

JOURNAL OF MEDICINAL CHEMISTRY

© Copyright 1995 by the American Chemical Society

Volume 38, Number 26

December 22, 1995

Perspective

Computational Methods to Predict Binding Free Energy in Ligand-Receptor Complexes

Ajay* and Mark A. Murcko

Vertex Pharmaceuticals Incorporated, 40 Allston Street, Cambridge, Massachusetts 02139-4211

Received June 30, 1995

Introduction

Most biological processes depend on the ability of molecules to bind and discriminate between one another. The ability to form noncovalent complexes is ubiquitous in reactions among biological macromolecules. This association is essential for many functions,¹ e.g., catalysis, transport, or signal transduction.

Structural information gathered over the past three decades has helped us understand many complexes at the molecular level providing a detailed description of the interaction between a receptor and its ligands. Binding can be remarkably sensitive to even small differences in structure. A quantitative knowledge of the binding process is therefore essential to understanding molecular recognition. Such a quantitative knowledge requires (1) a detailed understanding of the physical forces involved in the interaction and (2) a measure of the extent to which these forces contribute to the overall reaction procedure.

Thermodynamics governs the basic physical principles of molecular recognition. Therefore, the fundamentals processes involved in the binding of two or more molecules are similar to those for the folding of proteins. Binding, however, should be a simpler problem compared to folding as a smaller number (10–30) of amino acids, mostly in the active site, are involved. The affinity of two molecules that form a noncovalent complex is described by the change in enthalpy and entropy of the system. The system consists of the molecules and solvent (before complex formation) and the complex and solvent (see Figure 1). In general,

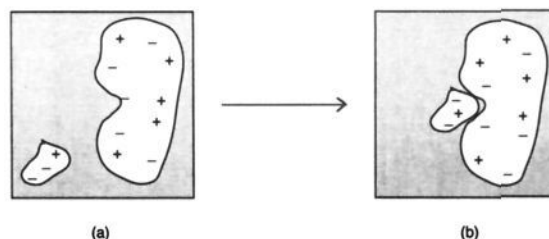


Figure 1. Process of general complex formation in solution. The “+” and “-” signs emphasize the importance of electrostatics, accurate evaluation of which is one of the major hurdles to accurate binding free energy predictions.

molecular association depends on ionic strength and pH of the solution.

In this perspective we will review some of the recent attempts at predicting binding affinities. Our primary focus is on noncovalent interactions between molecules; hence the binding of aspirin, for example, will not be considered. (Molecular mechanics—one of the fundamental tools for theoretical calculations—cannot treat the formation/breaking of covalent bonds, *i.e.*, the electronic structure, correctly. Hence, the binding energy of covalent inhibitors will be poorly represented.) We do not cover attempts to predict binding energies based on semiempirical quantum calculations in this perspective.² Another aspect of ligand binding that we do not address is metal ion binding. Ion binding can often be important for the stability and function of a variety of proteins. For zinc-binding proteins, it appears that sequence homology is sufficient to determine the binding domain.³ However, different calcium-binding proteins reveal strikingly different folds.⁴ Theoretical calculations on ion binding are plagued by very drastic approximations of hydration effects.

* Author to whom correspondence should be addressed. Tel: 617-499-0522. Fax: 617-576-2109. E-mail: ajay@vpharm.com.

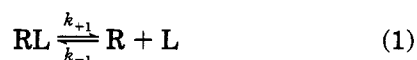
Almost all of the methods reviewed here use structural information on the complex. The structure is known either through X-ray or NMR methods or it can be modeled on the computer (of course, experimentally determined structures are preferable). As such the binding prediction methods form an integral part of structure-based drug design protocols. Examples of such structure-based design include HIV-protease⁵ for anti-AIDS agents, thymidylate synthase⁶ for anti-cancer agents, trypanosomal GAPDH⁷ for anti-parasitic agents, elastase⁸ for the treatment of emphysema, and glycogen phosphorylase⁹ for the treatment of diabetes.

One of the first groups to work on elucidating the forces involved in binding was that of Jencks.^{10,11} Their primary motivation was to delineate the entropic contributions to binding, namely, rotational and translational entropy losses and the loss in entropy on freezing bond rotations. Janin and Chotia¹² also contributed to understanding binding and protein folding using both entropic contributions and surface area burial (in an attempt to account for the "hydrophobic forces" involved in the reactions).

We begin with a brief overview of the concepts involved in binding, namely, the binding constant and free energy. Next the different contributions to binding free energy are described along with a set of assumptions that are used in their calculation. The following section is devoted to a brief review of the free energy perturbation method; this is the only method with an absolutely sound basis in statistical mechanics. The remainder of the paper introduces the recent (approximate) attempts at describing and delineating the components of the free energy that govern binding. Throughout our description we concentrate on the assumptions and, hence, some of the limitations of the methods. We conclude this perspective with a look at the immediate future of the prediction methods.

Binding Affinity

Protein–ligand interactions involve physical contact between the ligand and the receptor for a certain period of time. These contacts are specific and result in an attractive force. The receptor–ligand interactions are, in general, dependent on the concentration of the ligand, protein and salts in the solution.



represents a chemical interaction. Here, R is the receptor, L the ligand, RL the receptor–ligand complex, k_{-1} is the association rate constant for the reaction R + L going to RL, and k_{+1} is the dissociation rate constant for the reaction RL going to R + L.

An equilibrium constant (the dissociation equilibrium constant) can now be defined as

$$K_d = \frac{[\text{R}][\text{L}]}{[\text{RL}]} \quad (2)$$

Note that the units of K_d are moles/liter, which is a practically useful unit. Therefore, K_d is the equilibrium constant used for expressing binding affinity; the binding constant is usually expressed as a dissociation constant. When the magnitude of K_d is small, the

Table 1. Importance of Small Differences in Binding Energy (kcal/mol) for Binding Constant Values at Room Temperature

change in binding energy	change in binding constant	change in binding energy	change in binding constant
0.5	2×	2.0	29×
1.0	5×	2.5	68×
1.5	13×	3.0	158×

tendency of RL to dissociate is small, *i.e.*, RL tends to remain as a complex.

A fundamental thermodynamic equation relates free energy change to changes in enthalpy and entropy,

$$\Delta G = \Delta H - T\Delta S \quad (3)$$

where ΔG is the change in free energy of a reaction, ΔH and ΔS are the corresponding changes in enthalpy and entropy, and T is the temperature of the system. The equilibrium constant can be related to the free energy change of the dissociation of RL¹³ as,

$$\Delta G = \Delta G^\circ - RT \ln K_d \quad (4)$$

Here, G is the free energy, ΔG is the change in free energy for the reaction, R is the gas constant, and T is the absolute temperature. ΔG° is the free energy change associated with the reaction under standard conditions (all reactants and products are present at 1 M concentration, $T = 298$ K, and pressure is 1 atm). At equilibrium we have $\Delta G = 0$, therefore,

$$\Delta G^\circ = RT \ln K_d \quad (5)$$

For model (ideal) systems, K_d is only a function of temperature, hence the name equilibrium constant.

The following points must be kept in mind with respect to the above discussion:

(1) There is something not quite right about eq 5. Equation 2 shows that K_d has units of moles/liter and eq 5 seems to assume that K_d is unitless. The apparent contradiction is resolved by noting that eq 5 should really read as

$$\Delta G^\circ = RT \ln \frac{K_d}{c^\circ} \quad (6)$$

where c° is the ratio of the activities of each chemical species in the standard state.¹⁴ Therefore, ΔG° depends on c° , the standard state, and so when comparing ΔG° values we should be aware of the standard states used in the calculations.

(2) To be accurate, the quantities $[X]$ ($X = \text{R}, \text{L}, \text{RL}$) in eq 2 are only equal to concentration for ideal solutions. In real systems, concentrations have to be replaced by activities. However, because of the very dilute solutions in which these measurements are made, the conditions of ideality are closely approximated, and hence the differences are often ignored.

(3) The error in K_d measurements is usually about 10–20%, implying an error in ΔG° of 0.1–0.25 kcal/mol at room temperature (see Table 1). Experimental data on enthalpies and heat capacities usually have larger errors.

(4) Some authors provide thermodynamic parameter values for association and not dissociation. This results in a change of sign of all the numerical values given in this paper.

(5) An important point to note is that ΔH is independent while both ΔG and ΔS are dependent on the definition of standard state.¹³

In the rest of the paper, we drop the superscript "°" from all thermodynamic quantities, with the implicit understanding that they always refer to standard states.

It is instructive to note the quantitative relationship between a particular change in ΔG_{bind} and the corresponding change in K_i (see Table 1). Notice that a prediction of ΔG with an error of 2 kcal/mol (not considered unreasonable in most methods) implies an error of about 2900% in K_i at room temperature!

Measurement of Binding Affinity

K_i is measured in kinetic experiments. An important issue to keep in mind regarding K_i measurements is that experiments that involve both the substrate and the inhibitor are not equilibrium measurements unless steady state kinetics is assumed. More direct K_i measurements (usually called a K_d measurement) that do not involve competition with the substrate, however, are equilibrium measurements. All theoretical calculations strive to calculate the ensemble averaged free energy change. Therefore, all comparisons with experimental numbers should take place when the experiments are done in equilibrium conditions. This condition is often ignored, but this may not be inappropriate given the current limitations in the theoretical calculations and general difficulties with the experiments.

The interaction of the ligand with a protein may also be measured in terms of the IC_{50} value.¹⁵ The IC_{50} measures the concentration of the inhibitor required to reduce the binding of a ligand (or rate of reaction) by half. The IC_{50} value is, however, not very suitable for theoretical studies as it depends on the amount of ligand available to the receptor, and this makes comparisons between data obtained under different conditions impossible. The binding constants (K_i , K_d) on the other hand can be compared more easily. In principle, $K_i = K_d$. Determination of binding constants require more data than is needed for the calculation of the IC_{50} value. As mentioned, the binding constant is usually expressed in terms of a dissociation constant. The K_d is defined as the ligand concentration at which 50% of the receptor sites are occupied in a one to one complex. IC_{50} values are thus not true inhibition constants. Therefore, the ratio of two IC_{50} values for two enzymes are equivalent to K_i ratios *only* when the assays are performed under the same conditions and the enzymes have the same K_s (Michaelis constant) values. In general, quantification of inhibition is complicated by small values of K_s , substrate inhibition, low values of K_i , multiple conformations of the protein, and rate transients during assays.

Inhibitors bind to enzymes in many ways. Binding may be reversible or irreversible; further, the ligand may bind competitively, noncompetitively, or uncompetitively with respect to the substrate. If the ligand competes with the substrate for active site binding, then it is *competitive inhibition*. If the ligand binds at a site other than the substrate binding site (resulting in conformational change in the enzyme and hence inhibition of substrate turnover), then it is *noncompetitive inhibition*. If the ligand binds to the enzyme-substrate complex and not to the enzyme it is called *uncompetitive*

inhibition. In principle, the binding constants in all these processes are independent of ligand and/or substrate concentration.¹⁶ So comparisons across experiments can be reliably made.

Thermodynamics of Binding

The quantity of interest in determining binding constants is the free energy difference between the bound and the free states

$$\Delta G = G^c - G^u, \quad (7)$$

where the superscripts "c" and "u" are for complexed and uncomplexed, respectively. (We will use superscripts to designate the relevant part of the system, and subscripts to designate the nature of interaction.)

The calculation of free energy of binding is not yet an exact science. One of the basic steps that is often used is the factorization of the binding free energy into components:

(1) Overall loss in entropy due to association. That is, the loss in translational and rotational entropy (ligand and receptor are thought of as rigid entities). The changes in vibrational degrees of freedom also contribute.

(2) "Hydrophobic energy", *i.e.*, the entropy gain of water due to the binding of the ligand. The optimization of molecular interactions after a hydrocarbon has been removed from water has less entropic cost than the optimization of molecular interactions with the hydrocarbon in water.

(3) The entropy loss in both the enzyme and the ligand due to rotational constraints about single bonds on complex formation.

(4) The interaction energy between ligand and protein. Crystallographically identified water molecules are often integral to this. A major portion of the interaction energy is electrostatic. One important factor in a quantitative evaluation of the electrostatic interaction strength is the microscopic dielectric constant, which is almost never known. The greater the dielectric constant, the smaller the strength of the electrostatic interactions. Many approximations are made for the dielectric constant for the interior of the protein.

(5) Changes in steric interaction energy on binding.

(6) Change in conformational energy of the receptor and ligand upon binding.

The role of crystallographically identified water molecules, or water molecules with a large mean residence time in NMR experiments,¹⁷ is often very difficult to interpret, but these water molecules can play an important role in determining binding affinity. For example, it has been shown recently that noncovalent extensions of DNA bases by water molecules serve as selective recognition sites for specific protein-DNA interactions.¹⁸

The basic assumption in most of the work that is reviewed in this article is that different contributions to free energy of binding can be calculated separately and that they are additive. For example, we could separate out the electrostatic and nonpolar contributions, calculate each and combine the results to obtain the net change in free energy. There are other ways to dissect the interaction energies. Researchers have adopted different approaches based on the importance

of specific contributions to binding in the particular system under study and due to differences in the nature of these interactions.

The usual practice (following the model adopted in force fields) is to write the free energy of binding as an additive interaction of different parts. We call this the "master equation". We do this in the series of equations below by first writing it out conceptually.

$$\Delta G_{\text{bind}} = \Delta G_{\text{solvent}} + \Delta G_{\text{conf}} + \Delta G_{\text{int}} + \Delta G_{\text{motion}} \quad (8a)$$

This accounts for the contributions due to (1) the solvent, (2) conformational changes in the protein and the ligand, (3) the specific protein–ligand interactions that comes from their proximity, and (4) the "motion" in the protein and ligand once they are proximal.

$$\Delta G_{\text{solvent}} = \Delta G_{\text{hyd}} \quad (8b)$$

ΔG_{hyd} is the hydration free energy.

$$\Delta G_{\text{conf}} = \Delta G^{\text{r}} + \Delta G^{\text{l}} \quad (8c)$$

ΔG^{r} is the change in free energy of the receptor on complex formation, and ΔG^{l} is the same quantity for the ligand

$$\Delta G_{\text{int}} = \Delta G^{\text{r-l}} \quad (8d)$$

$\Delta G^{\text{r-l}}$ is the change in free energy due to specific (electrostatics and van der Waals) interactions between the ligand the receptor.

$$\Delta G_{\text{motion}} = \Delta G_{\text{rot}} + \Delta G_{\text{tr}} + \Delta G_{\text{vib}} \quad (8e)$$

ΔG_{rot} is the free energy contribution due to the freezing of the internal rotations of the receptor and ligand, ΔG_{tr} is the change in translational/rotational free energy, and ΔG_{vib} is the change in vibrational free energy due to complexation.

To reiterate, this is just *one* of many ways in which the free energy of binding can be partitioned. Care should be taken, however, to avoid double counting the contribution of certain interactions; these complications can arise when accounting for hydration interactions (see below).

To simplify the analysis, many assumption are usually made. We will detail them below. Parenthetically, we note that the assumptions used in free energy perturbation methods are of a different kind from those listed below.

(1) The first point to note is that, strictly, all the quantities in eq 8 should be ensemble averages as the complex, the free protein, and the ligand are dynamic entities. Flexibility of the molecules is often required for biological activity. However, the most common assumption is that ensemble averages are replaced by values corresponding to a single stable structure in evaluating the terms in eq 8.

(2) The ligand–protein complex itself is almost always assumed to be a unique structure, that is, it is assumed that the dynamic nature of the complex will not significantly affect binding constants.

(3) The receptor before complexation is also assumed to be a unique structure, that is, ΔG^{r} , and $\Delta G^{\text{r-l}}$, are the change in free energy on going from a single

uncomplexed receptor conformation to a single complexed receptor conformation.

(4) The ligand is sometimes assumed to be rigid. If the ligand is considered flexible then both ΔG^{l} and ΔG_{hyd} should involve conformational averages.

(5) ΔG_{rot} is usually approximated solely by entropy changes. In principle, enthalpy does make some contribution. But, it appears that the majority of the rotational free energy contribution derives from the changes in entropy and hence enthalpic contributions are ignored.

(6) ΔG_{tr} is often assumed to be a constant as it varies slowly with mass or moment of inertia. The entropic contribution to this comes from the loss of 6 degrees of freedom (3 translations and 3 rotations each for the receptor and ligand reduces to 6 degrees of freedom for the complex). The enthalpy loss comes mainly from the loss in the 3 degrees of rotational freedom.

(7) New vibrational modes created on complex formation are either ignored or approximated in a somewhat ad hoc manner.

(8) The errors and approximations in the force fields (CHARMM,¹⁹ AMBER,²⁰ ECEPP²¹) used for estimating the enthalpic and entropic (through Monte Carlo/molecular dynamics simulations) contributions are ignored.

(9) Strictly, there is a difference between enthalpy and energy. The additional $P\Delta V$ term is, however, negligible in solution.

At this point, it is important to point out that the term "energy" is used loosely, and often imprecisely, by different authors at different times. Free energy is the total energy of the system and includes both an enthalpic and entropic component (see eq 3). Some experiments yield the free energy of a system or a chemical process, while other experiments give energy or enthalpy. (Spectroscopic and calorimetric experiments are notable examples of the latter case.) Often, this distinction is overlooked by theoreticians. For example, molecular mechanics ("forcefield") calculations do not generally include the effects of zero-point energy, vibrational effects, or entropy; they should best be thought of as energies rather than enthalpies or free energies. However, the parameters that are developed for these forcefields are derived from fitting experimental data which are enthalpies or free energies! Thus the forcefield parameters implicitly account for these effects in some manner.

The different terms in eq 8 arise from both enthalpy and entropy contributions. The ΔG^{r} and ΔG^{l} terms are unfavorable for binding as, in general, the conformational enthalpy of the uncomplexed receptor and ligand will often be lower than that of the complex. The term, $\Delta G^{\text{r-l}}$ favors binding and is entirely enthalpic. The solvation energy (ΔG_{hyd}) is usually favorable for binding at room temperature. The rotational constraints in the term ΔG_{rot} , arising on binding, is unfavorable primarily because of the entropic cost of "freezing" the molecule into its bound conformation. However, there is both an entropic and enthalpic part to this, and some of the enthalpic release of heat to the surroundings can be taken up by new vibrational modes (ΔG_{vib}) in the complex. ΔG_{tr} , which gives the change in translational and rotational free energy, is also unfavorable to binding. The vibrational contribution, ΔG_{vib} , is favorable to binding (see below).

Table 2. Brief Overview of Some of the Recent Work on Predicting the Binding Free Energy

"Master Equation" Based Methods	
Honig <i>et al.</i> ^{50,52}	total free energy change is sum of electrostatic change and hydrophobic interactions electrostatics obtained from solving Poisson-Boltzmann equation hydrophobic interactions calculated based on surface area
Novotny and others ^{38,60}	fixed receptor fixed ligand surface area based hydrophobic cost molecular mechanics based electrostatic interaction energy (no vdW contribution) constant entropic conformational penalty constant translation/rotation penalty cratic entropy correction
Vajda <i>et al.</i> ⁶²	fixed receptor ensemble average of ligand conformations molecular mechanics interaction energy (no vdW contribution) conformational entropy cost based on X-ray data ASP-based hydration free energies constant T/R contribution no explicit vibrational contribution
Williams ^{61,65} and others ⁶⁴	fixed receptor fixed ligand vdW interactions for hydrocarbons and intrinsic binding constants for polar groups entropic loss for freezing rotors—a constant constant for translation/rotation surface area based hydration free energy the Andrews work accounts for the terms differently from that of Williams' group
Regression Methods	
Horton ⁴⁹	interaction energy calculated by an extension of the Eisenberg and McLachlan method to separate out atoms involved in hydrogen bonding from those not involved constant rotational/translational penalty
Böhm ⁶⁶	uses multiple linear regression to estimate the contributions of hydrogen bonding, ionic bonding, surface area buried, number of rotatable bonds in the ligand, and a constant; parameters estimated from crystal structure and experimental K_i values
Grootenhuis, ⁷⁸ Holloway, ⁷⁹ Ortiz ⁸⁰	use multiple linear regression based on a molecular mechanics potential function; the details of the terms included and regression methods differ substantially among the three groups
Head <i>et al.</i> ⁸³	a similar approach to Böhm; use different set of terms for the regression, a larger database, and two different regression techniques

Nonpartitioning Methods: Free Energy Perturbation

Free energy perturbation (FEP) is the only method that, in its basic formulation, does not start by partitioning the free energy of binding into different parts (as was done in the "master equation" above). The basic idea derives from the statistical mechanics, and it relates the free energy of a system and the ensemble average of an energy function that describes the system. There have been many excellent reviews of this method in the past few years (see Kollman²² and references therein), and so we will not go into many details. The method has been reasonably successful in rationalizing many experimental observations and in predictions for ligands binding to proteins²³ and less often for DNA binding ligands.²⁴ This is the only method that even attempts to deal seriously with calculating ensemble averages. In addition, it treats solvent molecules and ions explicitly.

In principle, the problems with free-energy perturbation arise from three factors: sampling difficulties, errors in the force field, and many "adjustable" parameters.

Sampling difficulty is effectively a technical problem in the sense that with better algorithms and faster computers the major problem of keeping the system in equilibrium during the simulation can be achieved. A

reasonable guess is that it will take about a decade to eliminate sampling difficulties in most routine problems. The force field errors (these are most likely small compared to sampling errors in most systems that are analyzed today) are more fundamental in nature and reflect, among other things, our inadequate understanding of electrostatics in water. However, progress is being made on this front. One promising direction is the technique of using Ewald sums^{25,26} to calculate electrostatic contributions without truncation. Another is the attempt to incorporate polarizability through nonadditive potential functions.²⁷ Another phenomenon that requires study is the coupling between the force fields of the explicit solvent and the solute.²⁸ Torsion parameters and point charges account for a majority of the "adjustable" parameters in FEP calculations.²⁹ Of the three, the sampling problem is the most important.

Sampling problems and difficulties with parameterization (of atom and bond types and atomic charges) make the routine use of free energy perturbation methods difficult. However, this is a very promising method with immense potential.

Individual Terms in the "Master Equation"

In this section we will consider each term in the "master equation" (eq 8) explaining the methods adopted to evaluate them. Table 2 summarizes the approach of the different groups.

Our discussion of the "master equation" approach follows eq 8. First we look at the methods adopted to account for the interactions of water molecules. Next we explore methods to account for the conformational contribution to binding. This has two parts: (1) conformational changes in the protein and (2) conformational changes in the ligand. The following section accounts for the interactions between the protein and the ligand once they are in close proximity to one another. Finally, we concentrate on accounting for molecular motion once the ligand and protein are close to one another.

Estimating Solvation Contributions

All protein–ligand interactions compete with interactions with water due to the fact that both the protein and the ligand are solvated before complexation and they lose their solvation shell on complex formation. But, the role of water molecules has, to date, been very difficult to characterize accurately. The properties of water do not make it amenable to simple pairwise additive interactions.³⁰ This has been one of the sources of theoretical difficulty. Both the bulk (long-range hydrophobic effect and dielectric shielding) and specific (hydrogen bonding) interactions of water are important in elucidating binding. This section is devoted to an overview of some of the attempts to relate the bulk properties of water with a description at the molecular level.

Water will modify both the electrostatic and dispersion interactions that are observed in the gas phase. The total free energy of interaction between two molecules can be written as

$$\Delta G(r) = U(r) + \Delta G_s(r) \quad (9)$$

Here $U(r)$ is the direct interaction in the absence of solvent and $\Delta G_s(r)$ is the contribution induced by the solvent.

High-level quantum mechanical (gas phase) calculations for dispersion and electrostatics forces are reasonably accurate with errors probably in the range of 5–25%. However, for theoretical calculations on the influence of water, there are disagreements even in the sign of the effect. The effects of approximations are most egregious in estimating solvation contributions.

(i) Interpreting the Results of Transfer Experiments. Almost all of the experimental knowledge of solvation is obtained by the so-called "transfer experiments" where changes in free energy, enthalpy, entropy, and heat capacity is measured on transferring compounds from organic liquids to water (experiments have also been done with vacuum to water transfer and solid phase to water). There is a large literature in this area; e.g., see Fauchere & Pliska.³¹ Interestingly, over 30 different scales (based on different model systems) are well correlated,³² suggesting that all of them capture similar information. Another method to evaluate the hydrophobic contributions to the stability of proteins is by introducing conservative point mutations which replace large hydrophobic side chain by smaller ones. The difference in stability can be related to free energy change. Interestingly, extraction of atomic solvation parameters (see below) from these methods yield numbers twice as large as those from octanol/water transfer experiments. Attempts to resolve the discrepancy have

not been very successful (see Lee³³ and references therein). One reason for the discrepancy may be that octanol is not a good model for protein interiors.

There are difficulties in interpreting these experiments. For example, the amino acid analog or small organic molecule is assumed to have the same conformation in the two phases, and there is dispute over whether or not (and if so, how?) to account for the entropy of mixing (also known as the Gibbs paradox). This controversy involves the extraction of solvation energies from transfer experiments: see Sharp *et al.*³⁴ and Chan and Dill³⁵ for details.

There are many different methods that have been adopted to account for solvation free energies. For a recent review of some of the ways solvation is modeled see Smith and Pettitt³⁶ and references therein. We will outline what we consider to be the most promising of these.

(ii) Surface Area Based Calculations. It has been observed empirically that in any association process there appears to be a corresponding large positive change in ΔC_p (the change in heat capacity). Both ΔC_p and ΔS_{solv} (the change in entropy of hydration) can be described within experimental error by changes in water-accessible surface area. The relationships usually take two forms. The first³⁷ (see eq 15 in the appendix) relates the entropy from the hydrophobic effect to the nonpolar surface area buried. It says that the complex is stabilized by a constant amount of 0.03 kcal/mol for every square angstrom of nonpolar area that is buried. A second (see eq 16 in the appendix) variant simply relates the hydration free energy to contact surface area on binding.³⁸

One of the confusing aspects of using surface area based methods is that it is not always clear in the published papers which buried surface area is being measured: (1) just the protein, (2) just the ligand, or (3) both the ligand the protein. From our "master equation" point of view, we are interested in the surface area buried in both.

(iii) Atomic Solvation Parameters (ASPs). The ASP method is a refinement of the surface area based method described above. Here, $\Delta G_{\text{hyd}} = \sum_i \sigma_i A_i$, where σ_i are the atomic solvation parameters for different types of atoms (five for the Eisenberg and McLachlan model;³⁹ seven for the Ooi *et al.*⁴⁰ model) and A_i is the averaged solvent exposed surface area for each type of atom. The same formulation is also used by Stouten *et al.*⁴¹ and Kang *et al.*⁴² where instead of accessible surface areas atomic occupancies (volume) are used. It is not clear which of these methods is better. The volume methods seem to have an edge in terms of the speed of calculations as it has been shown that it is possible to store the appropriate values on a grid.⁴³ The σ_i 's are calculated from experimental transfer energies, either from octanol to water or vapor to water. This method of calculating ΔG_{hyd} includes both enthalpic and entropic contributions as the least squares fit adopted to obtain averaged values of σ_i are obtained from experimental free energy data. Complications in incorporating this data into binding prediction schemes can arise, as some of the σ_i calculations assume that there is no conformational change upon transfer.

The original paper by Eisenberg and McLachlan³⁹ was parameterized such that it could not be used in conjunc-

tion with a molecular mechanics potential function. However, Ooi *et al.*⁴⁰ generated parameters that could be used in conjunction with a molecular mechanics (MM) potential function (data from water/vapor transfer experiments were used). In addition, fewer assumptions were made by Ooi *et al.* especially with respect to conformational changes due to the transfer. Wesson and Eisenberg,⁴⁴ in later work, also generated parameters that could be used with an MM potential function.

One can make two major objections to ASP-based methods. The first is that, in general, polar (or charged) groups in the interior of the protein would have different hydration energies compared to groups on or close to the surface.⁴⁵ The second is that hydration free energy of a group depends both on the hydration shell and the *entire* interface between protein and water.⁴⁶ The physical meaning of the numerical values obtained from the regression for each atom type is ambiguous. Another problem with the ASP parameters is that they may not compliment the molecular mechanics force field parameters. For details on both these points, see Schiffer *et al.*⁴⁷ More fundamentally, the validity of partitioning of solvation free energy into atomic contributions cannot be rigorously justified from a statistical mechanics perspective.⁴⁸

It is important to point out an attempt by Horton and Lewis⁴⁹ to predict binding. They used two criteria: (1) surface area term from Eisenberg and McLachlan and (2) loss of entropy on complex formation to fit experimental free energy of association of 15 protein complexes. Recognizing that the surface area terms for polar atoms should have different contributions, depending on whether or not they take part in hydrogen bonding or ionic interactions, they assign two additional parameters to the fitting procedure to account for this difference. The entropy loss was also an adjustable parameter in the regression. Excellent agreement (within a couple of kcal/mol) was obtained. Note, however, that a double fitting procedure (one by Eisenberg and McLachlan and the other by the authors) may not be as reliable as a single one.

(iv) An Approximate Treatment of Electrostatics in Water. Still *et al.* adopt the so-called generalized Born model (the GB/SA model). Here the electrostatic contribution to hydration free energy is modeled by the standard Coulombic interaction in addition to a polarization term of the two-charge system

$$\Delta G_{el} = \Delta G_{coul} + \Delta G_{pol} \quad (10)$$

The details can be found in eqs 17 and 18 in the appendix. This is an approximate continuum model (see the next item) formulation and avoids repeated solution of the Poisson equation. Next, a fitting procedure is adopted to determine the adjustable parameter in eq 18 so as to approximate the total hydration energy of two atoms in any molecule. It appears that this fitting procedure yields more accurate results for polar molecules compared to ionic molecules. In order to reproduce hydration energies of simple hydrocarbons, Still *et al.* introduced another term proportional to the surface area, $\Delta G_{cav} = \sigma \sum a_i$ (with a constant of proportionality, σ , independent of the atom type). The ΔG_{cav} term is added to the ΔG_{el} term to obtain the total contribution of the hydration effect. They refer to ΔG_{cav} as the free energy of cavity formation, which, strictly

speaking, should be positive and increases with solute size and does not include the dispersion interactions between the solute and water. However, Still *et al.* use it to represent the total hydration free energy of transfer of nonpolar solutes from gas phase to water.

(v) Continuum Model for Electrostatics in Water. A very substantial effort has been mounted to understand electrostatics in water over the past decade. The Honig group⁵⁰ has been one of the pioneers in this area (the program DELPHI is a contribution from this group). This is an important field as proteins are polar and an adequate understanding of inter- and intramolecular interactions require models that correctly reflect electrostatics. Most of the recent work focuses on continuum descriptions of solvent or protein-solvent systems and uses the boundary element method (BEM) or finite difference method (FDM). In a continuum description, water molecules are not treated explicitly; instead the emphasis is on a realistic treatment of the protein- or ligand-solvent boundary.

Most BEM applications solve the Poisson equation, and FDM approaches solve the linearized Poisson-Boltzmann (PB) equation. Absolute values of electrostatic energies are very hard to determine with these methods due to the approximations required to make the numerical methods tractable. One of the major problems with both these methods for calculating solvation energies is the errors induced because of the finite size of the grid; the grid points at which polarization charges are calculated may be at incorrect distances from the respective charged atom. The approach of the Honig group to binding free energy is to calculate the molecular mechanics potential function and two terms for solvation free energy (one from electrostatics and one nonpolar). (To our knowledge, no explicit binding free energy calculations have been reported in the literature by this group.) Their strength is the fairly reliable calculation of the electrostatics part of the free energy in the presence of ions and allowing for two dielectric constants (for different atomic polarizability) using the PB approach. However, the PB approach may not predict the electrostatic contribution to the difference in solvation free energy to better than 2 kcal/mol.⁵¹ The nonpolar interactions involve a comparatively crude calculation using an empirical surface area based term; this has, however, been improved recently. In recent work,⁵² they have extended their approach to obtain reasonably accurate values of hydration free energies of small organic molecules and amino acid side chains. As observed earlier, the polar contribution (electrostatics) to hydration free energy is obtained using finite difference solution to the Poisson equation. These results are added to the nonpolar solvation contribution which is obtained based on surface areas and experimental transfer (vacuum to water) data on hydrocarbons. One important reason for the reported success comes from the use of specifically optimized values for both the atomic radii and charge parameters for each atom. In this sense, this is also a regression type approach. A nice feature of this approach is that it is capable of dealing approximately with both polarization of functional groups and the ions present in the solvent.

(vi) Bound Waters in Hydration Free Energy Calculations. As mentioned previously, the role of bound water molecules can be quite important in

elucidating binding free energies. This is very difficult to evaluate. Recently, from experimental data on anhydrous and hydrated inorganic salts, Dunitz⁵³ estimated the entropic cost of a strongly bound water molecule to be roughly 2 kcal/mol at 300 K. This number has been disputed by Bryan⁵⁴ who showed that much larger increases (up to 6 kcal/mol at 300 K) are also possible. He also suggests that both the entropy of the protein and water can decrease. In our opinion, this contribution will be the hardest to calculate theoretically, compared to all the other contributions to binding free energy. The reason is that distinctions between strongly bound and not-so-strongly bound water molecules are extremely difficult even in experiments.

Contributions from Conformational Changes

In this part we review the methods adopted to account for contributions from changes in conformation of the protein and ligand due to their interactions.

(i) **Influence of Changes in Protein Conformation (ΔG^r).** In many complexes, the conformation of the receptor does not change much from the uncomplexed structure. In others, *e.g.*, HIV protease, the situation is different and the protein goes through a large change in its conformation on inhibitor binding.

The process of conformational change of the protein on binding involves the burial of hydrophobic surfaces (desolvation) which enhances binding and a change in entropy due to conformational changes (both backbone and side chain) which discourages binding. The enthalpic contribution to this can be obtained by changes in energy from the molecular mechanics potential function. The basic form of a molecular mechanics potential function is

$$E = E_{\text{bond}} + E_{\text{angle}} + E_{\text{torsion}} + E_{\text{vdW}} + E_{\text{Coulombic}} \quad (11)$$

The desolvation effects are modeled in two different ways, either using the atomic solvation parameters, ASPs, or by treating this contribution as purely entropic (ΔS_{solv} , the hydrophobic entropy) and estimating it from the change in heat capacity.

Many empirical observations about protein structure can be used in evaluating conformational changes of the protein. For example, certain residues have preferences for different conformational states (α -helix, β -sheet, *etc.*). However, there is controversy over the reasons for this. Four factors have generally been ascribed as the dominant physical reason for the observed preferences: (1) conformational entropy,^{55,56} (2) steric factors,⁵⁷ (3) hydrophobic effects,⁵⁸ and (4) electrostatics.⁵⁹ Knowing which of these factors is most important in a given system would simplify modeling the conformational changes upon binding.

(ii) **Influence of Changes in Ligand Conformation (ΔG^l).** If it is assumed that a single conformation dominates the free ligand (note that we have already assumed that the bound conformation of the ligand is unique and structural fluctuations of the bound ligand contribute negligible amounts to the free energy of binding), then the change in the self-energy of the ligand can be calculated in a fashion parallel to that of the receptor, as described in the previous section.^{60,61}

However, in general, it is incorrect to assume that the free ligand has a single dominant conformation in

solution. Therefore, ensemble averages (which are relatively easier to calculate here than for the receptor) become important for calculating the internal energy change, the backbone entropy change, and the hydrophobic transfer free energy change.⁶²

Statistical mechanics provides a recipe for calculating ensemble averages through the partition function (see eq 19). A rigorous evaluation of the partition function is almost impossible in a system of reasonable size. However, the average in eq 19 can be approximated by sampling a large enough number of conformations of the ligand. If a molecular mechanics potential function is used with explicit waters, then the enthalpic contribution, ΔE^l , to ΔG^l can be evaluated. The solvation contribution can also be accounted for using these conformations, that is, the ΔG_{hyd}^l (hydration free energy of the ligand) part of ΔG_{hyd} (see eq 8). If ASPs are used to account for the solvation contribution, then we have to remember that they are based on the assumption that the conformation of the amino acid residues does not change in the transfer experiments used to derive them. Therefore, an additional entropic term for the torsional freezing, $T\Delta S$, is also present. These entropies are evaluated using the microscopic definition of entropy

$$S = R \sum_j p_j \ln p_j \quad (12)$$

where R is the gas constant and p_j is the probability for conformation j and hence $\sum_j p_j = 1$. The estimation of entropies can, again, be based on extensive simulation data or may be based on experimentally observed distributions with the probabilities being calculated for each fragment of the ligand.^{60,62,63} Such entropy calculations are approximate due to difficulties in determining p_j .

Calculating Interaction Energies

The ΔG^{r-1} term in eq 8 represents the interaction energy. This is purely an enthalpic contribution, so ΔG 's are being replaced by ΔE 's. Therefore, $\Delta E^{c,r-1}$, the interaction energy in the complex is $E^{c,r-1} - E^{f,r} - E^{f,l}$ where $E_{c,r-1}$ is the total energy of the complex, $E^{f,r}$ is the energy of the free protein, and $E^{f,l}$ the energy of the free ligand.

There are two major contributions to the interaction energy: electrostatics and van der Waals interactions. The simplest way to calculate electrostatic interactions is to use Coulomb's law with atom-centered point charges. This is the standard molecular mechanics approach. There are also charge-dipole and dipole-dipole interactions, which are weaker individually, but are larger in number. In general, complex formation can also be accompanied by charge redistribution. Such charge redistribution results in a net attractive force. It is called polarization if it takes place within the ligand or the receptor, and it is called charge transfer if it takes place between the ligand and the receptor. Almost all calculations based on molecular mechanics potential functions, however, are restricted only to charge-charge interactions. The interactions between nonpolar molecules are parameterized as van der Waals interactions. This is a balance between attractive dispersion forces and short-range repulsion.

What is the Dielectric Constant? An important issue in treating electrostatic interactions is the answer to the question, "what is the dielectric constant, ϵ , in the active site?" Many answers to this question have been proposed, ranging from a constant value ($\epsilon = 1$ for vacuum and $\epsilon = 78$ for bulk water) to a value proportional to the distance between the atoms (a "distance dependent" dielectric, $\epsilon = r$) and more complicated functional forms. The interaction energies can change quite drastically as a result of changes in the dielectric constant. Different researchers use different values for mostly arbitrary reasons. There is neither a theoretically sound nor an experimentally compelling reason to prefer one value of dielectric constant over another. On the experimental front, gas phase, solution phase, and solid phase data may be used to justify different dielectrics. Most of the theoretical work has been done in improving the calculation of the electrostatic contribution, including the desolvation effect. Interestingly, the question of the form of the dielectric function has turned out to be important in our docking studies. For example, it is almost impossible to regenerate the docked conformation of methotrexate bound to DHFR (dihydrofolate reductase) if we use $\epsilon = 4r$. However, on using $\epsilon = r$, it is relatively easy to find low energy conformations close to the crystal structure. (The AMBER molecular mechanics force field has been used for these studies. Note, however, that docking studies are not necessarily a good guide for developing algorithms to predict K_i . For example, it is possible to obtain the correct binding orientations without including the crystallographically observed water molecules.) Another complication arises from the differences in the details of the different molecular mechanics potential energy functions. For example, the potential function may describe hydrogen bonds as purely electrostatic, or there may be an additional geometric term.

What is the Contribution of a Functional Group?

Some of the published work replaces the electrostatic contribution to ΔG^{r-1} by $\sum_i \Delta G_i$, where ΔG_i is the so-called "intrinsic (or apparent) binding energy"¹⁰ of the i th functional group in the ligand. Andrews *et al.*⁶⁴ have calculated the average binding energies of 10 functional groups (CO_2^- , OH, CO, sp^2 C, halogens, etc.) from 200 drugs and enzyme inhibitors. They assumed that all of the translational and rotational degrees of freedom of the ligand were lost completely on binding. This led to the assignment of unusually large values of intrinsic binding affinity to the functional groups unlike in the work of Williams' groups (see below). Usually, ΔG_i contains all the interactions between polar groups with the assumption that these values are transferable to any ligand interacting with any receptor (perhaps with the restriction of similar environment and solvent). The major motivation for this approach is to be able to answer the following question: "What is the change in binding energy if a particular functional group has been added to the ligand?"

Experimentally, the intrinsic binding energy for a functional group is obtained by comparing the binding energies for pairs of compounds which differ only in the functional group with the added assumption (or knowledge) that the two compounds bind similarly. For details, see Williams *et al.*⁶¹ The van der Waals contribution is calculated separately following the mo-

lecular mechanics framework. Another feature of this approach is the implicit accounting of the enthalpic contribution when estimating the contribution from desolvation effects. Therefore, care should be maintained to calculate only the entropic contributions to desolvation effects. The Williams' group has been very active in the field of binding energy prediction methods using the notion of intrinsic binding energy of functional groups (see Searle *et al.*⁶⁵ and references therein). They have also contributed to the elucidation of the entropic costs due to the translational, rotational, and conformational constraints. One of the peculiarities of attributing free energies to specific functional groups is that the specific values depend strongly on whether vibrational contributions on complex formation (see below) are assigned to functional groups or are treated along with the loss of translational, rotational entropies. It is not clear why an explicit accounting of ΔG_{vib} has not been performed to address this redundancy. One nice feature of their work (that is not found often in other work) is that they attempt to assess the influence of error in one contribution on the calculation of other contributions. Another important conclusion of the work by Williams' group is that the loss of entropy on binding is significantly smaller than that estimated by Page (*i.e.*, significant vibrational entropy remains in the complex; see the discussion of ΔG_{vib} below).

Some workers^{60,62} assume that protein–ligand, protein–solvent, and ligand–solvent interfaces are well packed and hence neglect any changes in the van der Waals interaction energy. Others assume that van der Waals interactions are better in a complex and therefore explicitly include them. Incorporation of vdW interactions depend on exactly how the other quantities are accounted for, *e.g.*, Williams *et al.*⁶¹ only take into consideration the nonpolar–nonpolar interactions as the well-packed polar–polar interactions are assumed to be accounted for in the ΔG_i calculations.

Crystallographic Waters. Explicit accounting as part of the protein of crystallographically well-defined water molecules can turn out to be very important for certain systems, *e.g.*, the "flap-water" in HIV protease. The presence of the water molecules should also be accounted for in the calculation of desolvation penalties, but the difference in this case may not be too significant. The role of the individual water molecules could turn out to be important in drug design when functional groups are modified which could lead to either the creation or removal of cavities large enough to accommodate water.

Accounting for Molecular Motion

This section considers the methods adopted to incorporate the effects of molecular motion. The approximations involved are underscored by the fact that in a number of methods the contributions are constant for a given ligand. Accounting for torsional constraints (due to binding) is covered first, and the following subsection covers translational, rotational, and vibrational contributions.

Torsional Constraints on Binding, ΔG_{rot} . The formation of a complex is associated with the freezing of the internal rotations of both the protein and the ligand. Hence there is a cost to free energy. This is mainly due to the adverse entropic change, although

there is some contribution due to enthalpy (changes in kinetic energy). This effect is usually incorporated using a constant^{60–62,66} value per dihedral angle that is frozen. Different constant values have been used. One approach to evaluate the loss of conformational entropy of immobilized side chains is to consider that there are three equal states (*i.e.*, three equienergetic states: trans and \pm gauche) per rotatable bond, therefore

$$R\Delta S_{\text{rot}} = -RT \ln 3 \quad (13)$$

from eq 12 or 14. This leads to a value of approximately 0.7 kcal/mol. Usually, the rotations of terminal CH₃ and NH₂ groups are not considered.

A third method, developed for treating peptides, but generally applicable, is to approximate the entropy of the free state using

$$S_i = -R \sum_j p_{ij} \ln(p_{ij}) \quad (14)$$

where p_{ij} is probability of side chain i to be in state j . Note that this is similar to eq 12. These have been estimated using observed frequencies of side chain conformations by (1) Pickett and Sternberg,⁶⁷ (2) Sternberg and Chickos⁶⁸ whose also attempt to estimate the errors involved in their calculations. Creamer and Rose⁵⁵ have also built a consistently calibrated conformational entropy scale using Monte-Carlo simulations. This conformational entropy scale is consistent with that of the hydrophobicity scale used in ASP formulations, in that the desolvation effects of binding are incorporated separately from the conformational effects. These calculations, as is true for all accounts of solvation contribution, has its own controversies; see, for example, the conclusions in Sternberg and Chickos.⁶⁸

Krystek *et al.*⁶⁰ argue that using eq 13 would be an *underestimate* of the real loss in entropy based on the argument that restricting one torsional angle would also restrict some others. However, from the analysis of Pickett and Sternberg,⁶⁷ one finds a range of values for the *effective* number of equienergetic states for amino acid side chains. Interestingly, it turns out that Krystek *et al.*⁶⁰ value is in fact an *overestimate*. This points out the difficulties with qualitative arguments. One important and useful lesson that points out the difficulties in K_i prediction schemes can be found in the Pickett and Sternberg⁶⁷ data. Errors generated in estimating ΔG_{rot} by using the Krystek *et al.*⁶⁰ approach can be as large as 1.5 kcal/mol. But, overall, the Krystek *et al.* results are quite good. One possible reason for this is the enthalpy–entropy compensation phenomenon.

Translation, Rotation, and Vibrational Contribution. The fundamental reason for treating all three contributions together is that often the translational and rotational degrees of freedom of the ligand and enzyme in the unbound state get mixed into the many vibrational states of the complex. However, we continue with our additive assumption. An extreme example of ligand mobility in complex is exemplified by some very interesting experimental results on the binding of thiocamphor and other ligands to cytochrome P-450_{CAM}. Raag and Poulos⁶⁸ show that the lack of hydrogen bonds and complementary vdW interactions leads to higher mobility of the ligand in the complex, the evidence being provided by the higher temperature factors in the

crystal structures as compared to camphor complex and the larger number of hydroxylation sites.

When two molecules bind, there is a loss of 3 rotational and 3 translational degrees of freedom. There is an enthalpic and entropic contribution to this free energy loss. The enthalpic contribution can be reasonably expected to be about $3RT$ ($RT/2$ for each degree of freedom). This is usually assumed to be a constant value independent of the size of the ligand. The entropic contribution, however, is not simple to evaluate accurately. It appears that an overall contribution between 7 and 11 kcal/mol is agreed upon irrespective of the ligand.^{10,61,65} The exact value of this contribution is, however, only important for absolute values of the free energy. This contribution would cancel out when comparing different ligands binding to the same receptor.

It is relatively simple to calculate rotational and translational free energies for small molecules, and immensely difficult for larger ones. See, for example, Hill⁶⁹ who has given expressions for translational, rotational, and vibrational entropies in the gas phase. It is believed to be a reasonable approximation to use gas phase equations to study solution behavior.^{10,70,71}

(1) Upon complex formation, parts of the translational and rotational freedom appears as new normal modes, namely six new internal vibrational modes. In addition to these six new modes, there can also be a change in the vibrational density of states (the frequency of the vibrational normal modes are altered). It has also been suggested that the complex has lower frequency modes compared to the uncomplexed protein.¹⁰ A general principle that could be kept in mind is that looser complexes have larger vibrational entropy compared to tighter ones.

(2) Finkelstein and Janin⁷⁰ evaluate that about half of the loss in translational and rotational entropy is compensated for by new vibrational modes upon binding of BPTI. This results in a net loss of approximately 15 kcal/mol. However, this number appears to be quite large when protein-cleaving experiments are taken into account.⁷² It has been shown that proteins can be cleaved at one or two sites without loss of stability. So if the Finkelstein and Janin estimate of the loss of translational and rotational entropy is correct, what compensates for the loss of 15–30 kcal/mol in entropy?

(3) Conformational entropy: Karplus and Kushick⁷³ found that both dihedral angle and bond angle variations contribute significantly to the entropy of proteins. However, see the detailed criticism of their methods by Edholm and Berendsen.⁷⁴

(4) The vibrational entropy should be treated quantum mechanically, whereas the translational and rotational contributions can be treated semiclassically. The vibrational contribution can be obtained by calculating the normal modes of the protein before and after complexation (assuming the anharmonic components cancel). However, this approach completely neglects protein motion. Vibrational contribution to binding is, to a very good approximation, independent of the size of the ligand. This is now the case for the solvation effects considered above. Therefore, for small ligands which have small solvation contribution, accurate evaluation of the vibrational entropy would be important in accessing binding free energies.

In an early attempt, Page and Jencks¹⁰ derived unfavorable $T\Delta S$ contribution at room temperature in solution to be 12–17 kcal/mol (if one of the components is about 100–1000 mass units). This value comes from reducing the 12 degrees of freedom (rotations and translations before association) to 6. This number has held up even in recent work with minor modifications.⁶¹ The loss of the 3 rotational degrees of freedom also results in an enthalpy loss of about 1 kcal/mol. The translational degrees of freedom do not contribute to enthalpy as the average translational kinetic energy of a molecule is $3RT/2$ irrespective of whether it is in a gas, liquid, or solid phase. The reason is that on going from the gas to the liquid phase the molecule's motion gets restricted to a narrower region of space.

Regression Approaches: Old and New

In the previous section we discussed approaches based on the "master equation". Here we discuss a very different method: methods that average out the contributions from many complexes. The simplest of these procedures have some value from an applications point of view due to their computational efficiency.

2D and 3D QSAR. One of the oldest methods that attempts to rationalize the binding of different molecules to a receptor is quantitative structure–activity relationships (QSAR). The basic idea is to build a regression model for the biological activity of a number of related compounds based on physical and chemical properties of the ligands. Many different properties have been used to build regression models.⁷⁵ In order to incorporate the three-dimensional nature of the binding process the notion of building regression models based on the spatial properties of the ligands (*e.g.*, electrostatics) has been introduced. There is now a large body of research on 3D QSAR methods.⁷⁶

3D QSAR can be quite powerful and useful. However, there are two major reasons why a regression model in 3D QSAR can be misleading:

(1) The binding mode of a given ligand is not known, *a priori*. This effectively has to be guessed either by aligning similar ligands based on their properties or aligning them based on the receptor interactions that are present (if it is known). Obviously, the regression will be more predictive if the alignment is done on the basis of the properties of the receptor binding site.

(2) The problem is underdetermined, *i.e.*, there are a much larger number of possible regressions compared to the ligands. This implies that standard regression methodology will yield incorrect results and some kind of biased regression method has to be used, the most common one being partial least squares (PLS). (The term "biased" is used here in a strict statistical sense and not in the literal sense of the word.) Usually, many equivalent models can be built using the regression approach. A method that has been recently developed that combines these different models⁷⁷ to build a higher level model could be useful to enhance predictive power.

Predictions Based on Molecular Mechanics Energy. Recently, there have been attempts to generate QSAR-type regression equations based on a molecular mechanics energy function^{78–80} yielding reasonable results. These methods ignore (1) the influence of the solvent almost completely and (2) the influence of

difference between bound and unbound conformations of the ligands to different degrees. These procedures assume that the bound conformation of the ligand and protein are known or can at least be well approximated. There are fundamental limits to this approach which arise from the neglect of entropic penalties and solvation contributions. Therefore, the success of these methods depend on the ability to carefully define the set of ligands on which predictions will be applied, which in turn is difficult to accomplish *a priori*. An interesting conclusion (stated without any details) in the Holloway *et al.*⁷⁹ work is that "attempts to incorporate these [flexibility of ligand and receptor, solvation effects] effects into a prediction model have been thus far unsuccessful in improving observed correlation". In any event, it appears that the purely molecular mechanics based approaches may be useful in a congeneric series.

Empirical Functions Approach. Recently, Böhm,⁶⁶ in an interesting and ambitious attempt adopted a simple linear regression approach to describe binding. A set of independent parameters (similar to the "master-equation" parameters) were used in conjunction with known (or modeled) crystal structures of about 45 protein–ligand complexes to obtain a linear equation. The free energy of binding was written as the sum of terms including a constant representing overall rotational and translational loss, a sum over all hydrogen bonds formed, a sum over all ionic interactions, the loss of lipophilic surface area on binding, and the number of torsions that are frozen. Linear regression yields the strength of the contribution of each term in the above equation. A plausible set of parameters was obtained, *e.g.*, a translational/rotational contribution of about 1.2 kcal/mol, a hydrogen bond interaction strength of -1.12 kcal/mol. The equation obtained also seems to have reasonable predictive ability in the examples considered.

This is of course a simplistic method and will have very obvious limitations in light of the above discussion. For example, in our experience and Böhm's (personal communication), the strength of contributions of different terms in the equation is highly dependent on the complexes included in building the regression. The method seems to perform well on predicting the change in free energy due to changes in surface area burial and, not surprisingly, not perform well due to changes in hydrogen bonds. The contribution to binding (or stability) due to the formation of a hydrogen bond seems to depend highly on the environment in which the hydrogen bond is formed and values ranging from 2 to 10 kcal/mol have been obtained (see Dill⁸¹ for details on the difficulty of estimating this number). In addition the contribution towards a hydrogen bond in the Böhm scheme depends on the details of the radial and angular interactions and will therefore be very dependent on how the hydrogens are added to donor atoms. Böhm also treats bonds to metal atoms as hydrogen bonds. Interestingly, recent work by Weber *et al.*⁸² attempts to explain the binding in streptavidin (an outlier in Böhm's analysis) by the number of crystallographic waters displaced.

Recently, Head *et al.*⁸³ in an analogous procedure have examined 61 protein–ligand complexes using partial least squares regression and neural network regression (they do not report any results using standard multiple linear regression procedure). They stud-

ied predictions on three independent sets of protein–ligand complexes and obtained mixed results. Their procedure uses 12 independent parameters including (1) the molecular mechanics energy, (2) solvent effects through $\log P$ calculations, (3) two surface area terms further subdivided into lipophilic, hydrophilic, polar, and nonpolar parts, (4) the number of rotatable bonds, and (5) the change in intramolecular ligand energy on going from a single unbound conformation to a single bound conformation. It would appear that different statistical procedures for eliminating unimportant variables would lead to different equations with similar predictive behavior. However, this analysis was not performed. Overall, both the neural network and the PLS method yield similar results. An interesting conclusion of this work is that the correlation between the electrostatic energy and binding was very weak in the regression equation. This is in contrast to work described earlier which has found good correlation between the electrostatic energy and binding affinity.

Future Aspects

All theories and rules are necessary for rationalizing and interpreting data, both for theorists and experimentalists. These theories and rules (however approximate) are needed also for designing new and potent inhibitors as all experiments cannot be done or in cases where direct experimental measurements are difficult.

Most contributions to protein–ligand interactions are electrostatic in nature.³⁰ Major nonelectrostatic contributions arise from chemical bonds, short-range repulsion from the Pauli principle, and partially entropic forces due to differences in number of substates. The least well understood, theoretically or experimentally, is solvation effects. A nice illustration of some of the contradictory conclusions that can be reached using different experimental setups is given in table 1 of Connelly⁸⁴ where different model systems used to mimic protein folding or protein–ligand binding predict opposite signs for contributions from same interaction (ΔC , ΔH , ΔS , and ΔG).

We mentioned in the Introduction that ligand binding should be a simpler problem to understand than protein folding. There is usually a large exchange of ideas between the two communities. It is, therefore, useful to study some of the newer approaches in protein folding. One of these is the method of Sippl²⁵ that starts by constructing a potential function based on crystal structures. The basic aim is to guarantee that the minimum of that potential function is the crystal structure. It is known that this is not true for molecular mechanics potential energy functions. As this approach involves extensive averaging to determine the probability distribution of different configurations (and depends critically on such averaging for its success), it will become more attractive as the number of X-ray- or NMR-determined structures of complexes increases.

One of the first attempts to understand the thermodynamics of protein–ligand and protein–protein interaction was by Janin and Chotia.¹² They used two criteria: (1) surface area buried on complexation and (2) loss of rotational–translational entropy. They obtained reasonable results. However, Finkelstein and Janin⁷⁰ noted that the earlier calculations neglected the effect of vibrational entropy (as large as 15 kcal/mol)

and hence would significantly change the original results. This example, to our mind, exemplifies the difficulties that have plagued attempts at predicting free energy of binding.

In all of the analysis presented above, it has been implicitly assumed that free energy can be broken up into components *meaningfully*. Critiques of this position are also available in the literature; see Mark and van Gunsteran⁸⁶ whose position is that free energy is a state function but the components of the free energy are not state functions and Janin⁸⁷ who provides an overview of some of the questions that are (1) usually assumed to be significant but are not and (2) important to keep in mind while predicting free energies of binding. We agree with both Mark and van Gunsteran and Janin. However, we believe that there is some utility in the attempts to understand free energies as described in this perspective based on three basic arguments. First, as stated above, even approximate rules may be useful for gaining a crude understanding. Second, it is hard to ignore the successes of some of the empirical schemes that we have described. Third (a point unrelated to theoretical arguments), it provides a reasonable rationale for choosing which compound to make next within a drug design program.

We would like to point out that “back-of-the-envelope” calculations of binding affinity are not necessarily very predictive when hydrogen bonds and salt bridges are involved in the binding process. For example, it is often believed that binding affinity increases about an order of magnitude per hydrogen bond. The glycogen phosphorylase–glucose complex⁹ shows that, despite many hydrogen bonds to the protein, the ligand exhibits only a very weak binding. The primary reason for this is that desolvation costs can entirely compensate for hydrogen bonds. These desolvation penalties, as we have seen, depend on the microscopic environment of the hydrogen bonding groups.

Another point that should be kept in mind is that the regression based approaches described do not account for the fact that unpaired buried polar groups in the protein–ligand interface are strongly adverse to binding. There is some correlation (empirically observed) between the number of hydrogen bonds formed and the K_i ; similarly there is some correlation between K_i and the lipophilic surface area buried. Neither correlations is perfect as seen in the results of Horton and Lewis⁴⁹ and Böhm.⁶⁶ It is possible, for example, that certain lipophilic groups on ligands stabilize the bound conformation of the ligand in solution even though the lipophilic area buried does not change.

An approach that is often used to improve binding is specific rigidification of the ligand. Ku *et al.*⁸⁸ provide an example of the successful design of a potent non-peptide fibrinogen receptor antagonist by analogy to a low-activity constrained peptide. On the other hand, the work by Weber *et al.*⁸⁵ provides a counterexample. They found that, for synthetic azobenzene ligands of streptavidin, ligand flexibility in the bound state contributes to the overall binding.

We have described the controversies involved in both the theoretical calculations and experimental interpretations of data. This is especially true with respect to the role of water. The question that becomes immediately apparent is: what are the reasons for the

published successes? These are not immediately apparent. It is clear that the methods work under certain conditions, e.g., when there is a large amount of fortuitous cancellations among the contributions that are neglected. This could be related to the enthalpy/entropy compensation in aqueous solutions which lets us predict the total free energy even though the components can not be predicted reliably.⁹⁰ In other words, increasing the enthalpic contribution to binding would require very specific positioning of functional groups which in turn would increase the entropic cost for binding. Under these conditions incorrect predictions of components of the free energy of binding may not lead to large inaccuracies in the prediction of binding free energies.

The basic conclusions from our analysis are as follows:

(1) Rule-of-thumb methods do not always work well for hydrogen bonding. They are better for lipophilic interactions.

(2) Straightforward forcefield methods are too simplistic except for congeneric series.

(3) The free energy perturbation method holds great promise if the convergence problems are solved.

(4) Solvation effects are still not well understood. (i) It constitutes the largest source of error in predictions. (ii) It has bulk effects on the environment, conformational preferences, etc. (iii) Specific water-mediated hydrogen bonds are hard to quantify. (iv) Addition of explicit waters really slows down calculations.

(5) Ligand flexibility is essential. It allows for estimating (i) entropy of binding (how many conformations are likely) and (ii) strain energy on binding.

(6) Protein flexibility is important, though it is often neglected. It is computationally intensive, but allows for estimating the consequences of (i) major conformational changes and (ii) many side chain motions.

(7) The vibrational states of the complex and unbound protein and ligand are different in general. Accounting for these differences is not easy.

In our estimation there are three steps that will be important in the immediate future to enhance our understanding of the different prediction algorithms:

(1) Reporting of prediction errors (even if they can only be assessed crudely) along with the predictions themselves.

(2) Setting up a database of ligand-receptor complexes chosen to emphasize differences in (i) magnitude of K_i and (ii) major contributions to binding can be used to analyze and compare the performance of all the available prediction algorithms. This will involve setting up a database that has both flexible and rigid ligands, ligands that differ in one or more hydrogen or ionic bonding possibilities, and ligands that have different amounts of polar and nonpolar surface area buried. A series of complexes that span a range of functionally and structurally different proteins will also be needed.

(3) Finally, only a limited number of microcalorimetric measurements that measure both enthalpy and entropy changes exist. More data along these lines will surely enhance theoretical studies as, given a protein, different ligand can bind through either enthalpically or entropically driven processes; an example is biotin binding to streptavidin is enthalpy driven whereas azobenzene ligands binding to streptavidin is entropy driven.^{82,88}

Appendix

Surface Area Based Equations in Estimating Solvation Contributions.

$$\Delta S_{\text{solv}} = 0.32\Delta A_{\text{np}} \ln(T/386) \quad (15)$$

where ΔA_{np} is the nonpolar surface area buried in \AA^2 , T is in Kelvin, and entropy is in $\text{cal K}^{-1} \text{mol}^{-1}$.

$$\Delta G_{\text{hyd}} = 25A_{\text{contact}} \text{ (in calories)} \quad (16)$$

Generalized Born Model Equations in Estimating Solvation Contributions.

$$\Delta G_{\text{coul}} = 166 \sum_{i \neq j} \frac{q_i q_j}{r_{ij}} \quad (17)$$

where q_i is the charge on atom i and r_{ij} is the distance between atoms i and j and the units are kcal/mol.

$$\Delta G_{\text{pol}} = -166(1 - 1/D_0) \sum_i \sum_j \frac{q_i q_j}{f_{ij}} \quad (18)$$

where $f_{ij} = r_{ij}$ for $i \neq j$ and $f_{ij} = a_i$ for $i = j$, q_i is the charge on atom i , and r_{ij} is the distance between atoms i and j , a_i is the radius of atom i , and the units are kcal/mol.

Ensemble Average Calculation in "Contributions from Conformational Changes".

$$\langle \Delta E^i \rangle = \frac{\sum_i \Delta E_i^i \exp[-\Delta E_i^i/RT]}{\mathcal{Z}} \quad (19)$$

where \mathcal{Z} is the partition function⁶⁵ and the subscript i is for an individual conformation.

Acknowledgment. We would like to thank our colleagues, Dr. Patrick Connelly and Dr. B. Govinda Rao, for helpful comments and discussion. We would also like to thank the anonymous reviewers for helpful suggestions.

Biographies

Ajay received his B.S. degree in physics from Delhi University in New Delhi, India, and his Ph.D. in theoretical condensed matter physics with Professor Richard Palmer at University of North Carolina, Chapel Hill, in 1991. After postdoctoral work with Professor Robert Langridge at University of California, San Francisco, developing neural network and genetic algorithm methods for molecular modeling, he joined Vertex Pharmaceuticals in 1994, where he is working on developing novel artificial intelligence tools to aid in computational drug design.

Mark A. Murcko received his B.S. degree in chemistry from Fairfield University, Connecticut, and his Ph.D. in organic chemistry with Professor Ken Wiberg at Yale in 1987. He then joined the molecular modeling group at Merck in West Point, PA. In 1990, he joined Vertex Pharmaceuticals, where he is involved in the application of structural information and a wide variety of modeling and computational tools to the process of drug design.

References

- (1) Creighton, T. E. *Proteins: Structures and Molecular Properties*, 2nd ed.; W. H. Freeman: New York, 1993.
- (2) Cramer, C. J.; Truhlar, D. G. AM1-SM2 and PM3-SM3 parameterized scf solvation models for free energies in aqueous solution. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 629-666.

- (3) Berg, J. M. Zinc finger domains: hypotheses and current knowledge. *Annu. Rev. Biophys. Biophys. Chem.* **1990**, *19*, 405–421.
- (4) Strynadka, N. C. J.; James, M. N. G. Crystal structures of the helix-loop-helix calcium-binding proteins. *Annu. Rev. Biol.* **1989**, *58*, 951–998.
- (5) Kim, E. E.; Baker, C. T.; Dwyer, M. D.; Murcko, M. A.; Rao, B. G.; Tung, R. D.; Navia, M. A. Crystal structure of hiv-1 protease in complex with vx-478, a potent and orally bioavailable inhibitor of the enzyme. *J. Am. Chem. Soc.* **1995**, *117*, 1181–1182.
- (6) Reich, S. H.; Fuhry, M. A.; Nguyen, D.; Pino, M. J.; *et al.* Design and synthesis of novel 6,7-imidazotetrahydroquinoline inhibitors of thymidylate synthase using iterative protein crystal. *J. Med. Chem.* **1992**, *35*, 847–858.
- (7) Verlindé, C. L.; Merritt, E. A.; den Akker, F. V.; Kim, H.; Feil, I.; Delboni, L. F.; Mande, S. C.; Sarfaty, S.; Petra, P. H.; Hol, W. G. Protein crystallography and infectious diseases. *Protein Sci.* **1994**, *3*, 1670–86.
- (8) Veale, C. A.; Bernstein, P. R.; Bryant, C.; Ceccarelli, C.; Damewood, J. R. Jr.; Earley, R.; Feeney, S. W.; Gomes, B. G.; *et al.* Nonpeptidic inhibitors of human leukocyte elastase. 5. Design, synthesis, and x-ray crystallography of a series of orally active 5-aminopyrimidin-6-one-containing trifluoromethyl ketones. *J. Med. Chem.* **1995**, *38*, 98–108.
- (9) Watson, K. A.; Mitchell, E. P.; Johnson, L. N.; Son, J. C.; Richard, C. J.; Orchard, M. G.; *et al.* Design of inhibitors of glycogen phosphorylase: a study of alpha- and beta-c-glucosides and 1-thio-beta-d-glucose compounds. *Biochemistry* **1994**, *33*, 5748–58.
- (10) Page, M. I.; Jencks, W. P. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U.S.A.* **1971**, *68*, 1678–1683.
- (11) Jencks, W. P. Binding energy, specificity, and enzymic catalysis: the circe effect. *Adv. Enzym. Related Areas Mol. Biol.* **1975**, *43*, 219–410.
- (12) Janin, J.; Chotia, C. Role of hydrophobicity in the binding of coenzymes. Appendix. translational and rotational contribution to the free energy of dissociation. *Biochemistry* **1978**, *17*, 2943–2948.
- (13) de Heer, J. *Phenomenological Thermodynamics with Applications to Chemistry*; Prentice Hall: Engelwood Cliffs, NJ, 1986.
- (14) Atkins, P. W. *Physical Chemistry*, 4th ed.; W. H. Freeman: New York, 1990.
- (15) Martin, Y. C. *Quantitative drug design: a critical introduction*; Marcel Dekker: New York, 1978.
- (16) Fersht, A. *Enzyme structure and mechanism*, 2nd ed.; W. H. Freeman: New York, 1985.
- (17) Levitt, M.; Park, B. H. Water: now you see it, now you don't. *Structure* **1993**, *1*, 223–226.
- (18) Shakked, Z.; Guzikovich-Guerstein, G.; Frolow, F.; Rabinovich, D.; Joachimiak, A.; Sigler, P. B. Determinants of repressor/operator recognition from the structure of the trp operator binding site. *Nature* **1994**, *368*, 469–473.
- (19) Brooks, B. R.; Brucoleri, R. E.; Olafson, B. D.; Swaminathan, D. J. S. S.; Karplus, M. Charmm: A Program for macromolecular energy minimization, and dynamics calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.
- (20) Weiner, S. J.; Kollman, P. A.; Nguyen, D. T.; Case, D. A. An all atom force field for simulations of proteins and nucleic acids. *J. Comput. Chem.* **1986**, *7*, 230–252.
- (21) Nemethy, G.; Pottle, M. S.; Scheraga, H. A. Energy parameters in polypeptides. 9. updating the geometrical parameters, nonbonded interactions, and hydrogen bond interactions for the naturally occurring amino-acids. *J. Phys. Chem.* **1983**, *87*, 1883–1887.
- (22) Kollman, P. Free energy calculations-applications to chemical and biochemical phenomena. *Chem. Rev.* **1993**, *93*, 2395–2417.
- (23) Rao, B. G.; Tilton, R. F.; Singh, U. C. Free energy perturbation studies on inhibitor binding to hiv-1 proteinase. *J. Am. Chem. Soc.* **1992**, *114*, 4447–4452.
- (24) Singh, S. B.; Ajay; Wemmer, D. E.; Kollman, P. A. Relative binding affinities of distamycin and its analog to d(CGCAAG-TTGGC)-d(GCCAATTGCG): Comparison of simulation results with experiment. *Proc. Natl. Acad. Sci. U.S.A.* **1994**, *91*, 7673–7677.
- (25) Schreiber, H.; Steinhauser, O. Cutoff size does strongly influence molecular dynamics results on solvated polypeptides. *Biochemistry* **1992**, *31*, 5856–60.
- (26) York, D. M.; Darden, T. A.; Pedersen, L. G. The effect of long range electrostatic interactions in simulations of macromolecular crystals: a comparison of Ewald and truncated list methods. *J. Phys. Chem.* **1993**, *99*, 8345–8348.
- (27) Caldwell, J. W.; Kollman, P. A. Cation-pi interactions-nonadditive effects are critical in their accurate representation. *J. Am. Chem. Soc.* **1995**, *117*, 4177–4178.
- (28) Smith, P. E.; van Gunsteran, W. F. Translational and rotational diffusion of proteins. *J. Mol. Biol.* **1994**, *236*, 629–636.
- (29) Shi, Y. Y.; Mark, A. E.; Wang, C. X.; Huang, F.; Berendsen, H. J.; van Gunsteran, W. F. Can the stability of protein mutants be predicted by free energy calculations? *Protein Eng.* **1993**, *6*, 289–95.
- (30) Israelachvili, J. N. *Intermolecular and surface forces*; Academic Press: London, 1985.
- (31) Fauchere, J. L.; Pliska, V. *Eur. J. Med. Chem.* **1983**, *18*, 369–375.
- (32) Cornette, J. L.; Cease, K. B.; Margalit, H.; Spouge, J. L.; Berzofsky, J. A.; DeLisi, C. Hydrophobicity scales and computational techniques for detecting amphipathic structures in proteins. *J. Mol. Biol.* **1987**, *195*, 659–685.
- (33) Lee, B. Estimation of the maximum change in stability of globular proteins upon mutation of a hydrophobic residue to another of smaller size. *Protein Sci.* **1993**, *2*, 733–738.
- (34) Sharp, K. A.; Nicholls, A.; Friedman, R.; Honig, B. Extracting hydrophobic free energies from experimental data: relationship to protein folding and theoretical models. *Biochemistry* **1991**, *30*, 9686–9697.
- (35) Chan, H. S.; Dill, K. A. Solvation—effects of molecular size and shape. *J. Chem. Phys.* **1994**, *101*, 7007–7026.
- (36) Smith, P. E.; Pettitt, B. M. Modeling solvent in biomolecular systems. *J. Phys. Chem.* **1994**, *98*, 9700–9711.
- (37) Spolar, R. S., Jr.; M. T. R. Coupling of local folding to site-specific binding of proteins to dna. *Science* **1994**, *263*, 777–784.
- (38) Novotny, J.; Brucoleri, R. E.; Saul, F. A. On the attribution of binding energy in antigen-antibody complexes MCPC 603, D1.3 and Hyhel-5. *Biochemistry* **1989**, *28*, 4735–4749.
- (39) Eisenberg, D.; McLachlan, A. D. Solvation energy in protein folding and binding. *Nature* **1986**, *319*, 199–203.
- (40) Ooi, T.; Oobatake, M.; Nemethy, G.; Scheraga, H. A. Accessible surface area as a measure of the thermodynamic parameters of hydration of peptides. *Proc. Natl. Acad. Sci. U.S.A.* **1987**, *84*, 3086–3090.
- (41) Stouten, P. F. W.; Frommel, C.; Nakamura, H.; Sander, C. An effective solvation term based on atomic occupancies for use in protein simulations. *Mol. Simul.* **1993**, *10*, 97–120.
- (42) Kang, Y. K.; Gibson, K. D.; Nemethy, G.; Scheraga, H. A. Free energies of hydration of solute molecules. 4. revised treatment of the hydration shell model. *J. Phys. Chem.* **1988**, *92*, 4739–4742.
- (43) Luty, B. A.; Wasserman, Z. R.; Stouten, P. F. W.; Hodge, C. N.; Zacharias, M.; McCammon, J. A. A molecular mechanics/grid method for evaluation of ligand-receptor interactions. *J. Comput. Chem.* **1995**, *16*, 454–464.
- (44) Wesson, L.; Eisenberg, D. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci.* **1992**, *1*, 227–235.
- (45) Rashin, A. A. Continuum electrostatics and hydration phenomena. *Int. J. Quant. Chem.: Quant. Biol. Symp.* **1988**, *15*, 103–118.
- (46) Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B. Focusing of electric fields in the active site of cu-zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins* **1986**, *1*, 47–59.
- (47) Schiffer, C. A.; Caldwell, J. W.; Kollman, P. A.; Stroud, R. M. Protein structure prediction with a combined solvation free energy-molecular mechanics force field. *Mol. Simulation* **1993**, *10*, 121–149.
- (48) Ben-Naim, A. *Solvation Thermodynamics*; Plenum Press: New York, 1987.
- (49) Horton, N.; Lewis, M. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1992**, *1*, 169–181.
- (50) Honig, B.; Sharp, K.; Yang, A.-S. Macroscopic models of aqueous solutions: Biological and chemical applications. *J. Phys. Chem.* **1993**, *97*, 1101–1109.
- (51) Rick, S. W.; Berne, B. J. The aqueous solvation of water: A comparison of continuum methods with molecular dynamics. *J. Am. Chem. Soc.* **1994**, *116*, 3949–3954.
- (52) Sitkoff, D.; Sharp, K. A.; Honig, B. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* **1994**, *98*, 1978–1988.
- (53) Dunitz, J. D. The entropic cost of bound water in crystals and biomolecules. *Science* **1994**, *264*, 670–670.
- (54) Bryan, W. P. The entropic cost of binding water to proteins. *Science* **1994**, *266*, 1726.
- (55) Creamer, T. P.; Rose, G. D. Side-chain entropy opposes alpha-helix formation but rationalizes experimentally determined helix-forming propensities. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 5937–5941.
- (56) Padmanabhan, S.; Baldwin, R. L. Straight-chain nonpolar amino acids are good helix-formers in water. *J. Mol. Biol.* **1991**, *219*, 135–137.
- (57) Hermans, J.; Anderson, A. G.; Yun, R. H. Differential helix propensity of small apolar side chain studied by molecular dynamics simulation. *Biochemistry* **1992**, *31*, 5646–5653.
- (58) Horowitz, A.; Matthews, J. M.; Fersht, A. Alpha-helix stability in proteins ii. factors that influence stability at an internal position. *J. Mol. Biol.* **1992**, *227*, 560–568.

- (59) Avbelj, F.; Moulton, J. Role of electrostatic screening in determining protein main chain conformational preferences. *Biochemistry* **1995**, *34*, 755–764.
- (60) Krystek, S.; Stouch, T.; Novotny, J. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J. Mol. Biol.* **1993**, *234*, 661–679.
- (61) Williams, D. H.; Cox, J. P. L.; Doig, A. J.; Garner, M.; Gerhard, U.; Kaye, P. T.; Lal, A. R.; Nicholls, I. A.; Salter, C. J.; Mitchell, R. C. Toward the semiquantitative estimation of binding constants. guides for peptide-peptide binding in aqueous solution. *J. Am. Chem. Soc.* **1991**, *113*, 7020–7030.
- (62) Vajda, S.; Weng, Z.; Rosenfeld, R.; DeLisi, C. Effect of conformational flexibility and solvation on receptor-ligand binding free energies. *Biochemistry* **1994**, *33*, 13977–13988.
- (63) Sternberg, M. J. E.; Chickos, J. S. Protein side-chain conformational entropy derived from fusion data-comparison with other empirical scales. *Protein Eng.* **1994**, *7*, 149–155.
- (64) Andrews, P. R.; Craik, D. J.; Martin, J. L. Functional group contributions to drug-receptor interactions. *J. Med. Chem.* **1984**, *27*, 1648–57.
- (65) Searle, M. S.; Williams, D. H.; Gerhard, U. 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.* **1992**, *114*, 10697–10704.
- (66) H.-J. Böhm. 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 Mol. Des.* **1994**, *8*, 243–256.
- (67) Pickett, S. D.; Sternberg, M. E. Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* **1993**, *231*, 825–839.
- (68) Raag, R.; Poulos, T. L. Crystal structure of cytochrome P-450_{CAM} complexed with camphane, tiocamphor, and adamantane: factors controlling P-450 substrate hydroxylation. *Biochemistry* **1991**, *30*, 2674–2684.
- (69) Hill, T. L. *Statistical mechanics: principles and selected applications*; Dover: New York, 1987.
- (70) Finkelstein, A. V.; Janin, J. The price of lost freedom—entropy of bimolecular complex formation. *Protein Eng.* **1989**, *3*, 1–3.
- (71) Chotia, C.; Janin, J. Principles of protein-protein recognition. *Nature* **1975**, *256*, 705–708.
- (72) Sancho, J.; Fersht, A. R. Dissection of an enzyme by protein engineering—the n-terminal and c-terminal fragments of barnase form a native-like complex with restored enzymic activity. *J. Mol. Biol.* **1992**, *224*, 741–747.
- (73) Karplus, M.; Kushick, J. N. *Macromolecules* **1981**, *14*, 325–332.
- (74) Edholm, O.; Berendsen, H. J. C. *Mol. Phys.* **1984**, *51*, 1011–1028.
- (75) Hansch, C.; Sammes, P. G.; Taylor, J. B., Eds. *Comprehensive medicinal chemistry: the rational design, mechanistic study & therapeutic application of chemical compounds*; Pergamon Press: Oxford, 1990.
- (76) Kubinyi, H., Ed. *3D QSAR in Drug Design*; ESCOM: Leiden, 1993.
- (77) Ajay. On better generalization by combining 2 or more models—a quantitative structure activity relationship example using neural networks. *Chem. Intl. Lab. Sys.* **1994**, *24*, 19–30.
- (78) Grootenhuys, P. D. J.; van Helden, S. P. Rational approaches towards protease inhibition: Predicting the binding of thrombin inhibitors. In *Computational approaches in supramolecular chemistry*; Wipff, G., Ed.; Kluwer Academic: Boston, 1994; pp 137–149.
- (79) Holloway, M. K.; Wei, J. M.; *et al.* A *Priori* prediction of activity for hiv-1 protease inhibitors employing energy minimization in the active site. *J. Med. Chem.* **1995**, *38*, 305–317.
- (80) Ortiz, A. R.; Pisabarro, M. T.; Gago, F.; Wade, R. C. Prediction of drug binding affinities by comparative binding energy analysis. *Preprint* **1995**.
- (81) Dill, K. A. Dominant forces in protein folding. *Biochemistry* **1990**, *29*, 7133–7155.
- (82) Weber, P. C.; Wendoloski, J. J.; Pantoliano, M. W.; Salemme, F. R. Crystallographic and thermodynamic comparison of natural and synthetic ligands bound to streptavidin. *J. Am. Chem. Soc.* **1992**, *114*, 3197–3200.
- (83) Head, R. D.; Smyte, M. L.; Oprea, T. I.; Waller, C. L.; Green, S. M.; Marshall, G. R. VALIDATE: A new method for the receptor-based prediction of binding affinities of novel ligands. *Preprint* **1995**.
- (84) Connelly, P. R. Acquisition and use of calorimetric data for prediction of the thermodynamics of ligand-binding and folding reactions of proteins. *Curr. Opin. Biotech.* **1994**, *5*, 381–388.
- (85) Sippl, M. J. Boltzmann's principle, knowledge-based mean fields and protein folding. An approach to the computational determination of protein structures. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 473–501.
- (86) Mark, A. E.; van Gunsteren, W. F. Decomposition of the free energy of a system in terms of specific interactions. *J. Mol. Biol.* **1994**, *240*, 167–176.
- (87) Janin, J. Elusive affinities. *Proteins: Structure Function and Genetics* **1995**, *21*, 30–39.
- (88) Ku, T. W.; Ali, F. E.; Barton, L. S.; *et al.* Direct design of a potent non-peptide fibrinogen receptor antagonist based on the structure and conformation of a highly constrained cyclic rgd peptide. *J. Am. Chem. Soc.* **1993**, *115*, 8861–8862.
- (89) Weber, P. C.; Pantoliano, M. W.; Simons, D. M.; Salemme, F. R. Structure-based design of synthetic azobenzene ligands for streptavidin. *J. Am. Chem. Soc.* **1994**, *116*, 2717–2724.
- (90) Lee, B. Enthalpy-entropy compensation in the thermodynamics of hydrophobicity. *Biophys. Chem.* **1994**, *51*, 271–277.

JM950484N