SURFCV algorithm (29). The polar contribution was calculated by solving the linearized PB equation using a cubic box with 161 grid points on each side, with a box fill of 80%, with a resulting scale, depending on the size of the complex, from 1.5 to 4 grids/Å. The external dielectric constant was either 80 or 1 for solvent and vacuum, respectively. For ion binding to the cryptate SC24, significant polarization of the cryptate ligand groups from the strong ion field is likely. Polarization cannot be included in the CVFF. However, to estimate its effects, we used the local dielectric constant (LDC) model (30) with the FDPB method. In this method, the calculated difference in electrostatic energy of the ion binding to the neutral cryptate in vacuum with a cryptate dielectric of 1 (no Sc24 polarization) and a cryptate dielectric of 2.31 (derived from the polarizability in the LDC model) is used to estimate the effect of polarization. Numerical precision in the solvation calculations was estimated by repeating the calculations with different mappings of the molecules onto the finite difference grid.

Results

The first system studied was the binding of an ion to Sc24, a tricyclic cryptate specific for monovalent anions. Relative binding data are available (31), which show that Cl⁻ binds about 1,000-fold more tightly than Br⁻, and one of the earliest applications of the free energy perturbation (FEP) method was to this system (32). Because there are no rotational and internal fluctuation contributions from the spherical ion, the number of factors contributing to the binding free energy is reduced. The binding is dominated by a strong electrostatic interaction (Table 1), and the stronger binding of Cl⁻ vs. Br⁻ is due to the more favorable intermolecular electrostatic interactions of the former: The smaller Cl⁻ radius permits the liganding groups to approach more closely (defined by the distance of the ligand groups from the ion center, the effective locus of the charge distribution). There are less internal vibrational fluctuations in Sc24 with Cl⁻ bound than Br⁻, a consequence of the stronger intermolecular charge–charge interactions with the smaller ion. Cl⁻ has

Table 2. Summary of translation and rotation contributions to binding at 298 K

Contribution	Cl⁻Sc24	NMA₂	26-10-DXN	BTN-STP
ΔG_{bind} at 1 M	−5.6–6.8	2.5	−12.3	−13.8
$\sigma_x \sigma_y \sigma_z$ Å³	0.002	1.9	0.08	0.08
σ_χ (132°)*	NA[†]	106°	10°	16°
$\Delta H^{trans} + \Delta H^{rot}$	0.89	1.78	1.78	1.78
$-T\Delta S^{trans}$ at 1 M	5.6	1.5	3.4	3.4
$-T\Delta S^{rot}$	0	<0.1	4.1	3.3
$-T\Delta S^{assoc}$ at 1 M	5.6	1.5	7.5	6.7

DXN, digoxigenin; BTN-STP, biotin–streptavidin.
*Expected for isotropic libration.
[†]Not applicable for a monoatomic ligand.

a greater desolvation penalty. Both these effects reduce the specificity for Cl⁻. However, the relative translational contribution favors the smaller ion because it has slightly more freedom to move in the cage formed by Sc24, because, even though the liganding groups are closer to the center of the Cl⁻ ion there is more space between the van der Waals surface of the ion and the liganding groups. Polarization of the Sc24 liganding NH groups by the strong field of the ion is significant and favors Cl⁻ binding because the liganding groups are closer to the ion. When polarization is included, good agreement with experiment is obtained. We note that in the original FEP study, the polarization effect could not be included, because standard molecular dynamics force fields are not polarizable. However, the FEP simulation also underestimated the relative solvation energies of Cl⁻ vs. Br⁻ as −3.35 kcal/mol. The experimental value is −6.0 kcal/mol (33), close to the value calculated by the FDPB method (Table 1 legend). It seems likely that in the earlier FEP study, these two factors almost canceled each other.

Dimerization of NMA in water and other solvents has been measured by Kresheck and Klotz (34), and this system has been analyzed theoretically as a model for hydrogen bonding in proteins (35, 36). This system is an example of binding where both partners are small and rigid, so internal fluctuations should play a small role in the binding energetics. Association is driven by a strong inter-molecular electrostatic interaction between the NH and CO groups, opposed to a large extent by desolvation [Table 1; Ben-Tal *et al.* (35)], leading to weak net binding. As expected, internal fluctuations are only slightly reduced on binding, leading to a small 1 kcal/mol penalty. Unlike this earlier study, we have included an estimate of the R/T penalty on binding. Because the binding is dominated by a single hydrogen bond, effectively a one-point attachment, a large amount of residual rotational motion remains in the complex (Table 2), leading to a very small rotational immobilization penalty of <1 kcal/mol. The translational immobilization is somewhat greater. The sum of all these factors results in weak binding, in reasonable agreement with experiment, considering the small numbers involved and the relative uncertainties.

The third system we examined was the antibody FAB fragment 26-10-digoxigenin complex, an example of tight binding of a very rigid ligand. The major part of the ligand consists of four fused rings, with only one major conformational degree of freedom: the bond joining the lactone ring substituent at C17. Binding is driven by electrostatic interactions, which are primarily hydrogen bonds with the hydroxyl and lactone oxygens. Because these H-bonds are with neutral protein groups, the electrostatic desolvation is not large, and the net electrostatic interaction is favorable (Table 1). Concomitantly, there are significant penalties for R/T immobilization in this tight-fitting complex. Internal vibrational fluctuation contributions are small, because the ligand is rigid.

The final system we studied was streptavidin–biotin binding, a less rigid ligand than digoxigenin, but one of the tightest protein–

ligand associations known and thus a paradigm for high-affinity protein binding. This system has been extensively studied most notably by a combination of molecular dynamics, FDPB/surface area and normal mode analysis (21, 37), a hybrid method developed in part to calculate absolute binding constants. Binding is again driven by a strong intermolecular interaction, worth about -107 kcal/mol. Of this interaction, -62 kcal/mol is electrostatic, which is opposed by $+81$ kcal/mol of desolvation, so the net electrostatic contribution is unfavorable (Table 1), because of the desolvation of the ureido and carboxylate groups. The binding is driven by the strong nonpolar interactions composed of van der Waals and hydrophobic terms, reflecting the extremely good match of shape and hydrophobic interactions for this complex. Again there are significant penalties for R/T immobilization in this tight-fitting complex, although less than for the more rigid digoxigenin ligand. Biotin has a flexible valeryl carboxylate tail whose mobility is significantly reduced on binding, adding a 2.5-kcal/mol vibrational fluctuation penalty. The general balance of interactions in the Bjerrum model is very similar those found by FEP (21), although we cannot compare the R/T term because it was not separated out in that study. We find a very strong net binding free energy, in agreement with experiment, although the magnitude is somewhat too low. This underestimate may well be because the quasiharmonic method used to evaluate the Bjerrum integral is oversimplified for this system, particularly in representing loss of entropy from bond rotations. Here a method in which snapshots from an extended molecular dynamics simulation are postprocessed using the PB method, permits a larger region of a nonharmonic binding well, e.g., from bond rotomerization, to be explored (38).

Discussion

A concise derivation of a general equation for the binding constant and the enthalpy and entropy components (Eqs. **4**, **6**, and **7**) is one result of this paper. We refer to this as the generalized Bjerrum approach: applied to the association of structureless (spherical) ions and by using a Coulombic potential to approximate ω, it was first used by Bjerrum to derive his well known expression for the ion association constant (39). In this application, the limits of $H = 1$ are from $r = a$, the distance of closest approach of the two ions, to $r = b$, the distance cutoff for ion pairing introduced by Bjerrum. The Bjerrum expression for ion pairing is usually presented without detailed derivation, obscuring its origin in a more exact statistical mechanical expression. It can, however, be generalized, as shown here, to the binding of polyatomic molecules.

Comparison with Chemical Potential-Based Derivation. Apart from notational differences, our final expressions for binding constant and entropy are almost identical to those derived previously by using the chemical potential approach [equations 49 and 50 of Gilson *et al.* (1)], except we have dispensed with symmetry numbers, because equivalent positions/orientations are implicitly included in the full integrals over $\mathbf{r}/\mathbf{\Omega}$, and have omitted the negligible pressure–volume term. Indeed, agreement is expected if both approaches are valid. Eq. **4** is also related to the expression derived previously for membrane association by Ben-Tal *et al.* by using the chemical potential approach (2), with the differences that for membrane and surface adsorption, the bound molecule is not confined translationally in the plane of the surface, and the appropriate quantity is not an association constant but the surface excess, which is equivalent in the notation used here to $\Gamma = \int H(e^{-\beta\omega} - 1)$. However, the derivation of K_{AB} we present here is shorter and arguably clearer in its assumptions and implications than the standard derivation based on chemical potentials. For example, one simplification is that it involves just configurational (coordinate) integrals of a fixed number of particles. There is no introduction or removal of particles, as implied in the chemical potential approach, with the attendant momentum terms. Although these cancel to

leave a mass independent binding constant (as required in the classical approximation), the mass independence is clearer if they are not introduced in the first place. Furthermore, it is easier to see that Eq. **4** is exact in the dilute A/B limit even for nonideal polyelectrolyte solutes such as DNA. This is not so obvious in the chemical potential approach since polyelectrolytes have nonunity activity coefficients even at infinite dilution because of ionic atmosphere effects. Other implications of our approach are discussed in more detail below.

Delineation of the Bound State. It is clear that a rigorous derivation of the binding free energy requires explicit delineation between the bound and unbound states. For mathematical convenience, we adopt the Heaviside function notation ($H = 1/0$ delineates bound/unbound states) used previously for gas phase-monoatomic association (12); alternatively, one could specify (rather complicated) integration limits on $\mathbf{r}/\mathbf{\Omega}$ in the integral in Eq. **4**. Because one is ultimately comparing to an experimental measurement of K_{AB}, the ideal situation would be to make $H = 1$ correspond exactly to those configurations that give the "AB" signal with the experimental probe (spectroscopy, calorimetry, etc.) used to measure the binding constant. In practice, one rarely knows enough about the experimental method to do this. From a computational point of view, one often finesses the issue, as in the quasiharmonic example given here: Most of the contribution to the binding phase volume integral comes near the minimum of the potential well in $\mathbf{r}, \mathbf{\Omega}$ phase space, and the numerical value is rather insensitive to where one cuts off the bound-state integration. Nevertheless, some definition of AB is always involved. This problem does not appear to arise in the usual derivation using chemical potentials, because one assumes there is a distinct AB species as soon as its chemical, μ_{AB}, is introduced. However, this assumption is misleading, because it merely postpones or hides some definition of what AB is. Also μ_{AB} is not quite a standard chemical potential, which is the thermodynamic work of insertion into pure solvent at constants T and P: If we insert one "molecule" of AB into water at infinite dilution, it will immediately dissociate, so μ_{AB} involves some additional constraint to T and P that restricts A + B to exactly the region defining the complex.

Role of Solvent. A distinction between R/T contributions to binding and the role of the solvent is introduced right from the start. It is inherent in the definition of the complex and the introduction of the potential of mean force between A and B. Although there is some degree of freedom (and care required) in choice of coordinates (40), the R/T coordinates of B are usually defined with respect to either its center of mass and moments of inertia or its centroid and geometric moments (nonmass weighted, used in this work). All solvent effects, including excluded volume, are fully accounted for by the solvent-mediated potential of mean force between A and B. Thus in the Bjerrum derivation, the range of integration for B is $4\pi^2 V$, where V is the total solution volume. This range is fully consistent with the widely accepted definition of the solute chemical potential as the sum of the thermodynamic work of introducing a solute into a fixed position plus a "liberational" term from solute translation/rotation in the *entire* solution volume (14, 41). The Bjerrum approach also provides a rigorous definition of solvent and association entropy contributions without appeal to the gas-phase Sackur Tetrode expression for translational entropy (often invoked in the chemical potential type derivations), which has been criticized as unrealistic for binding in solution (4, 42).

Absolute Binding Free Energies. Calculation of absolute binding constants has been discussed previously by Gilson *et al.* (1) and others, and it remains a computational challenge. As defined by Eq. **5**, it is the free energy change when one molecule of B is taken from a state in which it can rotate and translate freely in a phase volume $8\pi^2 V_{ref}$ ($= 1,660$ Å^3 at 1 M) to the bound state, where it undergoes restricted R/T motion under the influence of its solvent mediated

intermolecular interactions with A, in the Boltzmann-weighted phase volume (BPV) where $H(\mathbf{r}, \mathbf{\Omega}) = 1$. Thus, the free energy is the log of a unitless quantity, the volume ratio $V_{BPV}/8\pi^2 V_{ref}$. The appearance of V_{BPV} spells out the requirement for any absolute binding free energy calculation: it must either (*i*) evaluate a phase volume integral, preferably over the entire bound state, $H = 1$, but at least over the regions at the bottom of the well that make the most contribution to the integral, (e.g., refs. 13, 43–45, and this work) or (*ii*) compute a binding free energy relative to some system whose V_{BPV} is known (46, 47). A computational protocol that uses only the difference in free energy between A, B, and AB at some point(s) in the $\mathbf{r}, \mathbf{\Omega}$ potential well does not give the absolute binding free energy. Such protocols include "single point" FDPB calculations, double disappearance (48), binding energy scoring schemes, and many other empirical binding methods. These methods provide either ω or its equivalent at a single point (usually the minimum) on the $\omega(\mathbf{r}, \mathbf{\Omega})$ surface, or the average $\omega(\mathbf{r}, \mathbf{\Omega})$ for some region about the minimum.

It should also be noted that when "single point" methods are used for relative binding, this is equivalent to the assumption that the shape of the $\omega(\mathbf{r}, \mathbf{\Omega})$ well is the same, and only its depth is changed. This assumption may not work well comparing systems of widely differing affinity, because tighter binding usually implies less residual motion in the complex, i.e., a narrower $\mathbf{r}, \mathbf{\Omega}$ potential well, in which case the integration over the bound phase volume must be evaluated.

Practical Calculation of Binding Free Energies. For many systems of interest to the biochemist, it would be prohibitive to model the solvent explicitly. In the generalized Bjerrum approach, the pmf provides a natural and well defined way to introduce an implicit or continuum solvent model into the calculations. We have illustrated this by using the FDPB/surface area implicit solvent model for four binding reactions, choosing systems with diverse characteristics: the ligands ranged from a monoatomic ion to a large rigid drug, from very tight to very weak binding, and from electrostatically dominated to apolar dominated. We chose these systems primarily to illustrate the Bjerrum approach and to examine the issue of association entropy, not as a detailed analysis of the most accurate way to do binding free energy computations. Again, we emphasize that there are alternative reasonable approximations to those used here. Nevertheless, the general agreement with experiment is satisfactory enough that we can draw several conclusions about the association entropy. The relevant data from our calculations are summarized in Table 2. First, it is clear that the association entropy contributions vary quite widely (1.5–7.5 kcal/mol at 298 K), depending on the nature and strength of the binding. The upper limit is similar to that for association of the rigid cyclic diglycine molecule (19). The lower limit, for the partially immobilized NMA, is similar to estimates of 1.3 kcal/mol for another partially immobilized case, a lipid in a membrane (2). The translational contribution varies, but it does not simply increase with strength of binding, being largest for the tricyclic cryptate, which completely encapsulates the ion. For this case, the residual motion is about 0.2 Å in the *x*, *y*, and *z* directions, which is small compared to the value of about 0.8 Å for immobilization of a sodium ion in the GA channel (13). The translational part for high-affinity complexes is significantly greater than that estimated from the cratic equation or recent tethering experiments (9, 10) (2.4 and 1.5 kcal/mol at 298 K, respectively). However, we argue that tethering, although undoubtedly producing an enhancement in association rate and complex stabilization, cannot be used to extract an estimate of ΔS_{trans} for the untethered system. To obtain ΔS_{tran}, we need to know how much residual motion of B with respect to A there is in the bound state (Eq. **9**). This residual motion is determined by the specific interactions between A and B that drive them to form a complex AB with and without the tether and so are subtracted out when the difference in binding entropy or free energy is taken. The contribution of rotational immobilization, which has been omitted in some previous treatments, including the cratic entropy model, is at least as important as translational immobilization, particularly for a ligand that is bound by multiple attachments. Overall, we find that the association entropy penalty is significant for high-affinity complexes but smaller than early estimates in the literature of >14 kcal/mol (at 298 K). How general this conclusion is will be revealed by calculations on other systems. For calculating this and other contributions to the binding free energy, the generalized Bjerrum approach provides a rigorous and self-consistent theoretical basis.

1. Gilson, M. K., Given, J. A., Bush, B. L. & McCammon, J. A. (1997) *Biophys. J.* **72,** 1047–1069.
2. Ben-Tal, N., Honig, B., Bagdassarian, C. K. & Ben-Shaul, A. (2000) *Biophys. J.* **79,** 1180–1187.
3. Reddy, M. R., Erion, M. D. & Agarwal, A. (2000) *Rev. Comput. Chem.* **16,** 217–304.
4. Amzel, L. M. (1997) *Proteins* **28,** 144–149.
5. Amzel, L. M. (2000) *Methods Enzymol.* **323,** 167–177.
6. Gohlke, H., Hendlich, M. & Klebe, G. (2000) *J. Mol. Biol.* **295,** 337–356.
7. Sham, Y. Y., Chu, Z. T., Tao, H. & Warshel, A. (2000) *Proteins Struct. Funct. Genet.* **39,** 393–407.
8. David, L., Luo, R. & Gilson, M. K. (2001) *J. Comput. Aided Mol. Des.* **15,** 157–171.
9. Tamura, A. & Privalov, P. L. (1997) *J. Mol. Biol.* **273,** 1048–1060.
10. Yu, Y. B., Privalov, P. L. & Hodges, R. S. (2001) *Biophys. J.* **81,** 1632–1642.
11. Hill, T. L. (1985) *Cooperativity Theory in Biochemistry* (Springer, New York).
12. Chandler, D. (1979) *Faraday Discuss. Chem. Soc.* **66,** 184–190.
13. Roux, B. (1999) *Biophys. J.* **77,** 139–153.
14. McQuarrie, D. (1976) *Statistical Mechanics* (Harper & Row, New York).
15. Head, M. S., Given, J. A. & Gilson, M. K. (1997) *J. Phys. Chem. A* **101,** 1609–1618.
16. Karplus, M. & Kushnick, J. (1981) *Macromolecules* **14,** 325–332.
17. Levy, R., Karplus, M., Kushnick, J. & Perahia, D. (1984) *Macromolecules* **17,** 1370–1374.
18. Hamermesh, M. (1974) *Group Theory and Its Application to Physical Problems* (Dover, New York).
19. Brady, G. P. & Sharp, K. A. (1997) *Biophys. J.* **72,** 913–927.
20. Sitkoff, D., Sharp, K. & Honig, B. (1994) *J. Phys. Chem.* **98,** 1978–1988.
21. Miyamoto, S. & Kollman, P. (1993) *Proteins Struct. Funct. Genet.* **16,** 226–245.
22. Metz, B., Rosalky, J. M. & Weiss, R. (1976) *Chem. Commun.* **14,** 533–534.
23. Jeffrey, P. D., Strong, R. K., Sieker, L. C., Chang, C. Y., Campbell, R. L., Petsko, G. A., Haber, E., Margolies, M. N. & Sheriff, S. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 10310–10314.
24. Weber, P., Ohlendorf, C., Wendoloski, J. & Salemme, F. (1989) *Science* **243,** 85–88.
25. Press, W., Flannery, B., Teukolsky, S. & Vetterling, W. (1986) *Numerical Recipes* (Cambridge Univ. Press, Cambridge, U.K.).
26. Sharp, K. A. (1998) *Proteins* **33,** 39–48.
27. Gallagher, K. & Sharp, K. A. (1998) *Biophys. J.* **75,** 769–776.
28. Rashin, A. A. & Honig, B. (1985) *J. Phys. Chem.* **89,** 5588–5593.
29. Sridharan, S., Nicholls, A. & Sharp, K. A. (1995) *J. Comput Chem.* **16,** 1038–1044.
30. Sharp, K. A., Jean-Charles, A. & Honig, B. (1992) *J. Phys. Chem.* **96,** 3822–3828.
31. Graf, E. & Lehn, J. M. (1976) *J. Am. Chem. Soc.* **98,** 6403–6405.
32. Lybrand, T. P., Ghosh, I. & McCammon, J. A. (1985) *J. Am. Chem. Soc.* **107,** 7793–7794.
33. Marcus, Y. (1994) *Biophys. Chem.* **51,** 111–128.
34. Kresheck, G. & Klotz, I. (1969) *Biochemistry* **8,** 8–13.
35. Ben-Tal, N., Sitkoff, D., Topol, I. A., Yang, A.-S., Burt, S. K. & Honig, B. (1997) *J. Phys. Chem. B* **101,** 450–457.
36. Simonson, T. & Brunger, A. (1994) *J. Phys. Chem.* **98,** 4683–4694.
37. Wang, J., Dixon, R. & Kollman, P. A. (1999) *Proteins Struct. Funct. Genet.* **34,** 69–81.
38. Kuhn, B. & Kollman, P. A. (2000) *J. Am. Chem. Soc.* **122,** 3909–3916.
39. Bjerrum, N. (1926) *Kgl. Danske Videnskab. Selskab.* **7,** 1–48.
40. Potter, M. J. & Gilson, M. K. (2002) *J. Phys. Chem.* **106,** 563–566.
41. Ben-Naim, A. (1987) *Solvation Thermodynamics* (Plenum, New York).
42. Mammen, M., Shakhnovich, E., Deutch, J. & Whitesides, G. (1998) *J. Org. Chem.* **63,** 3821–3830.
43. Cieplak, P. & Kollman, P. A. (1988) *J. Am. Chem. Soc.* **110,** 3734–3739.
44. Tidor, B. & Karplus, M. (1994) *J. Mol. Biol.* **238,** 405–414.
45. Luo, R. & Gilson, M. K. (2000) *J. Am. Chem. Soc.* **122,** 2934–2937.
46. Hermans, J. & Shankar, S. (1986) *Isr. J. Chem.*, 225–227.
47. Roux, B., Nina, M., Pomes, R. & Smith, J. C. (1996) *Biophys. J.* **71,** 670–681.
48. Jorgensen, W. L., Buckner, J. K., Boudon, S. & Tirado-Rives, J. (1988) *J. Chem. Phys.* **89,** 3742–3746.
49. Schildbach, J., Near, R., Bruccoleri, R., Haber, E., Jeffrey, P., Ng, S., Novotny, J., Sheriff, S. & Margolies, M. (1993) *J. Biol. Chem.* **268,** 21739–21747.
50. Weber, P., Wendoloski, J., Pantoliano, M. & Salemme, F. (1992) *J. Am. Chem. Soc.* **114,** 3197–3200.