

# The Statistical-Thermodynamic Basis for Computation of Binding Affinities: A Critical Review

Michael K. Gilson,\* James A. Given,\* Bruce L. Bush,# and J. Andrew McCammon<sup>§</sup>

\*Center for Advanced Research in Biotechnology, National Institute of Standards and Technology, Rockville, Maryland 20850-3479;

#Department of Molecular Design and Diversity, Merck Research Laboratories, Rahway, New Jersey 07065; and <sup>§</sup>Departments of Chemistry and Biochemistry, and Pharmacology, University of California at San Diego, La Jolla, California 92093-0365 USA

**ABSTRACT** Although the statistical thermodynamics of noncovalent binding has been considered in a number of theoretical papers, few methods of computing binding affinities are derived explicitly from this underlying theory. This has contributed to uncertainty and controversy in certain areas. This article therefore reviews and extends the connections of some important computational methods with the underlying statistical thermodynamics. A derivation of the standard free energy of binding forms the basis of this review. This derivation should be useful in formulating novel computational methods for predicting binding affinities. It also permits several important points to be established. For example, it is found that the double-annihilation method of computing binding energy does not yield the standard free energy of binding, but can be modified to yield this quantity. The derivation also makes it possible to define clearly the changes in translational, rotational, configurational, and solvent entropy upon binding. It is argued that molecular mass has a negligible effect upon the standard free energy of binding for biomolecular systems, and that the cratic entropy defined by Gurney is not a useful concept. In addition, the use of continuum models of the solvent in binding calculations is reviewed, and a formalism is presented for incorporating a limited number of solvent molecules explicitly.

## INTRODUCTION

The noncovalent association of molecules is of central importance in biology and pharmacology. It underlies the action of hormones, the control of DNA transcription, the recognition of antigens by the immune system, the catalysis of chemical reactions by enzymes, and the actions of many drugs. Therefore, methods of predicting the affinity of such noncovalent associations would be of great practical value. For example, predictive computational models for the noncovalent association of biomolecules and their ligands would be useful in structure-based drug design and in the redesign of enzymes. As recently reviewed (Cohen et al., 1990; Kuntz et al., 1994; Balbes et al., 1994; Colman, 1994; Greer et al., 1994; Rosenfeld et al., 1995; Ajay and Murcko, 1995; Marrone et al., 1997), many groups have invested ingenuity and effort in the development of such models. However, although the physical chemistry of these association processes is rooted in statistical thermodynamics, few models are explicitly derived from these underlying principles. This has led to a certain amount of confusion in the field, as discussed below. The present paper therefore presents a statistical-thermodynamic derivation of the standard free energy of binding, and uses this to review the underpinnings of methods for computing noncovalent binding affinities.

There exists an extensive literature on the statistical thermodynamics of association. The first important attempts to characterize binding equilibria were mean field theories, which were originally developed to treat ion pairing in electrolytes (see, e.g., Pitzer and Schreiber, 1987; Laria et al., 1990). Mean field theories depend upon a small number of parameters that must be determined self-consistently. They perform well for cases where specific terms in the expression for enthalpy, such as Coulombic interactions, provide the dominant contribution to the binding free energy.

Modern liquid-state methods are based upon a systematic perturbation expansion of the free energy of binding. Such methods include both entropic and enthalpic contributions to the binding energy in a manner that is, in principle, exact. A formal extension of these methods to treat protein-protein and protein-ligand association is presented in Hill (1955a,b). The computation required to explicitly evaluate the resulting formulas is, however, daunting. The association of simple ligands can be studied with such methods by resummation of the terms corresponding to strongly attracting interactions (Stell, 1976; Andersen, 1973; Wertheim, 1984). Association between larger molecules can be treated in closed form, either exactly or approximately, by the use of correlation function methods (Chandler and Pratt, 1976; Stell, 1976; Andersen, 1973; Hoyer and Olausson, 1980; Wertheim, 1984). Such methods have been applied to a range of systems that contain a small number of different types of atoms (see, e.g., Chandler and Pratt, 1976; Hoyer and Olausson, 1980; Wertheim, 1984; Stell and Zhou, 1989; Zhou and Stell, 1995). However, for more complex systems, the number of distinct correlation functions that must be evaluated becomes prohibitively large. In particular, the binding of biological molecules has not yet been treated in any detail by liquid-state perturbation methods.

Received for publication 13 May 1996 and in final form 21 November 1996.

Address reprint requests to Dr. Michael K. Gilson, Center for Advanced Research in Biotechnology, 9600 Gudelsky Dr., Rockville, MD 20850-3479. Tel.: 301-738-6217; Fax: 301-738-6255; E-mail: gilson@indigo14.carb.nist.gov.

© 1997 by the Biophysical Society

0006-3495/97/03/1047/23 \$2.00

An alternative to perturbation expansions, and the exhaustive enumeration of configurations that they imply, is to use molecular dynamics or Monte Carlo simulations to sample the commonly occurring low-energy structures that give the dominant contribution to thermodynamic quantities. Such simulations typically include an explicit representation of the atoms of both the solutes and the solvent. These sampling methods may be used with thermodynamic integration or free energy perturbation methods to compute the free energy changes associated with binding. Such free energy simulation methods offer a microscopic picture of solute and solvent structure. They have been used to compute relative free energies for the binding of different ligands to a common receptor (Tembe and McCammon, 1984), as well as the binding free energy for a single ligand and receptor (Jorgensen et al., 1988). Excellent reviews of this material are available (Wong and McCammon, 1986; Bash et al., 1987; Beveridge and DiCapua, 1989; Lybrand, 1990; Straatsma and McCammon, 1992; Kollman, 1993; Warshel et al., 1994; Marrone et al., 1997). Free energy simulations are time-consuming, and it is not always clear that the results they yield are fully converged (Mitchell and McCammon, 1991; Balbes et al., 1994). These methods are likely to be most useful in the later stages of projects in molecular design, when the goal is to optimize promising lead compounds.

Another class of models of binding that is applicable to biomolecules avoids the problem of conformational sampling almost entirely. Instead, binding free energies are estimated by the use of phenomenological scoring functions applied to rigid or nearly rigid representations of the molecules that bind (see, e.g., Gurney, 1953; Chothia and Janin, 1975; Andrews et al., 1984; Erickson, 1989; Novotny et al., 1989; Searle et al., 1992a; Horton and Lewis, 1992; Murphy et al., 1993; Böhm, 1994; Weng et al., 1996). These scoring functions take into account a number of distinct contributions to the binding free energy. For example, stabilizing interactions may result from intermolecular hydrogen bonds and hydrophobic contacts, while destabilizing interactions typically include the entropic costs of restricting the relative position and orientation of the two ligands. Such "energy-component" models offer useful intuitions and are computationally tractable. On the other hand, they may be less accurate than the more detailed free energy simulations described above. Energy-component models are likely to be most useful in the early, exploratory stages of projects in molecular design.

As noted above, these and other computational models of biomolecular binding are, in principle, founded in the statistical thermodynamic of condensed phases. However, a number of models and methods in current use have been presented without an explicit statistical-thermodynamic derivation. In such cases, it is often difficult to discern what approximations are being made, or indeed whether the model is valid. Largely as a consequence, some uncertainties in this area have risen to the level of controversy. Notable examples include the computation of entropy

changes upon binding by a formula for the entropy of mixing (Kauzmann, 1959; Murphy et al., 1994; Holtzer, 1995), and the idea of an "absolute" binding free energy that does not depend upon the standard concentration (Jorgensen et al., 1988; Janin, 1996). Indeed, a recent paper appropriately notes that there is "considerable confusion" regarding the theory underlying the calculation of binding constants in solution (Tidor and Karplus, 1994). It is perhaps surprising that there should be confusion on this topic whose principles have already been laid out in the theoretical literature (see above). The problem appears to be that the relevant papers are extremely general in scope, so that subtle technical points must be resolved to apply them to biophysical problems.

We therefore believe that many uncertainties in this area can be resolved, and new controversies avoided, by an accessible treatment of the statistical thermodynamics underlying models of binding. The present paper provides an accessible derivation of the standard free energy of binding in terms of molecular properties, using elements gathered from existing publications. This derivation is then used as the basis for an examination of several important issues in existing models of binding. A complete review of methods for computing binding constants is beyond the scope of this work, however.

The paper is organized as follows. 'Statistical Thermodynamics of Binding' reviews the thermodynamics and the classical statistical thermodynamics of noncovalent association. The derivation of the standard free energy of binding is for a system at constant pressure, because most experiments are carried out under this condition. Particular attention is paid to the often-confusing issue of the standard state. 'Free Energy Simulations with Explicit Solvent' examines the theoretical underpinnings of calculations of relative and absolute binding free energies. The conditions under which parts of the protein may be treated as rigid also are defined. 'Other Representations of the Solvent' reviews the theoretical basis for implicit representations of the solvent in binding calculations, and also presents new material on hybrid implicit/explicit treatments of the solvent. Finally, 'Entropy and Energy Components of the Binding Free Energy' analyzes the statistical-thermodynamic basis of energy-component models for binding (see above). The literature on entropy changes associated with binding is reviewed, and the theory presented in the second section is used to define and discuss changes in translational, rotational, configurational, and solvent entropy. This section also analyzes the influence of molecular mass upon binding free energies.

## STATISTICAL THERMODYNAMICS OF BINDING

This section reviews the formula for a binding constant in terms of the standard chemical potentials of the molecular species involved. The concepts of the standard state and of activity coefficients are discussed in detail. Then classical statistical thermodynamics is used to derive expressions for

the standard chemical potentials of the two ligands and the bound complex. These expressions are combined to yield an expression for the standard free energy of binding in solution.

### Thermodynamics and Standard States

The reaction of interest is the noncovalent association of two ligands,  $A$  and  $B$ , to form a complex,  $AB$ :



In cases of biological interest, this reaction typically occurs in a mixed solvent, usually an aqueous electrolyte. This will be called the reaction solvent. The condition for equilibrium is that

$$\mu_{\text{sol},A} + \mu_{\text{sol},B} = \mu_{\text{sol},AB}, \quad (2)$$

where each  $\mu_{\text{sol},i}$  is the chemical potential of species  $i = A, B$ , or  $AB$ , in solution.

The chemical potential of species  $i$  in solution is given by

$$\mu_{\text{sol},i} = \mu_{\text{sol},i}^{\circ} + RT \ln \frac{\gamma_i C_i}{C^{\circ}}, \quad (3)$$

where  $\mu_{\text{sol},i}^{\circ}$  and  $C_i$  are, respectively, the standard chemical potential and the concentration of species  $i$ ,  $R$  is the gas constant,  $T$  is absolute temperature,  $\gamma_i$  is the activity coefficient of  $i$ , and  $C^{\circ}$  is the standard concentration in the same units as  $C_i$ . For example, if  $C_i$  is expressed in molecules/Å<sup>3</sup>, as may be convenient in discussing a molecular simulation, then the usual 1 M standard concentration must be expressed as  $C^{\circ} = 1 \text{ molecule}/1660 \text{ Å}^3$ . However, it is more common to express  $C$  in molar units (M), in which case  $C^{\circ} = 1 \text{ M}$ . Often,  $C^{\circ}$  is not included explicitly in Eq. 3. In such cases, it is implicit that the units of concentration,  $C_i$ , are “standard concentrations”; that is,  $C_i$  is then the ratio of the concentration to the standard concentration. Although concentrations are sometimes reported as mole fractions, in the present paper concentration will always imply a density with dimension volume<sup>-3</sup>. This point is discussed further in the last section.

The standard chemical potential of  $A$ ,  $B$ , or  $AB$  is its chemical potential in its standard state. It is convenient to define a hypothetical standard state in which each species is at standard concentration in the reaction solvent, defined above, but does not interact with other molecules of  $A$ ,  $B$ , or  $AB$ . The activity coefficients relative to this standard state,  $\gamma_i$ , approach unity as  $C_i$  approaches zero in the reaction solvent (Glasstone, 1947; Lewis and Randall, 1961a).

Equations 2 and 3 yield the following expression for the standard free energy of binding:

$$\begin{aligned} \Delta G_{AB}^{\circ} &\equiv \mu_{\text{sol},AB}^{\circ} - \mu_{\text{sol},A}^{\circ} - \mu_{\text{sol},B}^{\circ} \\ &= -RT \ln \left( \frac{\gamma_{AB} C^{\circ} C_{AB}}{\gamma_A \gamma_B C_A C_B} \right)_{\text{eq}} \equiv -RT \ln K_{AB}, \end{aligned} \quad (4)$$

where  $( )_{\text{eq}}$  implies a quantity at equilibrium. In this paper, attention is restricted to the low concentration case, for which  $\gamma_i = 1$  is a good approximation.

Because  $C^{\circ}$  is not always written explicitly in the expression for the binding constant  $K_{AB}$ , it might appear that  $K_{AB}$  has units of volume. It is therefore worth reemphasizing that  $K_{AB}$  is a dimensionless quantity (Alberty, 1994; Janin, 1995; Atkins, 1994). When  $C^{\circ}$  is not written explicitly, it is implicit that the units of concentration are “standard concentrations.” This occurs naturally when concentrations are in M and the standard concentration is 1 M. This issue is discussed further in Calculation of the ‘Standard Free Energy of Binding.’ Note, too, that  $C_{AB}$  depends slightly upon the geometric criteria used to define the complex, as discussed in ‘The Standard Chemical Potential of a Complex in Solution.’

### The Standard Chemical Potential of a Molecule in Solution

The standard chemical potential  $(\partial G/\partial n_A)_{T,P}$  of molecule  $A$  in solution is given by:

$$\mu_{\text{sol},A}^{\circ} = -RT \ln \left( \frac{1}{V_{N,A} C^{\circ}} \frac{Q_{N,A}(V_{N,A})}{Q_{N,0}(V_{N,0})} \right) + P^{\circ} \bar{V}_A. \quad (5)$$

A detailed derivation of this expression (Hill, 1985a), and related material (McMillan and Mayer, 1945; Hill, 1986), may be found elsewhere. Here  $Q_{N,A}(V_{N,A})$  is the canonical partition function for a system containing a large number  $N$  of solvent molecules and one solute molecule  $A$  at volume  $V_{N,A}$ .  $V_{N,A}$  is the volume of this system when it is at equilibrium at standard pressure  $P^{\circ}$ , typically one atmosphere. Similarly,  $Q_{N,0}(V_{N,0})$  is the canonical partition function for the  $N$  solvent molecules without the solute, now at a different equilibrium volume  $V_{N,0}$  that also corresponds to pressure  $P^{\circ}$ . In the rest of this paper, the volume for which each canonical partition function,  $Q$ , is to be evaluated will not be written explicitly, as it is in Eq. 5. Rather, it will be implicit that the volume is the equilibrium volume of the system for the specified pressure, unless otherwise noted. Finally,  $\bar{V}_A \equiv V_{N,A} - V_{N,0}$ . This is the change in the equilibrium volume when one molecule of solute is added to  $N$  molecules of solvent. Therefore, for  $N \gg 1$ ,  $\bar{V}_A$  is the partial molar volume of the solute at infinite dilution in the solvent (Hill, 1985a). The pressure-volume work associated with this volume change is  $P^{\circ} \bar{V}_A$ . The present expression for  $\mu_{\text{sol},A}^{\circ}$  can be interpreted as the standard chemical potential of the solute in the gas phase, plus the work of transferring it to the solvent isobarically. It is worth mentioning that, although Eq. 5 includes the standard concentration explicitly, the standard concentration is only implicit in the cited formula (Hill, 1985a). Agreement with the cited formula is achieved when the units of concentration are consistent with those of volume. For example, in MKS units, a 1 M standard concentration corresponds to  $C^{\circ} = 1000 N_{\text{sol}} \text{ m}^{-3}$ , where  $N_{\text{sol}}$  is Avogadro’s number.

Although the formulation of Eq. 5 differs from that of Widom (Widom, 1963, 1982), the expressions are equivalent in the thermodynamic limit (see Appendix B). The present form is convenient here because it makes the influence of pressure explicit. Note also that the pressure-volume term here is distinct from the volume-related contribution to solvation energy that may be calculated with Flory-Huggins theory (Kumar et al., 1995). The pressure-volume term found in Equation 5 is typically very small at standard pressure, because  $\bar{V}_A$  is typically small (Ben-Naim, 1992). In contrast, the Flory-Huggins contribution may be large, even when the partial molar volume of the solute is small.

We now proceed to simplify the ratio of partition functions on the right-hand side of Eq. 5. To do this, we require an expression for the energy of the system in terms of conjugate momenta and coordinates:

$$H(\mathbf{p}_A, \mathbf{p}_S, \mathbf{r}'_A, \mathbf{r}_S) = \sum_{i=1}^{M_A+M_S} \frac{p_i^2}{2m_i} + U(\mathbf{r}'_A, \mathbf{r}_S), \quad (6)$$

where  $M_A$  and  $M_S$  are respectively the numbers of atoms of the one solute molecule and the  $N$  solvent molecules;  $\mathbf{p}$  is a vector of the momenta of the  $M_A + M_S$  atoms;  $p_i^2$  is the squared magnitude of the momentum of atom  $i$ ;  $m_i$  is the mass of atom  $i$ ; and  $U(\mathbf{r}'_A, \mathbf{r}_S)$  is the potential energy as a function of all the atomic coordinates. If it is assumed that classical statistical thermodynamics is applicable (see discussion in last section), then (McQuarrie, 1973; Chandler and Pratt, 1976)

$$\frac{Q_{N,A}}{Q_{N,0}} = \frac{\int d\mathbf{p}_A d\mathbf{p}_S \int d\mathbf{r}'_A d\mathbf{r}_S e^{-\beta \left( \sum_{i=1}^{M_A+M_S} \frac{p_i^2}{2m_i} + U(\mathbf{r}'_A, \mathbf{r}_S) \right)}}{\sigma_A \int d\mathbf{p}_S \int d\mathbf{r}_S e^{-\beta \left( \sum_{i=M_A+1}^{M_A+M_S} \frac{p_i^2}{2m_i} + U(\mathbf{r}_S) \right)}}, \quad (7)$$

where  $\beta \equiv (RT)^{-1}$ ,  $\sigma_A$  is the symmetry number of the solute, and the symmetry numbers of the solvent molecules have cancelled. Atoms  $i = 1, \dots, M_A$  belong to the solute, and atoms  $i = M_A + 1 \dots M_A + M_S$  belong to the solvent. The integral over each momentum component extends from  $-\infty$  to  $\infty$ , and the position integrals range over all configurations that are consistent with molecules being intact and within their container. Equation 7 neglects a prefactor that does not contribute to the calculation of binding constants.

We now establish a molecular axis system that allows the lab-frame coordinates of the solute atoms,  $\mathbf{r}'_A$ , in Eq. 7 to be separated into internal and external coordinates. Any three atoms may be used to define the molecular axes. These atoms may be numbered 1, 2, and 3 without loss of generality. Atom 1 becomes the origin of the molecular coordinates. The vector joining atom 1 with atom 2 defines the  $x$ -axis. The direction of the  $y$ -axis is given by the direction of the vector joining atoms 2 and 3, minus the  $x$ -component of this vector. The  $z$ -axis is constructed as the cross-product of the  $x$ - and  $y$ -axes. The Cartesian internal coordinates of

each atom may then be specified relative to this molecular frame of reference. Note that, in internal coordinates, atom 1 is fixed at the origin; atom 2 always lies on the  $x$ -axis; and atom 3 lies in the  $z = 0$  plane. The six coordinates thus fixed correspond to the external coordinates of the molecule. Other definitions of the molecular frame are also possible. The set of  $3M_A - 6$  internal Cartesian coordinates will be termed  $\mathbf{r}_A$ . The position of the molecular frame—really the position of atom 1—will be termed  $\mathbf{R}_A$ . The three Eulerian angles that specify the orientation of the molecular frame relative to the lab frame—really the orientation of atoms 1, 2, and 3—will be termed  $\xi_{A,1}, \xi_{A,2}, \xi_{A,3}$ . The complete set of external coordinates will be termed  $\zeta_A \equiv (\mathbf{R}_A, \xi_{A,1}, \xi_{A,2}, \xi_{A,3})$ .

The integrals over the internal coordinates of the solute and over the coordinates of the solvent do not depend upon the position or orientation of the solute; viz., upon  $\zeta_A$ . Therefore, the integrals over  $\zeta_A$  may be carried out at once (Steinberg and Scheraga, 1963), yielding a factor of  $8\pi^2 V_{N,A}$ . In the classical approximation, the integral over momentum for each atom  $i$  yields a factor of  $(2\pi m_i RT)^{3/2}$ . The momentum integrals for the solvent atoms cancel in Eq. 7, leaving only the momentum integrals for the solute atoms. Therefore, from Eqs. 5 and 7, the standard chemical potential of species  $A$  may be written as:

$$\mu^\circ_{\text{sol},A} = -RT \ln \left( \frac{8\pi^2}{C^\circ \sigma_A} \prod_{i=1}^{M_A} (2\pi m_i RT)^{3/2} \frac{Z_{N,A}}{Z_{N,0}} \right) + P^\circ \bar{V}_A; \quad (8)$$

$$Z_{N,A} \equiv \int e^{-\beta U(\mathbf{r}_A, \mathbf{r}_S)} d\mathbf{r}_A d\mathbf{r}_S; \quad (9)$$

$$Z_{N,0} \equiv \int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S. \quad (10)$$

Here,  $Z_{N,A}$  and  $Z_{N,0}$  are configuration integrals.

### The Standard Chemical Potential of a Complex in Solution

Although Eqs. 8–10 apply directly to molecules  $A$  and  $B$ , two complications arise for the complex,  $AB$ . First, it is necessary to define the external and internal coordinates of the complex. This may be accomplished by using the external coordinates of molecule  $A$  as external coordinates of the complex. Then *the external coordinates of  $B$ ,  $\zeta_B$ , are taken to be defined relative to molecule  $A$ , so that the external coordinates of  $B$  become internal coordinates of the complex.* For simplicity of notation, these will still be called  $\zeta_B \equiv (\mathbf{R}_B, \xi_{B,1}, \xi_{B,2}, \xi_{B,3})$ .

Second, it is clear that configurations for which the two ligands are far apart should not be included in the configuration integral of the complex. For most cases of interest, this problem may be dealt with by restricting the configu-

ration integral of the complex to configurations for which the two molecules are, in fact, complexed (Hill, 1985b; Chandler and Pratt, 1976). Mathematically, this restriction may be implemented by including in the configuration integral a step function  $I(\zeta_B)$  that equals unity for complexed configurations and zero otherwise.

Thus, the standard chemical potential of the complex is

$$\mu_{\text{sol,AB}}^{\circ} = -RT \ln \left( \frac{8\pi^2}{C^{\circ} \sigma_{\text{AB}}} \prod_{i=1}^{M_A+M_B} (2\pi m_i RT)^{3/2} \frac{Z_{\text{N,AB}}}{Z_{\text{N,0}}} \right) + P^{\circ} \bar{V}_{\text{AB}}; \quad (11)$$

$$Z_{\text{N,AB}} \equiv \int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_A d\mathbf{r}_B d\zeta_B d\mathbf{r}_S. \quad (12)$$

Here  $M_{\text{AB}} = M_A + M_B$  is the number of atoms in both solutes and  $J_{\zeta_B}$  is the Jacobian determinant for the Eulerian rotation of molecule  $B$  relative to  $A$ .

It is now necessary to consider how the complex is to be defined for computational purposes; that is, to specify the range over which  $I(\zeta_B) = 1$ . This is particularly simple when  $B$  binds in a cavity of  $A$ . More generally, defining  $I(\zeta_B)$  is straightforward whenever the potential of mean force (defined in the last section) for the interaction of  $A$  and  $B$  is sharply peaked and negative in a small range of  $\zeta_B$ . Then the stable configurations of the complex will dominate the thermodynamic averages over a zone of configuration space in which  $A$  and  $B$  are close together. Under these circumstances,  $\mu_{\text{sol,AB}}^{\circ}$  will be insensitive to the precise range of configurations for which  $I(\zeta_B) = 1$ . The region in which  $I(\zeta_B)$  equals one must satisfy the following two requirements (Hill, 1955a): 1) The region should include all configurations contributing significantly to the chemical potential of the complexed state; i.e., those for which the Boltzmann factor of the potential of mean force (see last section) is large. 2) The region should not include so large a phase volume of uncomplexed configurations that these contribute appreciably to  $\mu_{\text{sol,AB}}^{\circ}$ .

It should also be pointed out that experimental measurements of the binding constant  $K_{\text{AB}}$  are based upon a two-state model in which solutes  $A$  and  $B$  are either complexed or not. For this to be a good approximation, binding must be strongly detected by a signal that changes when  $A$  and  $B$  form a complex. Also, the signal must be negligible when  $B$  is beyond the region around  $A$  dominated by the stable states of the complex.

The following idealized case illustrates the insensitivity of  $\mu_{\text{sol,AB}}^{\circ}$  for a tight complex to the range over which  $I(\zeta_B) = 1$ . Molecules  $A$  and  $B$  are taken to be structureless and spherically symmetric. They bind each other with a square-well potential of mean force that equals 0 at separations greater than 4 Å. Between 2 and 4 Å, there is an energy well of fixed depth  $w_b$ . The energy becomes infinite (hard wall) at distances of less than 2 Å. The step function  $I(\zeta_B)$  is set equal to 1 for intermolecular distances less than some value

$r_1$ , and zero otherwise. The sensitivity of the standard chemical potential of the complex to the value of  $r_1$  is examined in Fig. 1. This plots the standard chemical potential of the  $AB$  complex as a function of  $r_1$ , for several different depths of the well,  $w_b$ . The figure shows that the standard chemical potential of the complex is insensitive to  $r_1$ , so long as two conditions are met. First, the binding must be reasonably tight; here, this occurs for  $w_b < 5$  kJ/mol. Second, the entire energy well must be included in the integral; this is satisfied here when  $r_1 > 4$  Å. The same principles are expected to hold for the association of real molecules. In fact, the independence of the chemical potential upon the precise volume of integration might be viewed as the defining characteristic of a complex.

For loosely bound complexes, it becomes more difficult to define the complex (Lewis and Randall, 1961b). In such cases, a probabilistic definition of a bound complex may be useful (Weist and Glandt, 1994); here the indicator function,  $I(\zeta_B)$ , is generalized from a step function to an arbitrary positive decreasing function. Another treatment of loosely bound complexes assesses complexation through the reduction it causes in the computed osmotic pressure (Groot, 1992).

For completeness, it is worth considering whether the binding of large molecules in solution may be treated by the concept of "bound" states, which is used in the theory of imperfect gases (Hill, 1956, 1986). A bound state of two molecules is one in which their relative kinetic energy is less than the work required to separate them. Bound states are expected to be long-lived in the gas phase, where kinetic energy is exchanged relatively slowly. However, in solution, the relative kinetic energy of solutes is likely to fluctuate greatly because of collisions with the dense solvent.

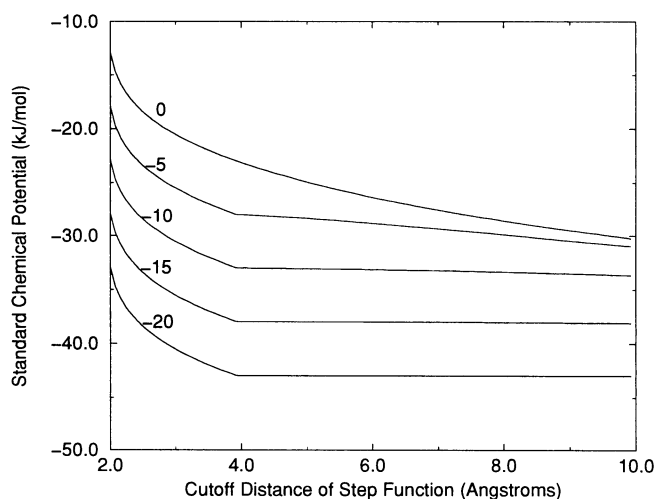


FIGURE 1 The standard chemical potential of a hypothetical complex of two spherically symmetric, structureless molecules that bind with a square-well potential, as a function of the radius of integration,  $r_1$  (see text). The numbers indicate the depth, in kJ/mol, of the binding potential. The results neglect a momentum term, but include a contribution of  $-RT \ln 8\pi^2/\sigma C^{\circ}$ , where  $C^{\circ} = 1660 \text{Å}^{-3}$ , and  $\sigma = 2$  for this symmetric complex.

As a consequence, the lifetime of a complex probably is not directly related to its instantaneous kinetic energy. Therefore, the concept of a kinetic bound state does not appear to be useful in this context.

### The Standard Free Energy of Binding

The expressions derived above for the standard chemical potentials of  $A$ ,  $B$ , and  $AB$  in solution permit Eq. 4 for the standard free energy change of binding to be written in terms of molecular properties:

$$\Delta G_{AB}^{\circ} = -RT \ln \left( \frac{C^{\circ}}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \frac{Z_{N,AB} Z_{N,0}}{Z_{N,A} Z_{N,B}} \right) + P^{\circ} \Delta \bar{V}_{AB}, \quad (13)$$

where  $\Delta \bar{V}_{AB} \equiv \bar{V}_{AB} - \bar{V}_A - \bar{V}_B$ . Note that all mass-dependent terms have now cancelled. A similar cancellation of mass terms was also observed in the context of the insertion of helices in membrane bilayers (Ben-Shaul et al., 1996). The issue of mass-dependence is discussed in the last section.

### FREE ENERGY SIMULATIONS WITH EXPLICIT SOLVENT

The derivation in the previous section is now used to review the statistical mechanical basis for the calculation of binding free energies via free energy simulations with an explicit treatment of the solvent. In these methods, some change in the system is divided into a series of stepwise perturbations. A molecular dynamics or Monte Carlo calculation is used to evaluate the work of taking each step. The sum of these work terms equals the overall change in free energy for the change in the system.

The first subsection here briefly considers the calculation of the relative binding constants of two different ligands  $B$  and  $C$  for the same receptor  $A$ . The second subsection considers in detail the calculation of the standard free energy of binding for a single ligand-receptor pair,  $A$  and  $B$ . The third subsection considers the circumstances under which parts of the protein may be treated as rigid in calculations of binding free energy.

#### Calculation of Relative Binding Free Energies

From Eq. 13, the difference between the binding free energy of two different ligands  $B$  and  $C$  for the same receptor  $A$  is

$$\begin{aligned} \Delta G_{AC}^{\circ} - \Delta G_{AB}^{\circ} = & -RT \ln \left( \frac{\sigma_{AB} Z_{N,AC}}{\sigma_{AC} Z_{N,AB}} \right) + P^{\circ} (\bar{V}_{AC} - \bar{V}_{AB}) \\ & - \left[ -RT \ln \left( \frac{\sigma_B Z_{N,C}}{\sigma_C Z_{N,B}} \right) + P^{\circ} (\bar{V}_C - \bar{V}_B) \right] \end{aligned} \quad (14)$$

The two pairs of terms on the right-hand side of Eq. 14 represent respectively the work of the "alchemical" trans-

formations of  $B$  to  $C$  in the receptor  $A$  and the work of the same transformation in solution (Tembe and McCammon, 1984). The use of multiconfiguration thermodynamic integration and free energy perturbation methods to compute these quantities has been discussed in detail elsewhere (Tembe and McCammon, 1984; Wong and McCammon, 1986; Bash et al., 1987; Beveridge and DiCapua, 1989; Lybrand, 1990; Straatsma and McCammon, 1992; Warshel et al., 1994), and is not reviewed further here. It is worth pointing out, however, that the two volume changes in parentheses in this equation are the changes in the equilibrium volume when  $B$  is mutated to  $C$  in the binding site of  $A$ , and in solution, respectively. These volume changes are small, so the associated pressure-volume work is negligible at standard pressure. However, these contributions become significant at pressures  $> \sim 100$  atmospheres, found in ultracentrifuge cells and in deep ocean (Schade et al., 1980a,b; Muller et al., 1981; Morild, 1981; van Eldik et al., 1989; Gross and Jaenicke, 1994). Calculations of binding energies at high pressure should therefore include these contributions.

#### Calculation of the Standard Free Energy of Binding

Although relative binding affinities suffice for the understanding of many biochemical processes, it would clearly be of interest to compute the free energy of binding for a single ligand and receptor. Such a free energy calculation was used in 1986 by Jan Hermans and Shankar Subramaniam to compute the affinity of a cavity in myoglobin for an atom of xenon (Hermans and Shankar, 1986). In that calculation, the affinity was computed as the sum of two terms. The first term was the work of reversibly replacing the interactions of the xenon atom with the protein by an artificial harmonic potential well that effectively trapped the xenon atom. The second term was the work done when the pressure of the trapped xenon was changed reversibly to a standard value of 1 atm. At 27 kJ/mol, the second term was not a negligible quantity. As noted in the paper, the second term is required if the simulation is to yield a standard free energy; i.e., a free energy referenced to a well-defined standard state. A closely related method has been used to compute the affinity of water for a protein cavity (Zhang and Hermans, 1996). These calculations for xenon and water were presented without a derivation. However, a very recent paper elucidates the statistical thermodynamic basis for such free energy calculations of the affinity of small molecules for specific sites inside a macromolecule, and applies the resulting formulas to water molecules in protein cavities (Roux et al., 1996).

At the same time as work on the binding of small molecules in protein cavities has progressed, a related set of approaches has evolved for computing the free energy of binding of larger molecules in protein binding sites. In 1988, the "double-annihilation" method (Jorgensen et al., 1988) was proposed as a way of computing the standard free

energy of binding for molecules in solution, and it has since been used in a number of studies (Jorgensen et al., 1988; Sneddon et al., 1989; Pranata and Jorgensen, 1991; Merz, Jr., 1991; Miyamoto and Kollman, 1992, 1993a and b; Mordasini Denti et al., 1996). The double-annihilation method resembles the method used for xenon and myoglobin, but it lacks the second term that corrects for the standard state. Moreover, it has been argued that the double-annihilation method yields binding free energies that do not depend upon the choice of standard concentration, and that it therefore violates the law of mass action (Janin, 1996). This is because the standard free energy of a reaction that replaces two molecules by one complex must depend upon the standard concentration, as reflected in the present expression for the binding free energy, Eq. 13. In contrast, relative binding free energies (previous section) and solvation free energies do not depend upon standard concentration, because these correspond to processes with fixed numbers of solute species.

The idea that the standard free energy of binding is independent of the standard concentration might appear to be supported by equations for the binding constant that do not explicitly include standard concentration (Prue, 1969; Justice and Justice, 1976; Chandler, 1979; Shoup and Szabo, 1982; Pranata and Jorgensen, 1991). For example, the following equation (Pranata and Jorgensen, 1991) relates the “association constant”  $K_a$  to a potential of mean force  $w(r)$  (see ‘Entropy External, Internal, and Solvent’) that depends upon the solute-solute distance  $r$ :

$$K_a = 4\pi \int_0^c r^2 e^{-\beta w(r)} dr. \quad (15)$$

Here  $c$  is a “cutoff limit . . . that defines association” (Pranata and Jorgensen, 1991). However, it can be shown that equations of this type implicitly use a standard concentration of 1 molecule/ $d^3$ , where  $d$  is the unit of distance in the integrals. Also, some equations yield not the binding constant, but the ratio of the binding constant to that which would be obtained for an ideal gas (Chandler, 1979). Such ratios of binding constants do not depend upon standard concentration.

This subsection examines the theoretical basis of the double-annihilation method, and shows that it is indeed incomplete. However, the double-annihilation method is a very valuable step toward a method for directly computing the standard free energy of binding. Therefore, the double-annihilation formalism is the starting point for the development of such a method, which is presented next. The present “double-decoupling” method is consistent with and somewhat more general than existing methods of computing the affinities of small molecules for specific sites in a macromolecule (Hermans and Shankar, 1986; Zhang and Hermans, 1996; Roux et al., 1996). Finally, published calculations with the double-annihilation method are discussed in light of the present analysis.

*A double-decoupling method of computing the standard free energy of binding*

The thermodynamic analysis that underlies the double-annihilation method is summarized in Fig. 2, which is from Pranata and Jorgensen (1991) with minor changes in notation. The top reaction is the “annihilation” of the ligand,  $B$ , from the solvated receptor-ligand complex,  $AB$ . The bottom reaction is the “annihilation” of the ligand from solvent. The free energy changes of the two reactions,  $\Delta G_1$  and  $\Delta G_2$ , combine as shown to yield what has been termed the “absolute free energy of binding,”  $\Delta G_{\text{abs}}$  (Jorgensen et al., 1988; Pranata and Jorgensen, 1991). The quantities  $\Delta G_1$  and  $\Delta G_2$  are computed by the method of multiconfiguration thermodynamic integration [see, for example, Straatsma and McCammon (1992)] or of free energy perturbation [see, for example, Lybrand, 1990].

With either approach, “annihilation” is accomplished by gradually turning off the interactions of the ligand with the rest of the system: the receptor-solvent system for  $\Delta G_1$ , and the solvent for  $\Delta G_2$ . A central observation of the present discussion is that this procedure does not actually annihilate the ligand; it merely decouples the ligand from its environment. In effect, “annihilation” converts the ligand into an ideal-gas molecule. With this in mind, the thermodynamic relation of Fig. 2 is redrawn in a more complete form in Fig. 3, where the subscripts (*sol*) and (*gas*) imply a species in solution or in the ideal gas phase, respectively. The free energies now bear a superscript that indicates a standard quantity. If the free energy changes shown in Fig. 3 can be computed, then the standard free energy of binding is their difference, as written in the figure. These free energy differences may be computed as follows.

The work of transferring the ligand from solution to the gas phase,  $\Delta G_2^\circ$ , does not depend upon the choice of standard concentration, so long as the same standard state is used for the two phases in comparing with experiment. This work of transfer may be computed by existing free energy perturbation or thermodynamic integration approaches, and need not be discussed here. However, it is not immediately obvious how to use computer simulation methods to find the standard free energy change associated with transferring the ligand from the receptor to the gas phase,  $\Delta G_1^\circ$ . This subsection therefore derives an expression for  $\Delta G_1^\circ$  in terms suitable for evaluation by thermodynamic integration. Mi-

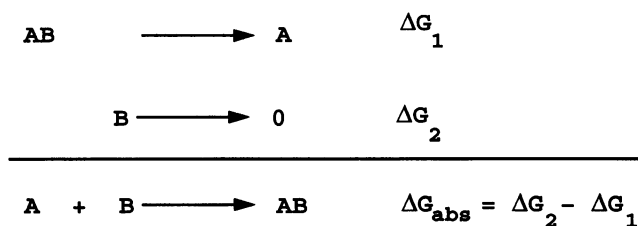


FIGURE 2 Thermodynamic analysis of double-annihilation method (Pranata and Jorgensen, 1991).

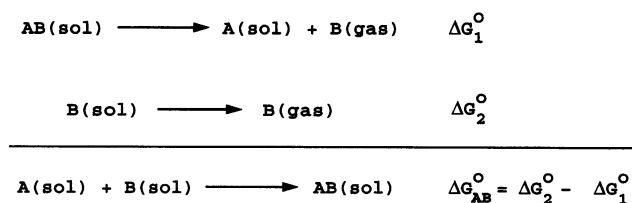


FIGURE 3 Thermodynamic analysis of "double-decoupling" method (see text).

nor modifications would permit its implementation by the method of free energy perturbation.

To begin, an expression is written for  $\Delta G_1^\circ$ . From Fig. 3 and Eq. 4,

$$\Delta G_1^\circ \equiv \mu_{\text{sol,A}}^\circ + \mu_{\text{gas,B}}^\circ - \mu_{\text{sol,AB}}^\circ, \quad (16)$$

where  $\mu_{\text{gas,B}}^\circ$  is the standard chemical potential of *B* in the ideal gas phase. From arguments analogous to those for a molecule in solution (second section), this standard chemical potential is given by

$$\mu_{\text{gas,B}}^\circ = -RT \ln \frac{Q_{0,B}}{V C^\circ} \quad (17)$$

$$= -RT \ln \left( \frac{8\pi^2}{\sigma_B C^\circ} \prod_{i=1}^{M_B} (2\pi m_i RT)^{3/2} Z_{0,B} \right), \quad (18)$$

where the configuration integral  $Z_{0,B}$  is defined by

$$Z_{0,B} \equiv \int e^{-\beta U(\mathbf{r}_B)} d\mathbf{r}_B. \quad (19)$$

Here  $Q_{0,B}$  on the right hand side of Eq. 17 is the molecular partition function of *B*, and  $V$  is the volume of the container. The factor of  $V$  is cancelled by a factor of  $V$  in  $Q_{0,B}$  which emerges when external and internal coordinates are separated. Here  $Z_{0,B}$  is the configuration integral of *B* in the gas phase, in internal coordinates.

Equations 8–12 provide expressions for  $\mu_{\text{sol,A}}^\circ$  and  $\mu_{\text{sol,AB}}^\circ$ , respectively. Combining these with Eqs. 17–19 yields

$$\Delta G_1^\circ = -RT \ln \left( \frac{8\pi^2}{C^\circ} \frac{\sigma_{\text{AB}}}{\sigma_A \sigma_B} \frac{Z_{\text{N,A}} Z_{0,B}}{Z_{\text{N,AB}}} \right) + P^\circ (\bar{V}_A - \bar{V}_{\text{AB}}) \quad (20)$$

Now, the use of thermodynamic integration to obtain the free energy difference between two states of a system requires the creation of an artificial energy function that interpolates smoothly between the energy functions of the initial and the final states. The precise functional form of the interpolation may be adjusted to optimize convergence of the simulations. Typically, the interpolation is controlled by a parameter  $\lambda \in [0, 1]$ , where 0 and 1 correspond to the starting and final energy functions, respectively. Here, the initial state is *AB(sol)*, and the final state is *A(sol) + B(gas)*.

Therefore, the artificial energy function  $U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)$  must satisfy:

$$U(0, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S) = U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S); \quad (21)$$

$$U(1, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S) = U(\mathbf{r}_A, \mathbf{r}_S) + U(\mathbf{r}_B). \quad (22)$$

A free energy function  $g(\lambda)$  is now constructed from the artificial potential energy function:

$$g(\lambda) \equiv -RT \ln \int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_A d\mathbf{r}_B d\zeta_B d\mathbf{r}_S. \quad (23)$$

As in Eq. 12,  $J_{\zeta_B}$  is the Jacobian determinant of the intermolecular coordinates. When  $\lambda = 1$ , the potential energy,  $U$ , depends upon  $(\mathbf{r}_A, \mathbf{r}_S)$ ,  $\mathbf{r}_B$ , and  $\zeta_B$  separately, so the integrals over these sets of coordinates can be separated. Therefore, the change in  $g(\lambda)$  when  $\lambda$  goes from 0 to 1 becomes

$$\begin{aligned}
 g(1) - g(0) &= -RT \ln \frac{\int e^{-\beta U(\mathbf{r}_A, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_B)} I(\zeta_B) J_{\zeta_B} d\mathbf{r}_A d\mathbf{r}_B d\zeta_B d\mathbf{r}_S}{\int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_A d\mathbf{r}_B d\zeta_B d\mathbf{r}_S} \\
 &= -RT \ln \frac{V_I \xi_I Z_{\text{N,A}} Z_{0,B}}{Z_{\text{N,AB}}}. \quad (24)
 \end{aligned}$$

$$= -RT \ln \frac{V_I \xi_I Z_{\text{N,A}} Z_{0,B}}{Z_{\text{N,AB}}}. \quad (25)$$

The second line shows the simplification that results when it is assumed that  $\int I(\zeta_B) J_{\zeta_B} d\zeta_B = V_I \xi_I$ , where  $V_I$  and  $\xi_I$  are from the integrals over position and orientation, respectively. This assumption leads to a formula that is particularly easy to interpret, but it is not actually needed to implement the thermodynamic integration method proposed here.

In the method of thermodynamic integration,  $g(1) - g(0)$  is evaluated using a numerical approximation to the following integral over  $\lambda$ :

$$g(1) - g(0) = \int \frac{\partial g(\lambda)}{\partial \lambda} d\lambda \quad (26)$$

$$= \int \left\langle \frac{\partial U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)}{\partial \lambda} \right\rangle_{\lambda, I(\zeta_B)=1} d\lambda \quad (27)$$

The subscripted angle brackets indicate a Boltzmann-weighted average in the conformational distribution appropriate to a given value of  $\lambda$ . In addition, from the definition of  $g(\lambda)$  in Eq. 23, it is clear that *the integrals used in computing this average range only over conformations for which  $I(\zeta_B) = 1$ ; i.e., for which the ligand is in or near the binding site.* This is indicated by the second subscript on the angle brackets. The Jacobian determinant,  $J_{\zeta_B}$ , is implicit in



the average, and is automatically accounted for by correct molecular dynamics or Monte Carlo sampling.

Equations 20, 25, and 27 permit  $\Delta G_1^\circ$  to be related to an integral that can be evaluated numerically in a computer simulation:

$$\begin{aligned} \Delta G_1^\circ = & \int \left\langle \frac{\partial U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)}{\partial \lambda} \right\rangle_{\lambda, I(\zeta_B)=1} d\lambda \\ & - RT \ln \left( \frac{\sigma_{AB}}{\sigma_A \sigma_B} \right) \\ & + RT \ln(C^\circ V_1) + RT \ln(\xi_l/8\pi^2) \\ & + P^\circ(\bar{V}_A - \bar{V}_{AB}) \end{aligned} \quad (28)$$

This is the chief result of this section. It states that the standard free energy change associated with the decoupling of the ligand from its binding site in solution may be evaluated from a free energy integration in which the ligand is *always* constrained to occupy a region in the binding site that is defined in the coordinate system of the receptor, even as the ligand is gradually decoupled from the receptor and the solvent. When the ligand is fully decoupled, it is a molecule of ideal gas still constrained to occupy the region where  $I(\zeta_B) = 1$ . It therefore possesses a well-defined chemical potential. The constraint can be implemented by an additional Hamiltonian component that is zero when the ligand is in the region where  $I(\zeta_B) = 0$ , and that rises sharply where  $I(\zeta_B)$  becomes zero. [A harmonic restraining potential could also be used (Hermans and Shankar, 1986; Roux et al., 1996; Zhang and Hermans, 1996).] Once the simulation is complete, however, it is necessary to correct for the fact that the chemical potential of the constrained ligand does not correspond to the standard concentration,  $C^\circ$ . Accordingly, the third term of Eq. 28 represents the change in free energy when the constrained, gas-phase ligand is allowed to expand to occupy a volume  $1/C^\circ$ . In Eq. 28, the form of this term is appropriate to a simulation in which the ligand is restrained by a hard-walled potential matching  $I(\zeta_B)$ . If the ligand is instead restrained in the binding site by a harmonic potential with force constant  $k$ , the corresponding term becomes  $-RT \ln [C^\circ(2\pi RT/k)^{3/2}]$  (Hermans and Shankar, 1986; Roux et al., 1996). Similarly, the fourth term in Eq. 28 represents the change in free energy when the rotationally constrained ligand is allowed to rotate freely. Again, a different form would be obtained if the orientation of the ligand were restrained harmonically. Thus, *these two terms correct the thermodynamic integration for the standard state.*

The last line of Eq. 28 contains the pressure-volume work associated with the change in volume of the receptor-solvent system when the ligand is decoupled from it. The volume change  $\bar{V}_A - \bar{V}_{AB}$  represents the volume change of the simulation box as  $\lambda$  goes from 0 to 1 in a constant-pressure simulation. This contribution is expected to be negligible at normal pressures.

It might appear that the results of the calculation prescribed here will depend upon the definition of  $I(\zeta_B)$ . The discussion in Standard Chemical Potential of a Complex in Solution shows that this should not be the case, so long as  $I(\zeta_B)$  meets certain rough guidelines. Nevertheless, it is worth discussing this point in the context of simulation. As the region where  $I(\zeta_B) = 1$  is made smaller, the two correction terms in Eq. 28 will become more negative. However, this change will be balanced by a larger work-integral, because the ligand will not be so free during the simulation to drift off into regions where  $\langle \partial U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S / \partial \lambda) \rangle_{\lambda, I(\zeta_B)=1}$  is small. Errors will occur only when the integration region defined by  $I(\zeta_B)$  becomes so small that conformations that ought to make important contributions to the work integral are missed. Errors will also occur if the region where  $I(\zeta_B)$  becomes so large that the simulation fails to converge.

To summarize, the double-decoupling method, like the double-annihilation method, requires two simulations. The first, a transfer of the ligand from solvent to gas, yields  $\Delta G_2^\circ$ . The second, just described, yields the standard free energy change for decoupling the ligand from the binding site of the solvated receptor. This yields  $\Delta G_1^\circ$  once a correction for the standard state is incorporated. The standard free energy of binding is then  $\Delta G_{AB}^\circ = \Delta G_2^\circ - \Delta G_1^\circ$ .

This methodology appears to be computationally feasible. In fact, as noted above, similar approaches have been used to compute the standard free energy of binding of xenon to a cavity in myoglobin (Hermans and Shankar, 1986) (see below), and to evaluate the thermodynamic stability of water molecules in protein cavities (Roux et al., 1996; Zhang and Hermans, 1996). These studies used a harmonic potential to restrain the ligand in the binding cavity, rather than the hard-walled potential matching  $I(\zeta_B)$  considered here. Also, the present treatment differs in allowing for an orientational restraint of the ligand. However, the various approaches are based upon the same principles and are equally valid, so long as they include the required correction terms, which correspond to the third and fourth terms of Eq. 28. Finally, the present analysis demonstrates that the same approach applies even when the binding site is not a sterically well-defined cavity. The constraint corresponding to  $I(\zeta_B)$  effectively defines the complex, and the precise definition of the complex is unimportant so long as binding is tight and all the energetically important regions of the binding site are sampled during the simulation.

#### *Review of Published Calculations with the Double-Annihilation Method*

It is of interest to consider the implications of the present analysis for calculations that have used the double-annihilation method. To review, the central observation of the present analysis is that the “annihilated” ligand is a molecule of ideal gas; that the computer simulation must give this molecule a well-defined chemical potential; and that a correction term is needed to account for the difference

between the standard state and the state assigned by the simulation. However, some calculations that use the double-annihilation method do not yield an ideal-gas molecule with a well-defined chemical potential, and none include a correction term, although the need for a correction term has been noted (Miyamoto and Kollman, 1993a; Janin, 1996). In contrast, the correction term has been included in the studies of xenon and water in protein cavities (Hermans and Shankar, 1986; Roux et al., 1996; Zhang and Hermans, 1996). In the double-annihilation method, the annihilated ligand has been treated in three different ways.

First, the original double-annihilation calculation (Jorgensen et al., 1988) computed the potential of mean force  $w(r)$  between two methane molecules as a function of the distance separating them. A minimum in  $w(r)$  was found at the contact distance of 4 Å. Because the distances examined ranged between 3.5 and 7.5 Å, the potential of mean force by itself did not yield the work of moving one methane from infinity to contact with the other methane. This quantity was obtained by a double-annihilation calculation in which a methane was annihilated from the contact position, and then from the bulk solvent. The resulting energy,  $-1.8$  kJ/mol, was referred to as the “absolute free energy of binding.” This energy value, and the entire potential of mean force  $w(r)$ , are indeed absolute, and have no concentration term, explicit or implicit. However, this single value of  $w(r)$  for  $r = 4$  Å is not the standard free energy of binding,  $\Delta G_{AB}^\circ$ . Obtaining  $\Delta G_{AB}^\circ$  from the potential of mean force  $w(r)$  requires integrating over  $r$  for a range defining the bound complex (see Eqs. 15 and 49). In practice, the methane-methane simulations converged slowly, and it was found necessary to integrate the “absolute binding free energy” while “reeling in” one methane toward the other. It is therefore somewhat difficult to interpret the results. Nonetheless, it is perhaps of interest that one can extract an approximate standard free energy of binding for two methanes by numerically integrating the potential of mean force plotted in Jorgensen et al. (1988). The results range from  $\sim 2$  to 4 kJ/mol depending upon whether one integrates over one or both minima in the curve. We are not aware of experimental data with which these results can be compared.

Second, some subsequent calculations allow the ligand to move without any imposed restraint while its coupling to the receptor and solvent is gradually decreased (Pranata and Jorgensen, 1991; Merz, Jr., 1991; Miyamoto and Kollman, 1992). An advantage of this method, compared to a calculation in which the ligand is fixed in position, is that the simulation can sample relevant intermolecular geometries. However, a difficulty arises in the final stages of the simulation, when the ligand is weakly coupled to the environment. The problem is that the ligand would have to explore the entire simulation box for the calculation to yield converged values of  $\langle \partial U(\lambda, \mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S) / \partial \lambda \rangle_\lambda$ . That the calculations in fact do not converge is evidenced by their irreversibility (Merz, Jr., 1991; Miyamoto and Kollman, 1992). On the other hand, if the decoupling ligand were, in effect, to sample a volume comparable to  $1/C^\circ = 1660 \text{ \AA}^3$ ,

no special term would be needed to correct for the standard concentration. Perhaps this explains why these calculations show little systematic error relative to experiment despite the lack of a correction term. However, because the calculations are unconverged, the results are not expected to be stable as a function of the length of the simulation.

Finally, in some calculations the position of the ligand has been restrained by artificial harmonic bonds to the starting position of the ligand and/or to the receptor. The restraints are left on as  $\lambda$  goes to 1 (Sneddon et al., 1989; Miyamoto and Kollman, 1993a, b; Mordasini Denti et al., 1996). This approach permits the ligand to sample at least part of the binding site, if the restraints are carefully defined. Another advantage of this approach is that the chemical potential of an ideal-gas molecule restrained by one or more harmonic “bonds” is well-defined. Therefore it is possible, at least in principle, to compute the standard-state correction. In practice, it is often difficult to determine from published work the precise form of the restraints. Nonetheless, it is striking that the reported “absolute binding free energies” computed by this approach are uniformly more negative (more favorable to binding) than measured standard free energies of binding. This is consistent with the neglect of a correction term that opposes binding. Thus, the computed absolute binding free energy of two formamide molecules in water is  $-12.1$  kJ/mol more favorable than the measured standard binding free energy of two *N*-methyl acetamide molecules in water (Sneddon et al., 1989), even though the two systems are chemically similar. (On the other hand, the bound conformation used in the simulations might not be the predominant bound conformation of the actual complex in solution. Therefore, the present comparison is subject to some uncertainty.) Similarly, for the binding of biotin with streptavidin, and of *N*-L-acetyltryptophanamide with chymotrypsin, the computed absolute free energies err with respect to measured standard free energies by  $-12.1$  and  $-16.3$  kJ/mol, respectively (Miyamoto and Kollman, 1993b). These systematic errors are consistent with the theory presented here, and suggest that calculations that account correctly for standard conditions will yield improved agreement with experiment.

#### *Treating Parts of a Receptor as Rigid*

The configuration integrals in the equations presented so far extend over all conformations of the two ligands. However, in many important cases, non-native conformations of a macromolecule are negligibly populated in both the free and bound form. It is thus appropriate that actual free energy calculations do not sample grossly non-native conformations of receptors. Furthermore, binding of a small ligand may produce little or no change in the conformation of the bulk of a large receptor. In other cases, binding may indeed alter the conformational distribution of the receptor, but substitution of one ligand for another in the binding site may produce little change in conformation (Appelt et al., 1991). In cases such as these, it is common practice to carry out

calculations in which parts of the protein are treated as rigid. This section examines the statistical thermodynamic basis for such approximations.

The molecule to be treated as partly rigid will be designated the receptor, and will be labeled A. The internal coordinates of this molecule are separated into those whose probability distributions change upon binding  $\mathbf{r}_a$ , and those whose probability distributions are unchanged by binding  $\mathbf{r}_a$ . The latter will tend to correspond to chemical groups far from the binding site, but may also correspond to rigid domains that move relative to each other, or to stiff degrees of freedom such as bond-lengths. The internal coordinates and external coordinates of the ligand will still be  $\mathbf{r}_B$  and  $\zeta_B$  (Fig. 4). With this notation, the binding free energy (Eq. 13) may be written as

$$\Delta G_{AB}^\circ = -RT \ln \left( \frac{C^\circ}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \frac{Z_{N,0}}{Z_{N,B}} \right) + P^\circ \Delta \bar{V}_{AB} \quad (29)$$

$$- RT \ln \left( \frac{\int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_a d\zeta_B d\mathbf{r}_B d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_a d\mathbf{r}_S} \right) \quad (30)$$

Treating the degrees of freedom  $\mathbf{r}_a$  as rigid in this expression means fixing these degrees of freedom in some conformation  $\mathbf{r}_a^*$ , and eliminating the integral over  $\mathbf{r}_a$ :

$$\Delta G_{AB}^\circ \approx -RT \ln \left( \frac{C^\circ}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \frac{Z_{N,0}}{Z_{N,B}} \right) + P^\circ \Delta \bar{V}_{AB} \quad (31)$$

$$- RT \ln \left( \frac{\int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a^*, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_a d\zeta_B d\mathbf{r}_B d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a^*, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_S} \right) \quad (32)$$

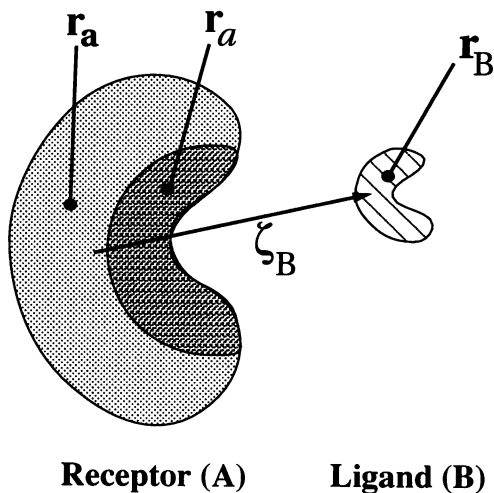


FIGURE 4 Diagram of separation of internal coordinates of complex.  $\mathbf{r}_a$ : internal coordinates of protein whose conformation distribution is assumed unchanged by binding.  $\mathbf{r}_a$ : internal coordinates of protein whose distribution may change upon binding.  $\mathbf{r}_B$ : internal coordinates of ligand.  $\zeta_B$ : position and orientation of ligand relative to protein.

This will be a good approximation when

$$\ln \frac{\int e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a^*, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_a d\mathbf{r}_S} \approx \ln \frac{\int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a^*, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_a d\zeta_B d\mathbf{r}_B d\mathbf{r}_S}{\int I(\zeta_B) J_{\zeta_B} e^{-\beta U(\mathbf{r}_a, \mathbf{r}_a, \mathbf{r}_B, \zeta_B, \mathbf{r}_S)} d\mathbf{r}_a d\mathbf{r}_a d\zeta_B d\mathbf{r}_B d\mathbf{r}_S} \quad (33)$$

That is, the logarithm of the probability density for  $\mathbf{r}_a^*$  must be unchanged by binding. Crystallographic or other structural studies are essential to determining which parts of a receptor meet this criterion and may therefore be treated as rigid.

The present treatment can also accommodate certain systems in which the receptor undergoes a large conformational change when it binds ligands. For example, when binding is associated with a hinge-bend conformational change that causes two rigid domains to close the binding site, it may be reasonable to include in  $\mathbf{r}_a$  only the torsion angles in the hinge region. The internal coordinates of the domains would then be fixed in conformation  $\mathbf{r}_a^*$ . Accomplishing this requires a change of coordinates.

It may also be concluded that, if binding does not affect the distributions of bond-lengths and bond-angles, it is permissible to treat these internal degrees of freedom as rigid also. This may be demonstrated formally if the Cartesian internal coordinates used here are replaced by internal coordinates consisting of bond-lengths, bond-angles, and torsion angles (Pitzer, 1946; Flory, 1988). It is worth emphasizing that the bonds and angles need not actually be rigid in order for them to be treated as rigid in calculations, so long as they are computationally held in a conformation the logarithm of whose probability does not change significantly upon binding. [For a detailed discussion of the separation of internal coordinates into "hard" and "soft" degrees of freedom, see Go and Scheraga (1969).]

In some cases, the conformation of a receptor does change significantly when it binds a ligand. Accounting for such changes in computational models of binding is difficult. Even in such cases, however, the conformation of a large part of the receptor may change little when one ligand is substituted for another in the binding site. A straightforward extension of the present analysis shows that it is then legitimate to keep that part of the receptor rigid when computing the relative binding constant of the two ligands.

Finally, this analysis is helpful in understanding the case where the receptor is tethered to a solid support, such as a separation column. So long as the tether and the solid support do not undergo conformational changes when the ligand is bound, their coordinates may be lumped with the  $\mathbf{r}_a$  internal coordinates of the receptor (above). Accordingly, their integrals will cancel when the binding constant is written, leaving a normal expression for the binding constant.

### OTHER REPRESENTATIONS OF THE SOLVENT

Successes in the development of implicit, or continuum, models of the solvent [see, e.g., Eisenberg and McLachlan

(1986); Kang et al. (1988); Still et al. (1990); Honig et al. (1993); von Freyberg et al. (1993); Stouten et al. (1993); Simonson and Brunger (1994); Madura et al. (1994); Cramer and Truhlar (1995); Gilson et al. (1995); Varnek et al. (1995); Marrone et al. (1996)] raise the possibility of making binding calculations more efficient by avoiding the explicit treatment of the solvent. The first subsection here reviews the statistical thermodynamic basis for the use of implicit solvent models in binding calculations. In some cases, it may be possible to make an implicit solvation model more accurate by including a few explicit solvent molecules. The second subsection therefore derives the theory necessary for using such a hybrid implicit/explicit solvent model in binding calculations.

Protein-ligand binding is often associated with the uptake or release of protons. In such cases, the apparent binding free energy is influenced by pH, and including the thermodynamic linkage between protonation and binding may be critically important. Therefore, the third subsection here provides a framework for including protonation equilibria in models of binding.

### Implicit Models of the Solvent

The formalism presented above is now used to isolate the chief effects of the solvent in a solvation energy term that depends upon the conformation of the solutes. Formalisms for separating solute and solvent degrees of freedom are discussed in a number of publications [see, e.g., Lifson and Oppenheim (1960); Ben-Naim (1975); Chandler and Pratt (1976)].

The standard free energy of binding in Eq. 13 contains three configuration integrals that involve both a solute and the solvent:  $Z_{N,A}$ ,  $Z_{N,B}$ , and  $Z_{N,AB}$ . These are of similar form, and we focus upon  $Z_{N,A}$ . To begin, the interaction of the solute with the solvent for a given configuration of the system is defined as

$$\Delta U(\mathbf{r}_A, \mathbf{r}_S) \equiv U(\mathbf{r}_A, \mathbf{r}_S) - U(\mathbf{r}_A) - U(\mathbf{r}_S). \quad (34)$$

It is then straightforward to show that

$$\mu_{\text{sol},A}^\circ = -RT \ln \left( \frac{8\pi^2}{C^\circ} \frac{M_A}{\sigma_A} \prod_{i=1}^{M_A} (2\pi m_i RT)^{3/2} Z_A \right) + P^\circ \bar{V}_A; \quad (35)$$

$$Z_A \equiv \frac{Z_{N,A}}{Z_{N,0}} = \int e^{-\beta[U(\mathbf{r}_A) + W(\mathbf{r}_A)]} d\mathbf{r}_A, \quad (36)$$

where

$$W(\mathbf{r}_A) \equiv -RT \ln \left( \frac{\int e^{-\beta\Delta U(\mathbf{r}_A, \mathbf{r}_S)} e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S}{\int e^{-\beta U(\mathbf{r}_S)} d\mathbf{r}_S} \right) \quad (37)$$

The parenthesized ratio in Eq. 37 is essentially the pure-solvent average of the Boltzmann factor for the solute-solvent interaction potential, for a given conformation  $\mathbf{r}_A$  of the solute. To within a small pressure-volume correction,

$W(\mathbf{r}_A)$  equals the work of transferring the solute in conformation  $\mathbf{r}_A$  from gas phase to the solvent. The solvation term  $W$  contains the most important effects of the solvent upon the standard chemical potential of the solute. It is a function of temperature and pressure, although this is not written explicitly here. In computations of binding free energies,  $W(\mathbf{r}_A)$  may be approximated by an implicit solvation model.

The standard free energy of binding given in Eq. 13 may now be rewritten as

$$\Delta G_{AB}^\circ = -RT \ln \left( \frac{C^\circ}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \frac{Z_{AB}}{Z_A Z_B} \right) + P^\circ \Delta \bar{V}_{AB}, \quad (38)$$

where  $Z_B$  and  $Z_{AB}$  are defined in exact analogy to  $Z_A$ . This equation yields the binding constant in terms of integrals over only the internal coordinates of the three solute species  $A$ ,  $B$ , and  $AB$ . It is worth pointing out that the methods of thermodynamic integration and free energy perturbation can be implemented with an implicit model of the solvent.

### Hybrid Models of the Solvent

For calculations with simple implicit models of solvent, it may be possible to improve accuracy by including a small number of solvent molecules explicitly (Beglov and Roux, 1994). We now derive expressions that permit the explicit treatment of a small number of solvent molecules that interact noncovalently with the solutes.

A small number  $n \ll N$  of solvent molecules that interact noncovalently with  $A$ ,  $B$ , or  $AB$  may be included explicitly by excluding these molecules from the  $N$  solvent molecules treated with the solvation term,  $W$ , and including them explicitly in the expression for the standard chemical potential of a solute. Thus, if  $n_A$  molecules are considered explicitly along with  $A$ , the argument of the logarithm of Eq. 5 is multiplied and divided by  $Q_{N-n_A,0}$ , and  $V_A = V_{N,A} - V_{N,0}$  is rewritten as  $(V_{N,A} - V_{N-n_A,0}) + (V_{N-n_A,0} - V_{N,0})$ , to yield:

$$\begin{aligned} \mu_{\text{sol},A}^\circ = & -RT \ln \left( \frac{1}{V_{N,A}} \frac{Q_{N,A}}{C^\circ Q_{N-n_A,0}} \right) \\ & + P^\circ (\bar{V}_{N,A} - \bar{V}_{N-n_A,0}) - RT \ln \left( \frac{Q_{N-n_A,0}}{Q_{N,0}} \right) \\ & + P^\circ (\bar{V}_{N-n_A,0} - \bar{V}_{N,0}), \end{aligned} \quad (39)$$

where  $V_{N-n_A,0}$  is the equilibrium volume of  $N - n_A$  molecules of pure solvent at standard pressure.

The first line of Eq. 39 is recognizable, by comparison with Eq. 5, as the standard chemical potential of an entity consisting of the solute,  $A$ , together with  $n_A$  molecules of solvent. The second line of Eq. 39 is directly related to the chemical potential of the solvent  $\mu_s$  as shown by the

following identity:

$$RT \ln \left( \frac{Q_{N-n_A,0}}{Q_{N,0}} \right) + P^\circ (\bar{V}_{N-n_A,0} - \bar{V}_{N,0}) \quad (40)$$

$$= RT \sum_{i=1}^{n_A} \left( \ln \frac{Q_{N-i+1,0}}{Q_{N-i,0}} \right) + n_A P^\circ \bar{V}_s$$

where  $\bar{V}_s$  is the partial molar volume of the solvent. The second equality holds because each term in the sum is the change in the Gibbs free energy upon removing one of the  $N$  molecules of solvent, under conditions where  $n_A \ll N$ . It is worth emphasizing that  $\mu_s$  is not the *standard* chemical potential of the solvent, but its *actual* chemical potential.

The first ratio of partition functions in Equation 39 may be treated in the same way as that in Eq. 5, except that now the coordinates of  $n_A$  solvent particles are treated explicitly along with the coordinates of the solute. Thus,

$$\mu_{\text{sol},A}^\circ = -RT \ln \left( \frac{8\pi^2}{C^\circ \sigma_A \sigma_S^{n_A}} \prod_{i=1}^{m_A+M_{n_A}} (2\pi m_i RT)^{3/2} \int e^{-\beta(U(\mathbf{r}_A, \mathbf{r}_{SA}) + W(\mathbf{r}_A, \mathbf{r}_{SA}))} d\mathbf{r}_A d\mathbf{r}_{SA} \right) \quad (41)$$

$$+ P^\circ (\bar{V}_A + n_A \bar{V}_s) - n_A \mu_s \equiv \mu_{n_A, A}^\circ - n_A \mu_s,$$

Here  $\mathbf{r}_{SA}$  represents all the coordinates of the  $n$  solvent molecules associated with  $A$ , defined in the frame of reference of  $A$ ; and  $M_{n_A}$  is the number of atoms belonging to the  $n_A$  solvent molecules. The final line defines  $\mu_{n_A, A}^\circ$ , the standard chemical potential of the solute together with  $n$  solvent molecules. Equation 41 demonstrates that a model of binding can use an implicit model of solvent as an approximation to  $W$  and still include some solvent molecules explicitly. The only adjustment is that the cost of removing these solvents from the bulk of the solvent,  $n_A \mu_s$ , must be accounted for.

The standard free energy of binding may now be written as

$$\Delta G^\circ = \mu_{AB, n_{AB}}^\circ - \mu_{A, n_A}^\circ - \mu_{B, n_B}^\circ - (n_{AB} - n_A - n_B) \mu_s, \quad (42)$$

where  $n_{AB}$ ,  $n_A$  and  $n_B$  are the numbers of solvent molecules included explicitly for  $AB$ ,  $A$ , and  $B$ , respectively. This formulation may be useful for calculations that combine an implicit model of solvent with a few explicitly treated solvent molecules. A disadvantage is that it requires integration over all possible placements of the explicitly treated solvent molecules. However, in cases where only a few solvent-sites contribute substantially, it should be possible to restrict the range of the integrals.

### The Influence of Solvent pH

The noncovalent binding of ligands by proteins may be associated with the binding or release of protons. As a

consequence, the pH of the solvent can affect the apparent binding constant. The treatment of binding presented above cannot yet account for this linkage. Therefore, this subsection extends the treatment of binding to account for the pH of the solvent. The resulting expressions are analogous to those for the explicit treatment of a few solvent molecules.

The binding polynomial formalism (Wyman, 1965; Schellman, 1975) provides a way to incorporate protonation equilibria into the present treatment of binding. Suppose that solute  $A$  exists in a number  $L$  of protonation states, designated  $A_p$ , each having a standard chemical potential,  $\mu_{A_p}^\circ$ , given by Eq. 5. Then the overall standard chemical potential of  $A$ , including contributions from all protonation forms,  $A_p$ , is

$$\mu_A^\circ = -RT \ln \sum_{p=0}^L e^{-\beta(\mu_{A_p}^\circ - \Delta n_p \mu_{H^+})}. \quad (43)$$

Here  $\mu_{H^+}$  is the chemical potential of the proton (not its standard chemical potential) and  $\Delta n_p$  is the difference between the number of protons bound in state  $A_p$  and the number bound in a reference state  $A_0$ . Which protonation state is designated  $A_0$  is purely a matter of convenience (Gilson, 1993). One possibility is the fully deprotonated state; another is the predominant protonation state at the pH of interest. The chemical potentials of  $B$  and  $AB$  may be modified in the same way to account for protonation equilibria. This approach has the disadvantage of prescribing direct computations of changes in free energy associated with covalent binding of protons. Fortunately, it is possible to avoid this difficulty by formulating the problem in terms of changes in the ionization energies of titratable groups of known initial pKa [see, e.g., Tanford and Kirkwood (1957); Matthew et al. (1985); Bashford and Karplus (1990); Yang and Honig (1993); Gilson (1993)]. Furthermore, if only a single protonation state,  $p$ ,  $q$ ,  $r$ , is highly occupied for each of  $A$ ,  $B$ , and  $AB$ , respectively, then the standard free energy of binding becomes

$$\Delta G_{AB}^\circ = \mu_{AB, p}^\circ - \mu_{A, p}^\circ - \mu_{B, q}^\circ - \Delta n \mu_{H^+}, \quad (44)$$

where  $\Delta n$  is the number of protons taken up upon binding. This formula is essentially the same as Eq. 42, except that covalently bound protons are substituted for noncovalently bound solvent molecules.

### ENTROPY AND ENERGY COMPONENTS OF THE BINDING FREE ENERGY

A number of models express the free energy of binding as a sum of attractive interaction terms and unfavorable entropy terms [e.g., Gurney (1953); Chothia and Janin (1975); Andrews et al. (1984); Erickson (1989); Novotny et al. (1989); Searle et al. (1992a); Horton and Lewis (1992); Murphy et al. (1993); Weng et al. (1996)]. Such “energy-component” models are simple and tractable. However, the theory underlying the various free energy components often

is unclear. In particular, a number of different theoretical approaches have been used to estimate the change in entropy associated with the loss of external freedom of two molecules upon binding; i.e., the changes in translational and rotational entropy. A major goal of the present paper is to establish clear definitions of the entropy changes upon binding. First, however, it is worth reviewing existing approaches to this problem.

One approach involves the use of the classical Sackur-Tetrode equation for the translational entropy of an ideal gas (McQuarrie, 1973) to compute the changes in translational entropy due to binding of molecules in solution (Doty and Myers, 1953; Steinberg and Scheraga, 1963; Page and Jencks, 1971; Chothia and Janin, 1975; Janin and Chothia, 1978; Finkelstein and Janin, 1989; Erickson, 1989; Searle et al., 1992b; Spolar and Record, Jr., 1994; Morton et al., 1995; Janin, 1995). A rotational entropy contribution, based upon ideal-gas formulas for rotational entropy, is often included in such models. This ideal-gas approach to estimating entropy changes is used without derivation in an early discussion of the dimerization of insulin that implicitly assumes that the translational entropy goes to zero upon binding (Doty and Myers, 1953). A subsequent paper justifies the use of ideal-gas theory for binding in solution, and points out that the translational entropy lost on binding is partly replaced by contributions from degrees of freedom internal to the complex (Steinberg and Scheraga, 1963). The ideal gas formulas for entropy have been applied in several different ways to compute binding entropies. Some applications yield entropy changes that depend upon the masses of the molecules involved (Doty and Myers, 1953; Page and Jencks, 1971; Chothia and Janin, 1975; Janin and Chothia, 1978; Searle et al., 1992b; Spolar and Record, Jr., 1994; Morton et al., 1995). Other applications yield entropy changes in which masses cancel out, leaving only effective volume terms (Finkelstein and Janin, 1989; Erickson, 1989; Janin, 1995). The existing literature does not contain a definitive analysis of the application of ideal-gas formulas to the problem of binding in solution. Moreover, current papers offer conflicting views on whether entropy changes upon binding depend upon molecular mass.

Another method of estimating the change in translational entropy upon binding is associated with the concept of the cratic entropy (Gurney, 1953; Kauzmann, 1959). In this approach, the free energy of binding is expressed as a sum of two contributions. The "unitary" contribution is related to the favorable work of assembling the complex at a fixed point in space, starting from two widely separated molecules at fixed points. The "cratic," or mixing, entropy is said to correct for the fact that the molecules and the complex are free to mix with solvent molecules. The mixing or cratic entropy term depends not upon concentration or number density of the solutes, but upon their mole fraction. This approach has been used in a model of protein-protein association (Novotny et al., 1989). It has been argued, based upon interpretation of experimental binding data (Murphy et al., 1994), that it is superior to approaches based upon the

Sackur-Tetrode equation. It has also been argued, based upon a thermodynamic analysis, that the cratic entropy is fallacious (Holtzer, 1995).

This section of the paper seeks to elucidate the statistical thermodynamic basis of energy-component models of binding, and pays particular attention to the entropy components. The first part of the section defines a partitioning of the entropy change upon binding. Three major components are identified: a contribution from the change in the external freedom of the molecules; a contribution from the change in the internal, or configurational, freedom of each molecule; and a term associated with the change of entropy of the solvent. It is also shown that the choice of interaction terms to be included in an energy-component model of binding depends upon which entropy terms are included. The second part of the section discusses in detail the changes in entropy that result from loss of external freedom of the solutes upon binding, and includes discussions of the "cratic" entropy term and the influence of molecular mass upon binding.

## External, Internal and Solvent Entropy

### *External Entropy*

The translational and rotational contributions to partition functions of molecules in the gas phase involve integrals over both momentum and position (McQuarrie, 1973). However, in classical statistical thermodynamics, the momentum parts of the partition functions cancel in the final expressions for the free energy of binding (see above). Thus, the purely classical treatment assumes that binding does not restrict the freedom of the molecules in the momentum part of phase space, but only in the spatial part. (Some of the physical limitations of this assumption are discussed in the last section.) As a consequence, the terms "translational" and "rotational" entropy are somewhat misleading. Here, "positional" and "orientational" entropy will be used instead. Together, these constitute the "external" entropy. In order to define the change in external entropy, it is necessary to rewrite the free energy of binding,  $\Delta G_{AB}^{\circ}$ , as an integral over only the external degrees of freedom of one of the molecules. This is accomplished by defining a potential of mean force (Lifson and Oppenheim, 1960; Go and Scheraga, 1969) for the interaction of the two solutes as a function of their relative position and orientation,  $\zeta_B$ . When the standard relationship between entropy and free energy (McQuarrie, 1973),

$$\Delta S^{\circ} = -\left(\frac{\partial \Delta G^{\circ}}{\partial T}\right)_p, \quad (45)$$

is applied to the resulting expression for  $\Delta G_{AB}^{\circ}$ , the chain rule yields the entropy of binding as a sum of contributions from the external, internal, and solvent degrees of freedom. The subscripted parentheses in Eq. 45 imply a derivative at constant pressure.

To begin, the in vacuo and solvent-mediated interactions of *A* and *B* are defined respectively as

$$\Delta U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) \equiv U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) - U(\mathbf{r}_A) - U(\mathbf{r}_B); \quad (46)$$

$$\Delta W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) \equiv W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) - W(\mathbf{r}_A) - W(\mathbf{r}_B). \quad (47)$$

These terms give the energy of the complex in a specified conformation, relative to the energy of the two solutes in the same conformations but widely separated. An interaction potential of mean force that depends only upon  $\zeta_B$  is now defined by

$$w(\zeta_B) \equiv -RT \ln \left( \frac{\int \int e^{-\beta(\Delta U(\zeta_B) + \Delta W(\zeta_B))} e^{-\beta(U(\mathbf{r}_A) + W(\mathbf{r}_A) + U(\mathbf{r}_B) + W(\mathbf{r}_B))} d\mathbf{r}_A d\mathbf{r}_B}{\int \int e^{-\beta(U(\mathbf{r}_A) + W(\mathbf{r}_A) + U(\mathbf{r}_B) + W(\mathbf{r}_B))} d\mathbf{r}_A d\mathbf{r}_B} \right) \\ = -RT \ln \langle e^{-\beta(\Delta U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) + \Delta W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B))} \rangle_{\mathbf{r}_A, \mathbf{r}_B, \text{ apart}}, \quad (48)$$

where the subscripted angle brackets indicate an ensemble-average over  $\mathbf{r}_A$  and  $\mathbf{r}_B$ . This average is based upon the conformational distributions of the two solutes when they are far apart. Because  $\Delta U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) + \Delta W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B)$  goes to zero as the solutes are widely separated,  $w(\zeta_B)$  does also. This potential of mean force is derived from an average over all solvent and internal solute degrees of freedom, and is directly related to the probability distribution of  $\zeta_B$ . Note that  $w$  includes not only solvent contributions, but also contributions from the direct interaction,  $\Delta U$ , between the two solutes.

The standard free energy of binding as given in Eq. 38 can now be expressed as an integral over only the external degrees of freedom:

$$\Delta G_{AB}^\circ = -RT \ln \left( \frac{C^\circ}{8\pi^2} \frac{\sigma_A \sigma_B}{\sigma_{AB}} \int I(\zeta_B) J_{\zeta_B} e^{-\beta w(\zeta_B)} d\zeta_B \right) \\ + P^\circ \Delta \bar{V}_{AB} \quad (49)$$

This equation generalizes similar formulations that apply only to rigid, spherical molecules (Prue, 1969; Justice and Justice, 1976; Chandler, 1979; Shoup and Szabo, 1982).

Inserting the expression for  $\Delta G_{AB}^\circ$  in Eq. 49 into Eq. 45 gives

$$\Delta S_{AB}^\circ = -\frac{1}{T} (\Delta G_{AB}^\circ - P^\circ \Delta \bar{V}_{AB}) \\ + \frac{1}{T} \langle w(\zeta_B) \rangle_{AB} - \left\langle \frac{\partial w(\zeta_B)}{\partial T} \right\rangle_{AB} + \Delta S_{PdV}^\circ, \quad (50)$$

where, as in what follows, it is implicit that the derivatives with respect to temperature are all at constant pressure.

The last term in Eq. 50,  $\Delta S_{PdV}^\circ \equiv P(\partial \Delta \bar{V}_{AB} / \partial T)_P$  is the temperature derivative of the very small pressure-volume work associated with binding at constant pressure. For a standard pressure of 1 atmosphere, it is expected to be

negligible, and it will not be discussed in detail here. However, it is worth mentioning that this term is distinct from what might be a much larger change in entropy if the reaction occurred at constant volume. At constant volume, changes in system pressure lead to changes in density and thus, potentially, to reorganization of the solvent (Lazaridis and Paulaitis, 1993; Yu and Karplus, 1988).

Equation 50 is now used to define the change in external entropy upon binding. Intuitively, the change in external entropy is that which results from the loss of positional and orientation freedom upon formation of the complex. It is thus the change in entropy that would result from the interaction potential of mean force  $w(\zeta_B)$  if the only degrees of freedom in the system were the external coordinates  $\zeta_B$ . Assuming the absence of other degrees of freedom means assuming that the temperature derivative of  $w(\zeta_B)$  equals zero and that  $\Delta S_{PdV}^\circ = 0$ . Therefore, the change in external entropy is obtained by setting to zero the third and fourth terms on the right hand side of Eq. 50:

$$\Delta S_{ext}^\circ \equiv -\frac{1}{T} (\Delta G_{AB}^\circ - P^\circ \Delta \bar{V}_{AB}) + \frac{1}{T} \langle w(\zeta_B) \rangle_{AB} \quad (51)$$

It is readily shown that, when  $\Delta S_{ext}^\circ$  is defined in this way, it equals the standard entropy change upon binding that would be obtained for two rigid molecules in the gas phase with an interaction potential  $U(\zeta_B) = w(\zeta_B)$ . Thus, theory appropriate to two structureless molecules in the gas phase can be used to compute changes in external entropy for binding in solution, so long as the intermolecular potential of mean force is substituted for the intermolecular potential energy. It is also worth pointing out that the change in external entropy depends upon the standard concentration.

### Internal Entropy

Now the change in entropy associated with the internal degrees of freedom of the solutes,  $\Delta S_{int}^\circ$ , is considered. Intuitively, the change in internal entropy results from changes in the conformational freedom of the solutes upon binding. That is, it is the change in entropy, over and above the external entropy change, that results from treating the internal degrees of freedom explicitly, but assuming that the solvation term,  $W$ , has no entropic component. This means assuming the temperature derivative of the solvation term to be zero. A formula for  $\Delta S_{int}^\circ$  is obtained by expansion of the temperature derivative of  $w(\zeta_B)$  in Eq. 50:

$$-\left\langle \frac{\partial w(\zeta_B)}{\partial T} \right\rangle_{AB} = -\frac{1}{T} \langle w(\zeta_B) \rangle_{AB} \\ + \frac{1}{T} [\langle U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) + W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) \rangle_{AB} \\ - \langle U(\mathbf{r}_A) + W(\mathbf{r}_A) \rangle_A - \langle U(\mathbf{r}_B) + W(\mathbf{r}_B) \rangle_B] \\ - \left( \left\langle \frac{\partial W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B)}{\partial T} \right\rangle_{AB} - \left\langle \frac{\partial W(\mathbf{r}_A)}{\partial T} \right\rangle_A - \left\langle \frac{\partial W(\mathbf{r}_B)}{\partial T} \right\rangle_B \right). \quad (52)$$

Assuming the temperature derivative of the solvation term to be zero yields the following definition of the internal entropy:

$$\Delta S_{\text{int}}^{\circ} \equiv -\frac{1}{T} \langle w(\zeta_B) \rangle_{\text{AB}} + \frac{1}{T} [\langle U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) + W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) \rangle_{\text{AB}} - \langle U(\mathbf{r}_A) + W(\mathbf{r}_A) \rangle_A - \langle U(\mathbf{r}_B) + W(\mathbf{r}_B) \rangle_B]. \quad (53)$$

When the entropy components are defined in this way, the binding free energy change may be rewritten as:

$$\Delta G_{\text{AB}}^{\circ} = \langle U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) + W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B) \rangle_{\text{AB}} - \langle U(\mathbf{r}_A) + W(\mathbf{r}_A) \rangle_A - \langle U(\mathbf{r}_B) + W(\mathbf{r}_B) \rangle_B - T\Delta S_{\text{ext}}^{\circ} - T\Delta S_{\text{int}}^{\circ} - T\Delta S_{\text{pdv}}^{\circ} + P^{\circ} \Delta \bar{V}_{\text{AB}}. \quad (54)$$

This says that the free energy change upon binding equals the sum of the changes in external and internal entropy, plus the change in the mean of  $U + W$  upon binding, plus minor pressure-volume terms. It can be shown that  $\Delta S_{\text{ext}}^{\circ} + \Delta S_{\text{int}}^{\circ}$  equals the standard entropy change upon binding for two flexible molecules in the gas phase with a potential energy function equal to  $U + W$ . Thus, an energy-component model of binding that includes explicit external and internal entropy contributions, but not solvent entropy contributions, must use the solvent-modified interaction potential  $U + W$  for the intra- and inter-molecular interactions of the solutes. Also, this interaction potential must be Boltzmann-averaged over the internal coordinates of the solutes, as indicated by the angle-brackets in Eq. 54.

### Solvent Entropy

Finally, it is clear from the above that the change in solvent entropy upon binding is appropriately defined as

$$\Delta S_{\text{solv}}^{\circ} \equiv -\left\langle \frac{\partial W(\mathbf{r}_A, \mathbf{r}_B, \zeta_B)}{\partial T} \right\rangle_{\text{AB}} + \left\langle \frac{\partial W(\mathbf{r}_A)}{\partial T} \right\rangle_A + \left\langle \frac{\partial W(\mathbf{r}_B)}{\partial T} \right\rangle_B \quad (55)$$

It can be shown from Eqs. 37, 54 and 55 that

$$\Delta G_{\text{AB}}^{\circ} = \langle U(\mathbf{r}_A, \mathbf{r}_B, \zeta_B, \mathbf{r}_S) \rangle_{\text{AB}} - \langle U(\mathbf{r}_A, \mathbf{r}_S) \rangle_A - \langle U(\mathbf{r}_B, \mathbf{r}_S) \rangle_B - T\Delta S_{\text{ext}}^{\circ} - T\Delta S_{\text{int}}^{\circ} - T\Delta S_{\text{solv}}^{\circ} - T\Delta S_{\text{pdv}}^{\circ} + P^{\circ} \bar{V}_{\text{AB}}, \quad (56)$$

where the potential energy terms  $U$  include contributions from the solutes and the  $N$  solvent molecules. This says that, when all three entropy contributions are made explicit, the only other contribution to the free energy is the enthalpy; i.e., the mean internal energy plus  $PV$ . This is as expected

because, at constant pressure,  $\Delta G \equiv \Delta H - T\Delta S \equiv \Delta U + P\Delta V - T\Delta S$ .

It is worth noting that the present analysis yields a way of computing the change of entropy upon binding when an implicit representation of the solvent is used. From Eq. 50 and 52,

$$\Delta S_{\text{AB}}^{\circ} = \frac{1}{T} (\langle U + W \rangle_{\text{AB}} - \langle U + W \rangle_A - \langle U + W \rangle_B - \Delta G_{\text{AB}}^{\circ}) - \left( \left\langle \frac{\partial W}{\partial T} \right\rangle_{\text{AB}} - \left\langle \frac{\partial W}{\partial T} \right\rangle_A - \left\langle \frac{\partial W}{\partial T} \right\rangle_B \right), \quad (57)$$

where, for brevity, arguments have been omitted from  $U$  and  $W$  and  $P\Delta V$  terms have been neglected. The first quantity in parentheses here is the entropy change of the solute degrees of freedom,  $\Delta S_{\text{int}}^{\circ} + S_{\text{ext}}^{\circ}$ . The last quantity in parentheses is the change in solvent entropy. In fact, because  $W$  is the solvation free energy for a solute in a fixed conformation, Eq. 45 implies that  $-\partial W/\partial T$  is the solvation entropy of a solute in a fixed configuration. Thus, Eq. 57 shows that the contribution of the solvent to the overall change in entropy upon binding is merely the change in the mean solvation entropy upon binding. This shows how an implicit solvation model that partitioned the solvation free energy into entropy and enthalpy could be used in computing the overall change in enthalpy and entropy upon binding. Such calculations might be compared with calorimetric results, for these yield both entropy and enthalpy. This approach has been used in an analysis of the thermodynamics of  $\alpha$ -helix propensities (Wang and Purisima, 1996).

### Discussion of Changes in Positional and Orientational Entropy

The change in external entropy is now analysed further. First, the concept of the volume of binding is discussed. It is then demonstrated that the change in external entropy depends upon the definition of the molecular coordinate frame. The "cratic entropy" of Gurney (Gurney, 1953) and Kauzmann (Kauzmann, 1959) is discussed briefly. Finally, it is argued that classical statistical thermodynamics is a good approximation for models of noncovalent binding and that, as a consequence, the binding entropy is essentially independent of molecular mass.

#### The Volume of Binding

The concept of the binding volume is illustrated by an artificial system. It is assumed that the potential of mean force,  $w(\zeta_B)$ , is a step-function having some negative value  $w_b$  when the relative position and orientation of  $B$  relative to  $A$ ,  $\zeta_B$ , fall within regions of volume  $V_b$  and angular volume  $\xi_b$  that are contained by the region where  $I(\zeta_B) = 1$ . So long



as  $w_b \ll -RT$ , as expected for strong binding, it will be a good approximation that  $\langle w(\zeta_B) \rangle_{AB} = w_b$ , and therefore

$$\Delta G_{AB}^\circ = w_b - RT \ln(C^\circ V_b) - RT \ln\left(\frac{\xi_b}{8\pi^2}\right) \quad (58)$$

and

$$\Delta S_{AB}^\circ = R \ln(C^\circ V_b) + R \ln\left(\frac{\xi_b}{8\pi^2}\right) - \left\langle \frac{\partial w(\zeta_B)}{\partial T} \right\rangle_{AB}. \quad (59)$$

The expression for  $\Delta G_{AB}^\circ$  consists of the free energy of interaction,  $w_b$ , along with terms clearly identifiable as changes in positional and orientational—i.e. external—entropy (Finkelstein and Janin, 1989; Erickson, 1989). It is important to recognize that  $V_b$  and  $\xi_b$  are the volumes of the region where  $w(\zeta_B) = w_b$ , not the volumes where  $I(\zeta_B) = 1$ . Thus, it is a physically meaningful interaction potential of mean force that determines the binding volumes. The step function that defines the complex,  $I(\zeta_B)$ , is not directly related to the binding volumes.

It would clearly be of interest to establish realistic values for the binding volumes  $V_b$  and  $\xi_b$  that determine the change in external entropy. However, the external entropy can be interpreted in terms of well-defined binding volumes only when the interaction potential of mean force is a step function. For real molecules in solution,  $w$  is a complicated function of  $\zeta_B$ , so the binding volumes found here and discussed elsewhere (Finkelstein and Janin, 1989; Erickson, 1989) are artificial constructs (Hill, 1985b). Still, the concept is intuitively helpful, and it would be of interest to establish approximate, or effective, binding volumes. It has been suggested that the crystallographic thermal factors of a bound complex provide information on the motions of one molecule relative to the other (Finkelstein and Janin, 1989; Janin, 1995). However, the position variations measured by thermal factors relate to the coordinate system of the crystal lattice. Thermal factors provide at best limited information on the mobility of one atom relative to another. For example, two atoms covalently bonded to each other may have large thermal factors, even though the bond between them sharply restricts their relative motion. Thus, thermal factors do not offer much information on the changes in external entropy that result from binding.

#### Ambiguity of the Change in External Entropy

The idea of a change in external entropy, although intuitively pleasing, is somewhat artificial. This is because the change in external entropy depends upon how the internal and external coordinates are defined. This ambiguity arises even if both molecules are rigid, but is illustrated in an extreme form by the case of a flexible linear ligand  $B$  of  $M_B$  atoms,  $i = 1 \dots M_B$  (Fig. 5). Atoms 1, 2, and 3 are bound restrictively by another molecule  $A$ , as shown, leaving the rest of the ligand mobile in solution. If the external coordinates of the ligand are defined with respect to atoms 1–3 (see the second section), the drop in external entropy upon

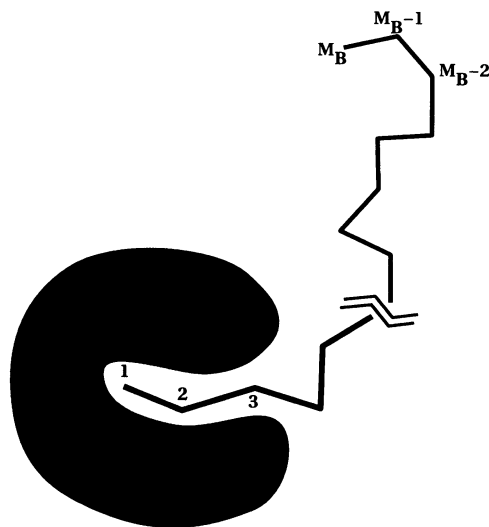


FIGURE 5 Diagram of binding of a flexible polymer to a protein by its first three atoms.

binding will be large, because these three atoms are tightly restricted in the complex. In contrast, if the external coordinates of the ligand are defined with respect to the atoms at its other end,  $M_B-2$ ,  $M_B-1$ , and  $M_B$ , the drop in external entropy upon binding will be far less, because these three atoms are not so tightly restricted in the complex. However, the total change in entropy upon binding must be independent of the choice of coordinates. It follows that this change in coordinates must shift the attribution of the drop in entropy associated with binding from external to internal degrees of freedom. This is reasonable, because when the polymer is bound to the protein, each accessible position and orientation of atoms  $M_B-2$ ,  $M_B-1$ , and  $M_B$  accommodates only a limited set of internal coordinates that place atoms 1, 2, and 3 in the binding site. Thus, models of binding that include an explicit term for the change in external entropy upon binding are incomplete unless they specify the coordinate system.

#### The Cratic Entropy

It has been stated that the standard entropy of a dissolved molecule equals the “unitary” entropy it possesses when it is forced to remain at a fixed position, plus a “cratic” entropy associated with its release that is given by

$$S_{\text{cratic}} \equiv -R \ln x, \quad (60)$$

where  $x$  is the mole fraction of the solute (Gurney, 1953; Kauzmann, 1959). Because the concentration of liquid water is 55 M, the cratic entropy of a solute at the standard 1 M concentration in water is  $\sim -R \ln 1/55$ . This concept has been used in models for binding, where the binding of molecules  $A$  and  $B$  to form  $AB$  decreases the number of solute particles from 2 to 1. This decrease in the number of particles is said to be associated with a contribution to the

standard free energy of binding of  $-RT \ln 1/55 = 10$  kJ/mol (Novotny et al., 1989; Murphy et al., 1994). The derivation in the present paper suggests that the concept is unnecessary, because it is not required for a complete theoretical treatment of binding; nor does it seem to emerge naturally from this treatment. Thus, the present analysis in terms of statistical thermodynamics is consistent with a recent critique that primarily used thermodynamics (Holtzer, 1995).

It is also worth noting that an uncritical use of cratic entropy as a component of the free energy of binding leads to the clearly incorrect result that the standard free energy of binding for two molecules in the gas phase does not depend upon the standard concentration. This is because the standard state in gas phase has each of the three molecular species,  $A$ ,  $B$ , and  $AB$ , at equal concentration  $C^\circ$ , and no solvent is present. Therefore, the mole fraction of each species is  $1/3$ , and the cratic entropy change is fixed at  $R \ln 3$ , regardless of the value of  $C^\circ$ .

#### *Molecular Mass and the Free Energy of Binding*

A number of papers consider binding constants to depend upon molecular mass (Doty and Myers, 1953; Page and Jencks, 1971; Chothia and Janin, 1975; Janin and Chothia, 1978; Searle et al., 1992b; Spolar and Record, Jr., 1994; Morton et al., 1995). This supposed dependence upon mass usually derives from the idea that binding "freezes out" the translational freedom of molecule  $B$  relative to  $A$ . It is sometimes assumed that the overall rotation of  $B$  also is frozen out. However, it is not possible to freeze out any degrees of freedom in classical statistical thermodynamics because confining a particle in an increasingly narrow energy well costs infinite entropy. For example, the translational entropy of a classical monatomic gas, given by the Sackur-Tetrode equation, decreases without limit as the volume  $V$  of its container goes to zero. On the other hand, when the Sackur-Tetrode equation is used to compute the entropy change for a volume change from  $V_1$  to some  $V_2 > 0$ , the entropy change does not depend upon mass, and is simply  $R \ln (V_2/V_1)$ .

Still, one may correctly use a mathematical device in which binding is viewed as the "freezing out" of three translational degrees of freedom, followed by the "adding back" of three degrees of freedom internal to the complex (Steinberg and Scheraga, 1963; Searle et al., 1992b; Tidor and Karplus, 1994). This approach does not actually freeze or neglect any degrees of freedom. Therefore, if the classical approximation to statistical thermodynamics is used, this approach must still lead to cancellation of all mass-dependence of the binding constant, and the results should be equivalent to those derived in the present paper. However, this freezing-out approach is more difficult to use mathematically, because it dictates carrying out the phase space integrals in an order that retains momentum contributions until a final cancellation occurs. If the final cancellation is not noted, the free energy of binding will be expressed as the sum of a translational contribution that depends upon the

molecular mass; a rotational contribution that depends upon the rotational moments of inertia; and internal contributions, often approximated with vibrational normal modes, that also depend upon mass [see, for example, Steinberg and Scheraga (1963); Tidor and Karplus (1994)]. When classical statistical thermodynamics is used, it is much more convenient to eliminate the momentum integrals at the start, as typically done in liquid theory (McQuarrie, 1973), and as done here.

On the other hand, the approach of freezing out translational degrees of freedom and then adding back internal degrees of freedom facilitates treating some degrees of freedom quantum-mechanically rather than classically. This approach is essential for treating *covalent* binding at room temperature, because classical statistical thermodynamics will be a poor approximation for the new vibrational modes associated with the new chemical bond. In such cases, the effects of mass upon the binding constant need not cancel. This approach has been used in a computational study of the entropy changes associated with the *noncovalent* dimerization of insulin (Tidor and Karplus, 1994), in which normal mode vibrations are treated quantum-mechanically. This study appropriately separates the entropy contributions into translational, rotational, and vibrational components that individually depend upon mass, as noted above. Therefore the net change in entropy upon binding appears to depend strongly upon mass. However, the degrees of freedom that need to be treated quantum-mechanically are precisely the "hard" degrees of freedom that are expected to be perturbed least upon noncovalent binding. Therefore, although changes in mass would alter the *individual* changes in translational, rotational, and vibrational entropy upon binding, the *net* dependence of these terms upon mass should again largely cancel, despite the quantum-mechanical treatment of the vibrations. The final results thus are likely to depend only weakly upon mass.

More generally, whether it is important to treat the dynamics of individual atoms quantum mechanically is a question that goes to the heart of the validity of classical simulations of molecular systems. Fortunately, quantum dynamical simulations of biomolecular systems suggest that the errors that result from assuming classical dynamics are modest (Zheng et al., 1988, 1989). Therefore, it is reasonable to construct models for binding based upon classical statistical thermodynamics. Still, for tight binding of some small ligands, such as light ions and water molecules, mass could be important. For these species, somewhat less entropy might be lost upon binding than predicted by classical statistical thermodynamics, because the quantization of vibrational motion limits the amount of entropy that can be lost upon binding. Also, substitution of deuterons for protons in the binding interface of two ligands is expected to strengthen hydrogen bonds somewhat (Chervenak and Toone, 1994). Nonetheless, the relative motions of the bulk of the molecules will not be significantly quantized because of their large mass. Therefore, the net molecular masses do not affect the binding constant.

Experimental data that shed light upon the possible mass-dependence of noncovalent binding are difficult to find, because differences in molecular mass are usually accompanied by substantial chemical differences at the binding interface. However, one could certainly imagine classes of systems, such as biotinylated proteins of various masses, that would permit these ideas to be tested experimentally.

## SUMMARY AND CONCLUSIONS

This paper reviews the classical statistical thermodynamics of noncovalent binding in solution and its relationship to commonly used computational models of binding. A special effort is made to define units and standard states clearly. In order to provide a basis for the discussion of specific computational methods, a fairly detailed derivation of the standard Gibbs free energy of binding is provided. The geometric definition of the bound complex is central to this derivation, and we follow previous authors in arguing that, for tight binding, the free energy depends only weakly upon the precise definition of the complex, so long as two conditions are met. The present derivation also makes explicit the influence of pressure.

It is emphasized that all dependence upon atomic or molecular mass has canceled in the final expression for the standard free energy of binding. It is the classical approximation that causes this cancellation to occur; we argue that this approximation is a good one for biophysical systems. Accordingly, models for the noncovalent binding of molecules in solution should not yield binding free energies that depend upon mass. Some appear to, however, and the reasons for this are analyzed.

The validity of the double-annihilation method (Jorgensen et al., 1988) for computing binding free energies has been questioned recently (Janin, 1996). Here, the method is reformulated, and it is argued that the method does not actually annihilate the ligand. Rather, it decouples the ligand from its surroundings, effectively creating an ideal-gas molecule. Obtaining a converged standard free energy of binding requires that this ideal-gas molecule be restrained by an artificial potential, and that a correction term be added to the computed free energy. The resulting “double-decoupling” method is consistent with and somewhat more general than existing methods (Hermans and Shankar, 1986; Zhang and Hermans, 1996; Roux et al., 1996) for computing the affinity of small molecules for protein cavities. These methods already account correctly for the standard state.

Implicit solvent models, in which some or all solvent degrees of freedom are suppressed, should allow more rapid calculations of binding free energies. Therefore, the theoretical basis for the use of such models is reviewed here. Treating parts of the solutes as rigid is another way of speeding binding calculations, and the conditions under which this is a good approximation are also reviewed.

Protonation equilibria are sometimes thermodynamically linked with binding equilibria, and accounting for this link-

age is important in such cases. Therefore, a binding polynomial formalism for treating the linkage between binding and protonation is presented.

Finally, a number of models yield binding free energies as sums of specific free energy components. However, it can be difficult to interpret the various components when they are not derived from the underlying statistical thermodynamics. The present derivation of the binding free energy is therefore used to define several important free energy components, including the commonly used translational and rotational entropy, configurational entropy, and solvent entropy. The present analysis also yields the following conclusions regarding the changes in entropy upon binding: 1) Ideal-gas equations may be used to compute changes in external entropy if the intermolecular potential of mean force is used in place of the intermolecular potential energy. 2) The change in external entropy upon binding is ambiguous for flexible molecules, because it depends upon an arbitrary choice of coordinate system. 3) The crystallographic thermal factors of a complex offer no direct information about the entropy change upon binding. 4) The concept of a change in cratic or mixing entropy upon binding lacks a well-defined theoretical basis.

## APPENDIX A: FREQUENTLY USED SYMBOLS

Here,  $B$  and  $C$  may replace  $A$  to indicate quantities appropriate to these species.

$\beta$	$(RT)^{-1}$
$C^\circ$	Standard concentration
$C_A$	Concentration of $A$ .
$\Delta G_{AB}^\circ$	Standard free energy of binding of $A$ and $B$ .
$\gamma_A$	Activity coefficient of $A$ in the reaction solvent.
$I(\zeta_B)$	Step function defining the complex.
$J_{\zeta_B}$	Jacobian determinant for the Eulerian external angles of $B$ .
$K_{AB}$	Equilibrium constant for binding of $A$ and $B$ .
$M_A$	The number of atoms in species $A$ .
$M_S$	The number of atoms in $N$ solvent molecules.
$\mu_{\text{gas},A}^\circ$	Standard chemical potential of $A$ as an ideal gas.
$\mu_{\text{sol},A}$	Chemical potential of $A$ in the reaction solvent.
$\mu_{\text{sol},A}^\circ$	Standard chemical potential of $A$ in the reaction solvent.
$\mu_s$	Chemical potential of the solvent.
$N$	The number of solvent molecules.
$N_{\mathcal{A}}$	Avogadro's number.
$P^\circ$	Standard pressure.
$\mathbf{p}_A$	The $3M_A$ momenta of $A$ in the laboratory frame.
$\mathbf{p}_S$	The $3N$ momenta of the solvent in the laboratory frame.
$Q_{N,A}$	Canonical partition function of $N$ solvents and $A$ , at volume $V_{N,A}$ .
$Q_{0,A}$	Gas-phase molecular partition function of $A$ .
$Q_{N,0}$	Canonical partition function of $N$ solvent molecules at volume $V_{N,0}$ .
$R$	The gas constant.
$\mathbf{R}_A$	The three laboratory-frame external Cartesian coordinates of $A$ .
$\mathbf{r}_A$	The $3M_A-6$ molecular-frame (internal) Cartesian coordinates of $A$ .
$\mathbf{r}'_A$	The $3M_A$ Cartesian coordinates of $A$ in the laboratory frame.

$\mathbf{r}_S$	The laboratory-frame Cartesian coordinates of $N$ solvent molecules.
$\mathbf{r}_a$	Those internal Cartesian coordinates of $A$ treated as mobile.
$\mathbf{r}_a$	Those internal Cartesian coordinates of $A$ treated as fixed.
$\mathbf{r}_a^*$	A specific fixed conformation of coordinates $\mathbf{r}_a$
$\sigma_A$	The symmetry number of $A$ .
$T$	Temperature (Kelvin).
$U$	Potential energy as a function of atomic coordinates.
$V_{N,A}$	Equilibrium volume of $N$ solvent molecules and one $A$ at standard pressure.
$V_{N,0}$	Equilibrium volume of $N$ solvent molecules at standard pressure.
$\bar{V}_A$	Partial molar volume of $A$ , $V_{N,A} - V_{N,0}$ .
$V_s$	Partial molar volume of solvent in the solvent ( <i>sic</i> ).
$W$	Solvation energy as a function of solute coordinates.
$w(\zeta_B)$	Potential of mean force for the interaction of $A$ with $B$ .
$\xi_{A,1}, \xi_{A,2}, \xi_{A,3}$	The three Eulerian angles specifying the orientation $A$ relative to the lab frame.
$\xi_A$	The six external coordinates of $A$ , $(\mathbf{R}_A, \xi_{A,1}, \xi_{A,2}, \xi_{A,3})$ .
$Z_A$	$Z_{N,A}/Z_{N,0}$ .
$Z_{N,A}$	Configuration integral in internal coordinates of $A$ and external coordinates of $N$ solvents at volume $V_{N,A}$ .
$Z_{N,0}$	Configuration integral of $N$ solvent molecules at volume $V_{N,0}$ .

## APPENDIX B: FORMULATION OF THE CHEMICAL POTENTIAL

This appendix demonstrates that Eq. 5 is equivalent to the constant-volume expressions of Widom (Widom, 1963). Equations 1 and 2 of Widom (1963) may be summarized as:

$$Q_N = V \langle \exp(-\Psi/RT) \rangle Q_{N-1} \quad (61)$$

$$n/z = \langle \exp(-\Psi/RT) \rangle. \quad (62)$$

Here, following the notation of Widom,  $Q_N$  is the configuration integral for a fluid of  $N$  identical particles in a volume  $V$  and at absolute temperature  $T$ ;  $n$  is the number density of particles;  $z$  is the activity of the particles, defined to approach  $n$  as  $n \rightarrow 0$ ; and  $\Psi$  is the potential energy of interaction with one particle with the other  $N - 1$  particles.

As noted in the Widom paper, these equations also can be applied to each molecular species in a mixture. Thus, we may consider the case of a single solute molecule mixed with  $N$  solvent molecules. Then Eq. 61 becomes

$$Q_{N,1} = V \langle \exp(-\Psi/RT) \rangle Q_{N,0} \quad (63)$$

where, as in the main text, the first subscript on  $Q$  indicates the number of solvent molecules in the configuration integral, and the second indicates the number of solutes. Note that both configuration integrals extend over the volume  $V$ . With the definition of the activity  $z \equiv \exp(-\mu/RT)$ , where  $\mu$  is the chemical potential of the solute, Eqs. 62 and 63 yield

$$\mu = -RT \ln \frac{Q_{N,1}}{nVQ_{N,0}} \quad (64)$$

The standard chemical potential,  $\mu^\circ$ , is the chemical potential for the usual hypothetical ideal solution at concentration  $C^\circ$ , and at standard pressure. The solution considered here will be highly dilute and effectively ideal so long as  $N \gg 1$ . Therefore, the standard chemical potential of the solute in

a hypothetical ideal solution at concentration may be written as

$$\mu^\circ = -RT \ln \frac{Q_{N,1}(V_{N,1})}{C^\circ V_{N,1} Q_{N,0}(V_{N,1})} \quad (65)$$

where, as in the body of this paper,  $V_{N,1}$  is the equilibrium volume of the system with  $N$  solvent molecules and 1 solute at the standard pressure  $P^\circ$ .

For  $N$  large, the addition of one solute molecule is a small perturbation of the system. Therefore,

$$RT \ln Q_{N,0}(V_{N,1}) = RT \ln Q_{N,0}(V_{N,0}) + [V_{N,1} - V_{N,0}] \left( \frac{\partial(RT \ln Q_{N,0})}{\partial V} \right)_{T,N} \quad (66)$$

and (McQuarrie, 1973)

$$\left( \frac{\partial(RT \ln Q_{N,0})}{\partial V} \right)_{T,N} = P^\circ, \quad (67)$$

when the derivative is evaluated for  $V = V_{N,0}$ . Therefore, Eq. 65 may be rewritten as

$$\mu^\circ = RT \ln \frac{Q_{N,1}(V_{N,1})}{C^\circ V_{N,1} Q_{N,0}(V_{N,0})} + P^\circ [V_{N,1} - V_{N,0}]. \quad (68)$$

This is the same as Eq. 5, which was to be shown.

*Note added in proof:* A paper submitted by J. Hermans and L. Wang also discusses the standard-state correction required in calculations of binding free energies by thermodynamic integration.

We are grateful to Drs. J. Hubbard, W. Jorgensen, P. Kollman, J. Moulton, K. Murphy, A. Roitberg, K. Sharp, A. Szabo, B. Tidor, and H.-X. Zhou for helpful discussions and insights, and to Drs. H. S. R. Gilson, P. Shenkin, J. Hermans, and two anonymous reviewers for valuable comments on the manuscript.

This work was supported by the National Institute of Standards and Technology, the National Institutes of Health, the National Science Foundation, and the National Science Foundation Supercomputer Centers.

## REFERENCES

- Ajay, and M. A. Murcko. 1995. Computational methods to predict binding free energy in ligand-receptor complexes. *J. Med. Chem.* 38:4953-4967.
- Alberty, R. A. 1994. Recommendations for nomenclature and tables in biochemical thermodynamics. *Pure and Appl. Chem.* 66:1641-1666.
- Andersen, H. C. 1973. Cluster expansion for hydrogen-bonded fluids. I. Molecular association in dilute gases. *J. Chem. Phys.* 59:4714-25.
- Andrews, P. R., D. J. Craik, and J. L. Martin. 1984. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* 27: 1648-1657.
- Appelt, K., R. J. Bacquet, C. A. Bartlett, C. L. J. Booth, S. T. Freer, M. A. M. Fuhry, M. R. Gehring, S. M. Herrmann, E. F. Howland, C. A. Janson, T. R. Jones, C. Chen Kan, V. Kathardekar, K. K. Lewis, G. P. Marzoni, D. A. Matthews, C. Mohr, E. W. Moomaw, C. A. Morse, S. J. Oatley, R. C. Ogden, M. R. Reddy, S. H. Reich, W. S. Schoettlin, W. W. Smith, M. D. Varney, J. E. Villafranca, R. W. Ward, S. Webber, S. E. Webber, K. M. Welsh, and J. White. 1991. Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* 34:1925-1934.
- Atkins, P. W. 1994. *Physical Chemistry*. Oxford University Press and W. H. Freeman, New York.

- Balbes, L. M., S. W. Mascarella, and D. B. Boyd. 1994. A perspective of modern methods in computer-aided drug design. *Rev. Comput. Chem.* 5:337–379.
- Bash, P. A., U. C. Singh, R. Langridge, and P. A. Kollman. 1987. Free energy calculations by computer simulation. *Science*. 236:564.
- Bashford, D., and M. Karplus. 1990. pKa's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry*. 9:327–335.
- Beglov, D., and B. Roux. 1994. Finite representation of an infinite bulk system: solvent boundary potential for computer simulations. *J. Chem. Phys.* 100:9050–9063.
- Ben-Naim, A. 1975. Solute and solvent effects on chemical equilibria. *J. Chem. Phys.* 63:2064–2073.
- Ben-Naim, A. 1992. *Statistical Thermodynamics for Chemists and Biochemists*, 1st ed. Plenum Press, New York. 677.
- Ben-Shaul, A., N. Ben-Tal, and B. Honig. 1996. Statistical thermodynamic analysis of peptide and protein insertion in lipid bilayers. *Biophys. J.* 71:130–137.
- Beveridge, D. L., and F. M. DiCapua. 1989. Free energy via molecular simulation: application to chemical and biomolecular systems. *Annu. Rev. Biophys. Chem.* 18:431–492.
- Böhm, 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. *J. Comput.-Aided. Molec. Design*. 8:243–256.
- Chandler, D. 1979. Effects of liquid structures on chemical reactions and conformational changes of non-rigid molecules in condensed phases. *Faraday Discuss. Chem. Soc.* 66:184–190.
- Chandler, D., and L. R. Pratt. 1976. Statistical mechanics of chemical equilibria and intramolecular structures of nonrigid molecules in condensed phases. *J. Chem. Phys.* 65:2925–2940.
- Chervenak, M. C., and E. J. Toone. 1994. A direct measure of the contribution of solvent reorganization to the enthalpy of ligand binding. *J. Am. Chem. Soc.* 116:10533–10539.
- Chothia, C., and J. Janin. 1975. Principles of protein-protein recognition. *Nature*. 256:705–708.
- Cohen, N. C., J. M. Blaney, C. Humblet, P. Gund, and D. C. Barry. 1990. Molecular modeling software and methods for medicinal chemistry. *J. Med. Chem.* 33:883–894.
- Colman, P. M. 1994. Structure-based drug design. *Curr. Opin. Struct. Biol.* 4:868–874.
- Cramer, C., and D. Truhlar. 1995. Continuum solvation models: classical and quantum-mechanical implementations. In *Reviews in Computational Chemistry*, Vol. 6, Chap. 1. VCH Publishers, New York. 1–72.
- Doty, P., and G. E. Myers. 1953. Low molecular weight proteins: thermodynamics of the association of insulin molecules. *Disc. Farad. Soc.* 13:51–58.
- Eisenberg, D., and A. D. McLachlan. 1986. Solvation energy in protein folding and binding. *Nature*. 319:199–203.
- Erickson, H. P. 1989. Cooperativity in protein-protein association. The structure and stability of the actin filament. *J. Mol. Biol.* 206:465–474.
- Finkelstein, A. V., and J. Janin. 1989. The price of lost freedom: entropy of bimolecular complex formation. *Prot. Engin.* 3:1–3.
- Flory, P. J. 1988. *Statistical Mechanics of Chain Molecules*. Hanser Publishers; Oxford University Press, New York.
- Gilson, M. K. 1993. Multiple-site titration and molecular modelling: two rapid methods for computing energies and forces for ionizable groups in proteins. *Proteins: Struct. Func. Gen.* 15:266–282.
- Gilson, M. K., J. A. McCammon, and J. D. Madura. 1995. Molecular dynamics simulation with a continuum electrostatic model of the solvent. *J. Comput. Chem.* 16:1081–1095.
- Glasstone, S. 1947. *Thermodynamics for Chemists*, van Nostrand, New York, p. 351.
- Go, N., and H. Scheraga. 1969. Analysis of the contribution of internal vibrations to the statistical weights of equilibrium conformations of macromolecules. *J. Chem. Phys.* 51:4751–4767.
- Greer, J., J. W. Erickson, J. J. Baldwin, and M. D. Varney. 1994. Application of the three-dimensional structures of protein target molecules in structure-based drug design. *J. Med. Chem.* 37:1035–1053.
- Groot, R. D. 1992. The association constant of a flexible molecule and a single atom: theory and simulation. *J. Chem. Phys.* 97:3537–3549.
- Gross, M., and R. Jaenicke. 1994. Proteins under pressure. The influence of high hydrostatic pressure on structure, function and assembly of proteins and protein complexes. *Eur. J. Biochem.* 221:617–630.
- Gurney, R. W. 1953. *Ionic Processes in Solutions*. McGraw-Hill, Inc., New York.
- Hermans, J., and S. Shankar. 1986. The free energy of xenon binding to myoglobin from molecular dynamics simulation. *Isr. J. Chem.* 27:225–227.
- Hill, T. L. 1955a. Theory of protein solutions. I. *J. Chem. Phys.* 23:623–636.
- Hill, T. L. 1955b. Theory of protein solutions. II. *J. Chem. Phys.* 23:2270–2274.
- Hill, T. L. 1956. *Statistical Mechanics. Principles and Selected Applications*. McGraw-Hill, New York.
- Hill, T. L. 1985a. *Cooperativity Theory in Biochemistry*, 1st ed. Springer Series in Molecular Biology. Springer-Verlag, New York. p. 8.
- Hill, T. L. 1985b. *Cooperativity Theory in Biochemistry*, 1st ed. Springer Series in Molecular Biology. Springer-Verlag, New York.
- Hill, T. L. 1986. *An Introduction to Statistical Thermodynamics*. Dover, New York.
- Holtzer, A. 1995. The cratic correction and other fallacies. *Biopolymers*. 35:595–602.
- Honig, B., K. Sharp, and A.-S. Yang. 1993. Macroscopic models of aqueous solutions: biological and chemical applications. *J. Phys. Chem.* 97:1101–1109.
- Horton, N., and M. Lewis. 1992. Calculation of the free energy of association for protein complexes. *Prot. Sci.* 1:169–181.
- Hoye, J. S., and K. Olausson. 1980. Statistical mechanical model with chemical reaction. *Physica*. 104A:435–446.
- Janin, J. 1995. Elusive affinities. *Proteins: Struct. Funct. and Gen.* 21:30–39.
- Janin, J. 1996. For Guldberg and Waage, with love and cratic entropy. *Proteins: Struct. Funct. Gen.* 24(4):i–ii [sic].
- Janin, J., and C. Chothia. 1978. The role of hydrophobicity in the binding of coenzymes. *Biochemistry*. 17:2943–2948.
- Jorgensen, W. L., J. K. Buckner, S. Boudon, and J. Tirado-Rives. 1988. Efficient computation of absolute free energies of binding by computer simulations. Application to methane dimer in water. *J. Chem. Phys.* 89:3742–3746.
- Justice, M.-C., and J.-C. Justice. 1976. Ionic interactions in solutions: I. The association concepts and the McMillan-Mayer theory. *J. Sol. Chem.* 5:543–561.
- Kang, Y. K., K. D. Gibson, G. Némethy, and H. A. Scheraga. 1988. Free energies of hydration of solute molecules. 4. Revised treatment of the hydration shell model. *J. Phys. Chem.* 92:4739–4742.
- Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* 14:1–63.
- Kollman, P. 1993. Free energy calculations—applications to chemical and biochemical phenomena. *Chem. Rev.* 93:2395–2417.
- Kumar, S. K., I. Szleifer, K. Sharp, P. J. Rossky, R. Friedman, and B. Honig. 1995. Size dependence of transfer free energies. 1. A Flory-Huggins approach. *J. Phys. Chem.* 99:8382–8391.
- Kuntz, I. D., E. C. Meng, and B. K. Shoichet. 1994. Structure-based molecular design. *Acc. Chem. Res.* 27:117–123.
- Laria, D., H. Corti, and R. Fernandez-Prini. 1990. The cluster theory for electrolyte solutions. *J. Chem. Soc. Farad. Trans.* 86:1051–1056.
- Lazaridis, T., and M. E. Paulaitis. 1993. Reply to the comment on "Entropy of hydrophobic hydration: a new statistical mechanical formulation." *J. Phys. Chem.* 97:5789–5790.
- Lewis, G., and M. Randall. 1961a. *Thermodynamics*, McGraw-Hill, 2 edition, Revised by K. S. Pitzer and L. Brewer. pp. 244–249.
- Lewis, G., and M. Randall. 1961b. *Thermodynamics*, McGraw-Hill, 2 edition, Revised by K. S. Pitzer and L. Brewer. pp. 307–308.
- Lifson, S., and I. Oppenheim. 1960. Neighbor interactions and internal rotations in polymer molecules. IV. Solvent effect on internal rotations. *J. Chem. Phys.* 33:109–115.

- Lybrand, T. 1990. Computer simulation of biomolecular systems using molecular dynamics and free energy perturbation methods. *In Reviews in Computational Chemistry*. Vol. 1. VCH Publishers, New York. 295–320.
- Madura, J. D., M. E. Davis, M. K. Gilson, R. C. Wade, B. A. Luty, and J. A. McCammon. 1994. Biological applications of electrostatic calculations and Brownian dynamics simulations. *In Reviews in Computational Chemistry*. Vol. 1. VCH Publishers, New York. 229–267.
- Marrone, T., M. Gilson, and J. McCammon. 1996. Comparison of continuum and explicit models of solvation-potentials of mean force for alanine dipeptide. *J. Phys. Chem.* 100:1439–1441.
- Marrone, T. J., J. M. Briggs, and J. A. McCammon. 1997. Structure-based drug design: computational advances. *Annu. Rev. Pharm. Tox.* In press.
- Matthew, J. B., F. R. N. Gurd, B. E. Garcie-Moreno, M. A. Flanagan, K. L. March, and S. J. Shire. 1985. pH-dependent processes in proteins. *Critical Reviews in Biochemistry*. 18:91–197.
- McMillan, W. G., and J. E. Mayer. 1945. The statistical thermodynamics of multi-component systems. *J. Chem. Phys.* 13:276–305.
- McQuarrie, D. A. 1973. *Statistical Mechanics*, Harper & Row, New York.
- Merz, Jr., K. M. 1991. CO<sub>2</sub> binding to human carbonic anhydrase II. *J. Am. Chem. Soc.* 113:406–411.
- Mitchell, M. J., and J. A. McCammon. 1991. Free energy difference calculations by thermodynamic integration: difficulties in obtaining a precise value. *J. Comput. Chem.* 12:271.
- Miyamoto, S., and P. A. Kollman. 1992. Molecular dynamics studies of calixspherand complexes with alkali metal cations: calculations of the absolute and relative free energies of binding of cations to a calixspherand. *J. Am. Chem. Soc.* 114:3668–3674.
- Miyamoto, S., and P. A. Kollman. 1993a. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbation approaches. *Proteins: Struct. Func. Gen.* 16:226–245.
- Miyamoto, S., and P. A. Kollman. 1993b. What determines the strength of noncovalent association of ligands to proteins in aqueous solution? *Proc. Natl. Acad. Sci. U.S.A.* 90:8402–8406.
- Mordasini Denti, T. Z., W. F. van Gunsteren, and F. Diederich. 1996. Computer simulations of the solvent dependence of apolar association strength: Gibbs free energy calculations on a cyclophane-pyrene complex in water and chloroform. *J. Am. Chem. Soc.* 118:6044–6051.
- Morild, E. 1981. The theory of pressure effects on enzymes. *Adv. Prot. Chem.* 34:93–163.
- Morton, A., W. A. Baase, and B. W. Matthews. 1995. Energetic origins of specificity of ligand binding in an interior nonpolar cavity of T4 lysozyme. *Biochemistry*. 34:8564–8575.
- Muller, K., H. Ludemann, and R. Jaenicke. 1981. Pressure-induced structural changes of pig LDH. *Biophys. Chem.* 14:101–110.
- Murphy, K. P., D. Xie, K. C. Garcia, L. M. Amzel, and E. Freire. 1993. Structural energetics of peptide recognition: Angiotensin II/antibody binding. *Proteins: Struct. Func. Gen.* 15:113–120.
- Murphy, K. P., D. Xie, K. S. Thompson, L. M. Amzel, and E. Freire. 1994. Entropy in biological binding processes: estimation of translational entropy loss. *Proteins: Struct. Func. Gen.* 18:63–67.
- Novotny, J., R. E. Bruccoleri, and F. A. Saul. 1989. On the attribution of binding energy in antigen-antibody complexes MCPC 603, D1.3 and HyHEL-5. *Biochemistry*. 28:4735–4749.
- Page, M. I., and W. P. Jencks. 1971. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. USA.* 68:1678–1683.
- Pitzer, K. S. 1946. Energy levels and thermodynamic functions for molecules with internal rotation. II. Unsymmetrical tops attached to a rigid frame. *J. Chem. Phys.* 14:239–243.
- Pitzer, K. S., and D. R. Schreiber. 1987. The restricted primitive model for ionic fluids. Properties of the vapor and the critical region. *Mol. Phys.* 60:1067–1078.
- Pranata, J., and W. L. Jorgensen. 1991. Monte-Carlo simulations yield absolute free energies of binding for guanine-cytosine and adenine-uracil base pairs in chloroform. *Tetrahedron*. 47:2491–2501.
- Prue, J. E. 1969. Ion pairs and complexes: free energies, enthalpies, and entropies. *J. Chem. Ed.* 46:12–16.
- Rosenfeld, R., S. Vajda, and C. DeLisi. 1995. Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.*
- Roux, B., M. Nina, R. Pomes, and J. C. Smith. 1996. Thermodynamic stability of water molecules in the Bacteriorhodopsin proton channel: a molecular dynamics free energy perturbation study. *Biophys. J.* 71: 670–681.
- Schade, B. C., H. Ludemann, and R. Jaenicke. 1980a. Reversible high-pressure dissociation of LDH from pig muscle. *Biochemistry*. 19: 1121–1126.
- Schade, B. C., H. Ludemann, R. Rudolph, and R. Jaenicke. 1980b. Kinetics of reconstitution of porcine muscle LDH after reversible high pressure dissociation. *Biophys. Chem.* 11:257–263.
- Schellman, J. A. 1975. Macromolecular binding. *Biopolymers*. 14: 999–1018.
- Searle, M. S., D. H. Williams, and U. Gerhard. 1992a. Partitioning of free energy contributions in the estimation of binding constants: residual motions and consequences for amide-amide hydrogen bond strengths. *J. Am. Chem. Soc.* 114:10697–1070.
- Searle, M. S., D. H. Williams, and U. Gerhard. 1992b. Partitioning of free energy contributions in the estimation of binding constants: residual motions and consequences for amide-amide hydrogen bond strengths. *J. Am. Chem. Soc.* 114:10697–10704.
- Shoup, D., and A. Szabo. 1982. Role of diffusion in ligand binding to macromolecules and cell-bound receptors. *Biophys. J.* 40:33–39.
- Simonson, T., and A. T. Brunger. 1994. Solvation free energies estimated from macroscopic continuum theory: an accuracy assessment. *J. Phys. Chem.* 98:4683–4694.
- Sneddon, S. F., D. J. Tobias, and C. L. Brooks III. 1989. Thermodynamics of amide hydrogen bond formation in polar and apolar solvents. *J. Mol. Biol.* 209:817–820.
- Spolar, R. S., and M. T. Record, Jr. 1994. Coupling of local folding to site-specific binding of proteins to DNA. *Science*. 263:777–784.
- Steinberg, I. Z., and H. A. Scheraga. 1963. Entropy changes accompanying association reactions of proteins. *J. Biol. Chem.* 238:172–181.
- Stell, G. 1976. Correlation functions and their generating functionals. *In Phase Transitions and Critical Phenomena*. Domb, C. & Green, M. S., editors. Volume 5b, p 205, Academic Press, London.
- Stell, G., and Y. Zhou. 1989. Chemical association in simple models of molecular and ionic fluids. *J. Chem. Phys.* 91:3618–3623.
- Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.
- Stouten, P. F. W., C. Frömmel, H. Nakamura, and C. Sander. 1993. An effective solvation term based on atomic occupancies for use in protein simulations. *Molecular Simulation*. 10:97–120.
- Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. *Annu. Rev. Phys. Chem.* 43:407–435.
- Tanford, C., and J. G. Kirkwood. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J. Am. Chem. Soc.* 79: 5333–5339.
- Tembe, B. L., and J. A. McCammon. 1984. Ligand-receptor interactions. *Comput. Chem.* 8:281–283.
- Tidor, B., and M. Karplus. 1994. The contribution of vibrational entropy to molecular association, the dimerization of insulin. *J. Mol. Biol.* 238: 405–414.
- van Eldik, R., T. Asano, and W. J. Le Noble. 1989. Activation and reaction volumes in solution. *Chem. Rev.* 89:549–688.
- Varnek, A. A., G. Wipff, A. S. Glebov, and D. Feil. 1995. An application of the Miertus-Scrocco-Tomasi solvation model in molecular mechanics and dynamics simulations. *J. Comput. Chem.* 16:1–19.
- von Freyberg, B., T. J. Richmond, and W. Braun. 1993. Surface area included in energy refinement of proteins. A comparative study on atomic solvation parameters. *J. Mol. Biol.* 233:275–292.
- Wang, J., and E. O. Purisima. 1996. Analysis of thermodynamic determi-

- nants in helix propensities of nonpolar amino acids through a novel free energy calculation. *J. Am. Chem. Soc.* 118:995–1001.
- Warshel, A., H. Tao, M. Fothergill, and Z.-T. Chu. 1994. Effective methods for estimation of binding energies in computer-aided drug design. *Isr. J. Chem.* 34:253–256.
- Weist, A. O., and E. D. Glandt. 1994. Equilibrium polymerization and gelation. I. Integral-equation theory. *J. Chem. Phys.* 101:5167–5171.
- Weng, Z., S. Vajda, and C. DeLisi. 1996. Prediction of protein complexes using empirical free energy function. *Prot. Sci.* 5:614–626.
- Wertheim, M. S. 1984. Fluids with highly directional attractive forces. I. Statistical thermodynamics. *J. Stat. Phys.* 35:19–34.
- Widom, B. 1963. Some topics in the theory of fluids. *J. Chem. Phys.* 39:2808.
- Widom, B. 1982. Potential-distribution theory and the statistical mechanics of fluids. *J. Phys. Chem.* 86:869–872.
- Wong, C., and J. McCammon. 1986. Dynamics and design of enzymes and inhibitors. *J. Am. Chem. Soc.* 108:3830–3832.
- Wyman, J. 1965. The binding potential, a neglected linkage concept. *J. Mol. Biol.* 11:631–644.
- Yang, A.-S., and B. Honig. 1993. On the pH dependence of protein stability. *J. Mol. Biol.* 231:459–474.
- Yu, H.-A., and M. Karplus. 1988. A thermodynamic analysis of solvation. *J. Phys. Chem.* 89:2366–2379.
- Zhang, L., and J. Hermans. 1996. Hydrophilicity of cavities in proteins. *Proteins: Struct., Func., Gen.* 24:433–438.
- Zheng, C., C. F. Wong, J. A. McCammon, and P. G. Wolynes. 1988. Quantum simulation of ferricytochrome C. *Nature* 334:726–728.
- Zheng, C., C. F. Wong, J. A. McCammon, and P. G. Wolynes. 1989. Classical and quantum aspects of ferricytochrome C. *Chem. Scripta* 29A:171–179.
- Zhou, Y., and G. Stell. 1995. Chemical association in simple models of molecular and ionic fluids. IV. New approximation for the cavity function and an application to the theory of weak electrolytes. *J. Chem. Phys.* 102:8089–8093.