Free Energy Decomposition of Protein-Protein Interactions

Sergey Yu. Noskov* and Carmay Lim*†
*Institute of Biomedical Sciences, Academia Sinica, 11529 Taipei, and †Department of Chemistry, National Tsing-Hua University, 300 Hsinchu, Taiwan

ABSTRACT A free energy decomposition scheme has been developed and tested on antibody–antigen and protease–inhibitor binding for which accurate experimental structures were available for both free and bound proteins. Using the x-ray coordinates of the free and bound proteins, the absolute binding free energy was computed assuming additivity of three well-defined, physical processes: desolvation of the x-ray structures, isomerization of the x-ray conformation to a nearby local minimum in the gas-phase, and subsequent noncovalent complex formation in the gas phase. This free energy scheme, together with the Generalized Born model for computing the electrostatic solvation free energy, yielded binding free energies in remarkable agreement with experimental data. Two assumptions commonly used in theoretical treatments; viz., the rigid-binding approximation (which assumes no conformational change upon complexation) and the neglect of vdW interactions, were found to yield large errors in the binding free energy. Protein–protein vdW and electrostatic interactions between complementary surfaces over a relatively large area (1400–1700 Å^2) were found to drive antibody–antigen and protease–inhibitor binding.

INTRODUCTION

To understand molecular recognition of protein surfaces, it is important to obtain the quantitative contributions of the individual forces governing binding affinity and specificity. Consequently, it is useful to develop reliable methods for computing absolute or relative free energies that can dissect the relative magnitudes of different thermodynamics effects. Many such methods have been developed and they fall generally into three classes. The first class includes free energy simulation methods (Kollman, 1996; Tembe and McCammon, 1984) that treat solute and solvent atoms explicitly with an atomic force-field description. These simulation methods have a rigorous statistical mechanics basis and can be used to compute the difference between the binding free energies of two closely related medium-sized compounds (e.g., ligands differing only in a single residue) in solution fairly accurately (Jorgensen and Tirado-Rives, 1995). However, they are less reliable for ligand–protein or protein–protein association due mainly to the problem of sampling the relevant configurations of the solute and solvent (McCammon, 1998).

The second class encompasses empirical approaches that use parameters obtained from a "training set" of known interactions (Andrews et al., 1984; Bohm, 1994; Horton and Lewis, 1992; Weng et al., 1997; Williams et al., 1993). The binding free energy is taken to be a sum of different terms associated with hydrophobic contacts, hydrogen bonds or salt bridges, and conformational entropy loss. The various terms are usually not derived from theoretical backgrounds (e.g., statistical thermodynamics or molecular mechanics), but from available binding data in a statistical manner. The key advantage of empirical methods is their speed. However, these phenomenological scoring functions rely critically on the quality of the training database.

The third class includes methods that treat the solute atoms explicitly and the solvent as a continuum dielectric medium (Kollman et al., 2000). The electrostatic contribution to the solvation free energy is computed using finite difference Poisson–Boltzmann (PB) (Honig and Nicholls, 1995) methods or the Generalized Born (GB) approximation (Still et al., 1990; Qiu et al., 1997), whereas the nonelectrostatic contribution is treated as an empirical function of the solvent-accessible surface area (SASA) (Still et al., 1990). These methods appear promising for certain systems because they represent a good balance between speed (by avoiding sampling of solvent configurations) and accuracy (by incorporating long-range electrostatics, ionic strength, and polarization effects).

Although the computed absolute/relative binding free energies for various systems have been reported to be in close agreement with the respective experimental values, this does not necessarily mean that the underlying theory or scheme for computing the binding free energy is accurate. This is because the agreement may be fortuitous and errors due to experimental issues may mask the errors inherent in computing the various free energy components. For example, in many cases the x-ray structure of the complex, but not its unbound counterparts, is known. In such cases, the binding free energy is often estimated from the atomic coordinates of the complex alone assuming no conformational change upon complexation (rigid-binding approximation) (Novotny et al., 1989; Williams et al., 1991; Horton and Lewis, 1992; Jackson and Sternberg, 1995; Friedman and Honig, 1995; Froloff et al., 1997; Novotny et al., 1997; Olson, 1999). In cases where the atomic coordinates of both free and bound proteins are available, the protein may have

Received for publication 22 January 2001 and in final form 7 May 2001.
Address reprint requests to Carmay Lim, Institute of Biomedical Sciences, Academia Sinica, 11529 Taipei, Taiwan. Tel.: 886-2-2652-3031; Fax: 886-2-2788-7641. E-mail: carmay@gate.sinica.edu.tw. Dr. Noskov's permanent address is Institute of Solution Chemistry, Russian Academy of Sciences, Akademicheskaya str. 1, 153045 Ivanovo, Russia.

0006-3495/01/08/737/14 $2.00

been crystallized at a pH that is significantly different from the pH at which the binding free energy was measured. Furthermore, NMR structures are often solved at low pH. However, the free energy for protein association that is mediated by ionizable residues can be very sensitive to pH (Gibas, 1997; Xavier and Wilson, 1998). Discrepancy between the computed and experimental binding free energies has been attributed to the use of low pH structures rather than shortcomings in the free energy function/calculations, like the neglect of van der Waals (vdW) interactions (Krystek et al., 1993).

Partly due to the above issues, the various theoretical studies have not reached a consensus regarding the major force driving the binding process. Pauling concluded that the cooperation of weak vdW and hydrogen-bonding interactions between complementary surfaces over an area sufficiently large to overcome the disrupting influence of thermal agitation govern protein–protein recognition (Pauling, 1974; Pauling and Pressman, 1945). In contrast, Chothia and Janin suggested that vdW and polar interactions contribute little to complex stability, and it is hydrophobicity (the release of interfacial water) that is the major factor stabilizing protein–protein associations (Chothia and Janin, 1975; Kauzmann, 1959). This view is also implicitly assumed in studies that fit the experimental binding free energy to buried surface areas (Horton and Lewis, 1992). Honig and coworkers and other groups found that nonpolar interactions, represented by a free energy–surface area relationship that is assumed to account for the hydrophobic effect and enhanced interfacial packing, provide the major driving force for MHC class I protein–peptide, trypsin-inhibitor and λcl repressor protein–DNA association (Froloff et al., 1997; Jayaram et al., 1999; Misra et al., 1998). Note that Chothia and Janin as well as Honig and coworkers did not take vdW interactions into account explicitly in computing the absolute binding free energy.

In this work, Scheme 1 (see next section), in conjunction with continuum dielectric models, has been implemented to compute the absolute free energy of protein–protein association. Our initial goals are twofold: to assess the accuracy of our scheme for computing absolute binding free energies, and to assess the validity of the various assumptions in previous theoretical treatments like the rigid-binding approximation or the neglect of vdW interactions and vibrational entropy (Novotny et al., 1989; Williams et al., 1991; Horton and Lewis, 1992; Vajda et al., 1994; Jackson and Sternberg, 1995; Novotny et al., 1997; Froloff et al., 1997; Olson, 1999). To address our first goal, we chose systems for which experimental binding free energies are available for comparison with the computed values, and accurate structures are available to minimize errors due to experimental issues. In this way, any observed discrepancy between the calculated and experimental results would reflect primarily the assumptions or deficiencies in our methodology rather than inaccuracies in the experimental structures. Specifically, we chose systems for which the crystal structures of both the complex and its components have been solved to ≤1.9-Å resolution to eliminate errors arising from poor quality structures or absence of free protein structures. Having the x-ray structures of the unbound and bound proteins also enable us to account for conformational changes that accompany binding (especially at the interaction surfaces). Also, we verified that the pH at which the free and complex structures were solved is close to the pH at which the corresponding binding free energy was measured to minimize errors due to structural changes at different pH values. To address our second goal, the systems chosen satisfied not only the above criteria, but have also been studied using various approximations.

Only two systems were found to satisfy all of the aforementioned criteria. They are trypsin complexed with bovine pancreatic trypsin inhibitor (BPTI), and the fragment variable (Fv) region of mouse monoclonal antibody, D1.3, bound to hen egg-white lysozyme (HEL). Their experimental binding free energies, Protein Data Bank (PDB) entries (Bernstein et al., 1977) and various properties are listed in Table 1. The interface area is defined as the SASA difference between the complex and its components. Note that both systems correspond to association of positively charged proteins to form a +12e complex with only small changes in conformation, as evidenced by root-mean-square deviation (RMSD) of the backbone atoms in the free x-ray structure from that in the corresponding complex crystal structure of <1 Å (see Table 1). The excellent agreement found between the computed and experimental binding free energies (see Results) allowed us to identify the key forces governing protease-inhibitor and antibody–antigen complexation with interface areas in the range of 1400–1700 Å^2. Finally, we compare our findings with results obtained in previous studies for the same or homologous systems as well as for other protein–protein complexes (see Discussion).

METHODS

Theory

The binding free energy calculations were based on the thermodynamic cycle

$$
\begin{array}{ccccc}
[r^{\min}]_{\mathrm{gas}} & + & [l^{\min}]_{\mathrm{gas}} & \xrightarrow{\Delta G_{\mathrm{gas}}} & [(R\cdot L)^{\min}]_{\mathrm{gas}} \\
\uparrow {\scriptstyle -\Delta G_{\mathrm{isom}}(r)} & & \uparrow {\scriptstyle -\Delta G_{\mathrm{isom}}(l)} & & \downarrow {\scriptstyle \Delta G_{\mathrm{isom}}(R\cdot L)} \\
[r]_{\mathrm{gas}} & + & [l]_{\mathrm{gas}} & & [R\cdot L]_{\mathrm{gas}} \\
\uparrow {\scriptstyle -\Delta G_{\mathrm{solv}}(r)} & & \uparrow {\scriptstyle -\Delta G_{\mathrm{solv}}(l)} & & \downarrow {\scriptstyle \Delta G_{\mathrm{solv}}(R\cdot L)} \\
[r]_{\mathrm{sln}} & + & [l]_{\mathrm{sln}} & \xrightarrow[\Delta G^{\circ}]{} & [R\cdot L]_{\mathrm{sln}}
\end{array}
$$

In Scheme 1, lower case r and l denote the unbound receptor and ligand conformations, whereas upper case R and L denote the bound receptor and ligand conformations. Note that $[x]_{\mathrm{sln}}$ and $[x]_{\mathrm{gas}}$, where $x = r$, l, or $R \cdot L$,

TABLE 1 Properties of protein–protein complexes studied in this work

Property	D1.3·HEL	D1.3	HEL	Trypsin·BPTI	Trypsin	BPTI
ΔG_{exp} (kcal/mol)*		-11.4 ± 0.1[†]			-18.1 ± 0.1[‡]	
(pH of binding)		(7.1)			(8.0)	
PDB entry	1vfb	1vfa	1dpx	2ptc	2ptn	1bpi
Resolution (Å)	1.80	1.80	1.65	1.90	1.55	1.10
pH of crystallization	7.1	7.1	8.0	7.0	7.0	8.2
Number of atoms	5,384	3,424	1,960	4,112	3,220	892
Net charge	12	4	8	12	6	6
SASA (Å^2)[§]	15,049	10,027	6,470	11,441	9,095	4,013
Conformational		1.07	1.27		0.83	1.23
Change (Å)[¶]		(0.87)	(0.88)		(0.61)	(0.93)

*Experimental binding free energy, ΔG_{exp}, in kcal/mol, at 298 K and the pH of the ΔG_{exp} measurement in brackets.
[†]From Verhoeyen et al. (1988).
[‡]From Vincent & Lazdunski (1972).
[§]SASA values are based on the x-ray structures.
[¶]The RMSD of the heavy atoms (or backbone atoms) of the bound protein from the free one.

correspond to the same structure; viz., the relaxed "all-hydrogen" x-ray structure (see next section). The latter was energy minimized in the gas phase to yield $[x^{min}]_{gas}$, which corresponds to a local energy minimum that is required to compute gas-phase vibrational frequencies and thus the vibrational entropy change accompanying binding (see Gas-phase Binding Free Energy).

The standard free energy change (ΔG^0) for the noncovalent association of two molecules in solution according to Scheme 1 is given by

$$\Delta G^\circ = \Delta\Delta G_{solv} + \Delta\Delta G_{isom} + \Delta G_{gas}, \quad (1)$$

where

$$\Delta\Delta G_{solv} = \Delta G_{solv}(R \cdot L) - \Delta G_{solv}(r) - \Delta G_{solv}(l) \quad (2)$$

and

$$\Delta\Delta G_{isom} = \Delta G_{isom}(R \cdot L) - \Delta G_{isom}(r) - \Delta G_{isom}(l). \quad (3)$$

The desolvation free energy, $-\Delta G_{solv}$, corresponds to the work of transferring the molecule in its solution conformation to the same conformation in the gas phase at 298 K. The isomerization free energy, $-\Delta G_{isom}$, is the work of transforming the molecule in its x-ray conformation to a nearby local minimum structure in the gas phase. ΔG_{gas} is the standard free energy change per mole for the noncovalent association of the molecules in the gas phase at 298 K. The calculations of ΔG_{solv}, ΔG_{isom}, and ΔG_{gas} are described below.

Structures

In computing the absolute free energy according to Scheme 1, x-ray structures of the free proteins and their respective complex were used (see Table 1). The ionizable groups in the free proteins and respective complex were protonated or deprotonated according to available experimental pK_a values and the pH of crystallization. All aspartic acid and glutamic acid residues as well as COOH-terminal groups were deprotonated, whereas arginine, lysine, cysteines, and NH_2-terminal groups were protonated. Histidine residues were protonated only if their side-chain nitrogen atoms were within 3.5 Å of a hydrogen acceptor in the crystal structures. However, for the proteins studied here, the histidine side chains were found not to be involved in hydrogen bonding. Thus, they were treated as neutral by protonating at $N^{\epsilon 2}$ or $N^{\delta 1}$ according to their local environment in the crystal structure. In trypsin, His^{40} and His^{91} were protonated at $N^{\epsilon 2}$, but His^{57} was protonated at $N^{\delta 1}$ (Jackson and Sternberg, 1995). BPTI has no histidine residues. For the Fv region of mouse monoclonal D1.3 antibody and lysozyme, all the histidines were protonated at $N^{\epsilon 2}$ (Tanokura, 1983). Only one histidine is located in the interface region in the complex crystal structures; viz., His^{57} in the trypsin/BPTI complex.

First, hydrogen atoms were added to the crystal structure using the HBUILD module of the CHARMM program (Brooks et al., 1983). The resulting structure was subjected to several steps of minimization using steepest descent followed by adopted-basis Newton–Raphson with strong (10 kcal/mol/Å^2) harmonic constraints on all heavy atoms. This relieved close contacts in the protein without disrupting its overall conformation (Philippopoulos and Lim, 1995), as evidenced by RMSDs from the respective crystal structure of $\leq$0.064 Å (see Table 2). The relaxed all-hydrogen x-ray structures, denoted by $[r]_{sln}$, $[l_{sln}$, and $[R \cdot L]_{sln}$, were used to compute the solvation free energies (see below). For the isomerization step, each all-hydrogen x-ray structure was energy minimized (initially with constraints, then without) using a dielectric constant of one for $\geq$2500 steps of steepest descent, followed by ~7500 steps of adopted-basis Newton–Raphson until the average force was $<2 \times 10^{-6}$ kcal/mol/Å. The fully minimized $[r^{min}]_{gas}$, $[l^{min}]_{gas}$, and $[(R \cdot L)^{min}]_{gas}$ structures, which remain close to their respective crystal structure (see Table 2), were used to obtain the gas-phase complexation energies and entropies. All energy minimizations were carried out using the CHARMM program (Brooks et al., 1983) and the version 22 all-hydrogen forcefield (MacKerell et al., 1998).

In computing ΔG^0 using the rigid-binding approximation, the starting conformations of the unbound r and l proteins were obtained from the

TABLE 2 RMSD of heavy atoms from crystal structures*

Complex Structure	Free Structures	D1.3·HEL	D1.3	HEL	Trypsin·BPTI	Trypsin	BPTI
$[(R \cdot L)]_{sln}$	$[r/l]_{sln}$	0.06	0.06	0.04	0.05	0.06	0.05
$[(R \cdot L)]_{sln}$	$[R/L]_{sln}$	0.06	1.05	1.29	0.05	0.97	1.20
$[(R \cdot L)]_{sln}$	$[R/L^{min+cons}]_{sln}$	0.06	1.05	1.25	0.05	0.97	1.18
$[(R \cdot L)^{min}]_{gas}$	$[r/l^{min}]_{gas}$	1.23	1.15	1.48	1.37	1.23	1.35

*The RMSD of the heavy atoms in the free or bound protein structure from the respective x-ray structure (see Methods).

relaxed all-hydrogen x-ray structure of the $R \cdot L$ complex; i.e., $[r]_{sln} = [R]_{sln}$ and $[l]_{sln} = [L]_{sln}$. These rigid free structures exhibit heavy atom RMSDs from the respective crystal structures that range from 0.97 to 1.29 Å (see Table 2). The latter value is greater than the motion of protein crystal atoms, 0.3–0.5 Å, derived from their B-factors (Froloff et al., 1997; Luzzati, 1952). Thus, in the rigid-binding approximation, ΔG^0 is computed according to

$$\begin{array}{ccccc} [R^{min}]_{gas} & + & [L^{min}]_{gas} & \xrightarrow{\Delta G_{gas}} & [(R\cdot L)^{min}]_{gas} \\ \uparrow -\Delta G_{isom}(R) & & \uparrow -\Delta G_{isom}(L) & & \downarrow \Delta G_{isom}(R\cdot L) \\ [R]_{gas} & + & [L]_{gas} & & [R\cdot L]_{gas} \\ \uparrow -\Delta G_{solv}(R) & & \uparrow -\Delta G_{solv}(L) & & \downarrow \Delta G_{solv}(R\cdot L) \\ [R]_{sln} & + & [L]_{sln} & \xrightarrow[\Delta G^\circ]{} & [R\cdot L]_{sln} \end{array}$$

The ΔG^0 was also evaluated using free receptor and free ligand structures that have been modeled from the respective complex by minimizing the receptor R and ligand L conformations using CHARMM with a distance-dependent dielectric constant and constraints on all heavy atoms (denoted by min + cons in Scheme 3). In this case, $[r]_{sln} = [R^{min+cons}]_{sln}$ and $[l]_{sln} = [L^{min+cons}]_{sln}$. Their heavy-atom RMSDs from the respective crystal structures are similar to those of the rigid structures (see Table 2).

$$\begin{array}{ccccc} [R^{min}]_{gas} & + [L^{min}]_{gas} & \xrightarrow{\Delta G_{gas}} & [(R\cdot L)^{min}]_{gas} \\ \uparrow -\Delta G_{isom}(R^{min+cons}) & \uparrow -\Delta G_{isom}(L^{min+cons}) & & \downarrow \Delta G_{isom}(R\cdot L) \\ [R^{min+cons}]_{gas} & + [L^{min+cons}]_{gas} & & [R\cdot L]_{gas} \\ \uparrow -\Delta G_{solv}(R^{min+cons}) & \uparrow -\Delta G_{solv}(L^{min+cons}) & & \downarrow \Delta G_{solv}(R\cdot L) \\ [R^{min+cons}]_{sln} & + [L^{min+cons}]_{sln} & \xrightarrow[\Delta G^\circ]{} & [R\cdot L]_{sln} \end{array}$$

Calculations

Solvation free energy ΔG_{solv}

The standard solvation free energy, ΔG_{solv}, can be expressed as a sum of 3 terms,

$$\Delta G_{solv} = \Delta G^{cav}_{solv} + \Delta G^{vdW}_{solv} + \Delta G^{elec}_{solv} \quad (4)$$

The first two terms in Eq. 4 constitute the nonelectrostatic contribution (ΔG^{nonel}_{solv}) to the solvation free energy due to the work required to create the solute cavity in solution (ΔG^{cav}_{solv}) and to vdW interactions between the solute and solvent (ΔG^{vdW}_{solv}). The last term in Eq. 4 gives the electrostatic contribution (ΔG^{elec}_{solv}) to the solvation free energy.

The nonelectrostatic solvation free energy, ΔG^{nonel}_{solv}, was approximated by a linear function of the SASA (Lee and Richards, 1971),

$$\begin{aligned} \Delta G^{nonel}_{solv} &= \Delta G^{vdW}_{solv} + \Delta G^{cav}_{solv} \\ &\approx (\gamma^{vdW} + \gamma^{cav}) \times SASA = \gamma \times SASA. \end{aligned} \quad (5)$$

γ was set to 7.2 cal/mol/Å^2. This value in Eq. 5 and the GB model (Eq. 6 below) could reproduce the experimental hydration free energies of a wide range of neutral small molecules (Still et al., 1990) and the experimental free energies of dihydrofolate reductase and trypsin binding (Zou et al., 1999). It is also in accord with that (6.4 ± 1.7 cal/mol/Å^2) (Friedman and Honig, 1995) derived from the partition coefficients of linear alkanes between water and the gas phase using AMBER radii to compute the SASA. γ^{vdW} was set to −38.8 cal/mol/Å^2 (Jayaram et al., 1999), the value obtained from experimental vaporization enthalpies of hydrocarbons (Ohtaki and Fukushima, 1992). The difference between γ and γ^{vdW} yields γ^{cav} = 46.0 cal/mol/Å^2 (Sharp et al., 1991). The SASA of the molecule was computed using the GEPOL program (Pascual-Ahuir and Silla, 1990) and CHARMM vdW radii (MacKerell et al., 1998). Hence, $\Delta\Delta G^{nonel}_{solv}$ is proportional to the loss in SASA at the protein–protein interface.

The electrostatic solvation free energy, ΔG^{elec}_{solv}, was computed using two continuum solvent approaches: the analytical GB model (Qiu et al., 1997; Still et al., 1990) and numerical PB methods (Honig and Nicholls, 1995). The former is denoted by $\Delta G^{elec}_{solv,GB}$, whereas the latter by $\Delta G^{elec}_{solv,PB}$. In the GB model (Still et al., 1990), $\Delta G^{elec}_{solv,GB}$ is expressed as a sum of Coulombic interactions between each pair of charges (q_i, q_j) in a solvent of dielectric constant ϵ and the Born self solvation energy of each individual charge,

$$\Delta G^{elec}_{solv,GB} = -166\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{n}\sum_{j=1}^{n}\frac{q_i q_j}{f_{GB}},$$

$$f_{GB} = \sqrt{r_{ij}^2 + \alpha_i\alpha_j \exp\left\{-\frac{r_{ij}^2}{4\alpha_i\alpha_j}\right\}}. \quad (6)$$

f_{GB} is an effective distance function that depends on the interatomic distances (r_{ij}) and the effective Born radii of atoms i and j (α_i, α_j). In computing the electrostatic solvation free energies, ϵ was set to 80, the dielectric constant for bulk water, and the atomic charges were taken from the CHARMM version 22 forcefield (MacKerell et al., 1998). The effective Born radii were computed using an empirical scheme first proposed by Still and co-workers (Qiu et al., 1997) and modified for protein solvation calculations by Dominy and Brooks (1999).

The effective Born radius of an atom depends on the atom's specific molecular environment (Babu and Lim, 1999, 2001). It is defined as the atomic radius that would give the Born solvation free energy (Born, 1920) of the atom with unit charge, if all other atoms in the molecule were uncharged and served only to displace the solvent dielectric medium; i.e.,

$$\alpha_i = -\frac{166}{G_{pol,i}}, \quad (7)$$

where

$$G_{pol,i} = \left(1 - \frac{1}{\varepsilon}\right)\left[\frac{1}{\lambda}\left(-\frac{166}{R_{vdW,i}}\right) + P_1\left(\frac{166}{R^2_{vdW,i}}\right) + \sum_{j}^{Bond}\frac{P_2 V_j}{r_{ij}^4} + \sum_{j}^{Angle}\frac{P_3 V_j}{r_{ij}^4} + \sum_{j}^{Nonbond}\frac{P_4 V_j}{r_{ij}^4}\,CCF\right] \quad (8)$$

and

$$CCF = 1.0, \quad \text{if} \quad \left(\frac{r_{ij}}{(R_{vdW,i} + R_{vdW,j})}\right)^2 > \frac{1}{P_5}, \quad (9a)$$

$$CCF = \left\{0.5\left[1.0 - \cos\left(\left(\frac{r_{ij}}{(R_{vdW,i} + R_{vdW,j})}\right)^2 P_5\pi\right)\right]\right\}^2 \quad \text{if} \quad \left(\frac{r_{ij}}{(R_{vdW,i} + R_{vdW,j})}\right)^2 \le \frac{1}{P_5} \quad (9b)$$

In Eq. 8, r_{ij} is the distance from the point charge of interest, i, to an uncharged atom j in the molecule; $R_{vdW,i}$ is the vdW radius of atom i, and $V_j = \frac{4}{3}\pi R^3_{vdW,j}$ is the volume of atom j. The vdW radii were taken from the CHARMM version 22 all-hydrogen forcefield with the polar hydrogen radii set to 0.8 Å following previous works (Dominy and Brooks, 1999; Qiu et al., 1997). P_1, P_2, P_3, P_4, P_5 and λ are fitting parameters. They have been optimized by Dominy and Brooks (1999) for a combined protein and nucleic acid structure database and the CHARMM version 22 forcefield. The parameters used in this work are $P_1 = 0.448$, $P_2 = 0.173$, $P_3 = 0.013$, $P_4 = 9.015$, $P_5 = 0.9$, and $\lambda = 0.705$ (Dominy and Brooks, 1999). The reader is referred to the original works for the derivation of Eq. 8 (Qiu et al., 1997) and the fitting procedure to obtain the parameters (Dominy and Brooks, 1999).

The ΔG^{elec}_{solv} energies have also been evaluated by solving numerically the linearized Poisson–Boltzmann equation using the DelPhi program (Gilson and Honig, 1988). The protein was treated as a low dielectric medium ($\epsilon_{int} = 1$) surrounded by a high dielectric solvent ($\epsilon_{out} = 80$ for water). The ionic strength was set to zero, but increasing it to 0.1 M had little effect on the magnitude of the solvation free energy, as found in previous works (Froloff et al., 1997; Jackson and Sternberg, 1995). The low-dielectric region of the protein was defined as the region inaccessible to contact by a 1.4-Å sphere rolling over the molecular surface, defined by the atomic coordinates and vdW radii. The vdW radii and atomic charges used in the DelPhi calculations were taken from the CHARMM version 22 forcefield. The electrostatic potentials were calculated on a cubic grid with a spacing of 0.5 Å and a 90% grid fill to avoid grid boundary artifacts. The difference between the electrostatic potential calculated in solution ($\epsilon_{out} = 80$) and in the gas phase ($\epsilon_{out} = 1$) yielded the electrostatic contribution to the solvation free energy.

Isomerization free energy ΔG_{isom}

The isomerization free energy for each molecule was estimated from the energy difference between the relaxed all-hydrogen structure and the respective fully minimized structure (see above); i.e.,

$$\Delta G_{isom} \approx \Delta E_{isom} = E(X) - E(X^{min}), \tag{10}$$

where $X = r$, l, or $R \cdot L$, and

$$E \approx E^{vdW} + E^{elec} \tag{11}$$

The energy E was calculated using the CHARMM program (Brooks et al., 1983) with the version 22 all-hydrogen forcefield (MacKerell et al., 1998) with $\epsilon = 1$ and an atom-based force-shifting function, which shifts the nonbonded forces to zero at 12 Å.

Gas-phase binding free energy, ΔG_{gas}

The standard gas-phase free energy change is given by

$$\Delta G_{gas} = \Delta E_{gas} + p\Delta V - T(\Delta S^{trans} + \Delta S^{rot} + \Delta S^{vib} + \Delta S^{conf}) \tag{12}$$

where the intramolecular energy change, ΔE_{gas} is computed from Eqs. 11 and 13,

$$\Delta E_{gas} = E((R \cdot L)^{min}) - E(r^{min}) - E(l^{min}) \tag{13}$$

and $p\Delta V = -RT = -0.6$ kcal.mol at 298 K.

The gas-phase translational and rotational entropies for the free proteins and their complex were calculated from ideal gas partition functions (Q_{trans}, Q_{rot}) using classical statistical mechanics (McQuarrie, 1976).

$$TS^{trans} = \left[\frac{5}{2} + \frac{3}{2}\ln\left(\frac{2\pi m k_B T}{h^2}\right) - \ln(\rho)\right]k_B T,$$
$$TS^{rot} = \left[\frac{3}{2} + \frac{1}{2}\ln(\pi I_a I_b I_c) + \frac{3}{2}\ln\left(\frac{8\pi^2 k_B T}{h^2}\right) - \ln(\sigma)\right]k_B T. \tag{14}$$

In Eq. 14, m is the total mass of the molecule, k_B is Boltzmann's constant, T is the temperature (298 K), h is Planck's constant, ρ is the number density (1 M per liter), I_A, I_B, I_C are the 3 principal moments of inertia, and σ is the symmetry number (which is equal to 1 for nonsymmetric molecules).

The gas-phase vibrational entropies for the free proteins and the respective complex were calculated from normal mode frequencies ν_j based on the expression (McQuarrie, 1976):

$$TS_{vib} = \sum_{j=1}^{3N-6} \frac{h\nu_j}{e^{h\nu_j/k_B T} - 1} - k_B T \ln(1 - e^{-h\nu_j/k_B T}). \tag{15}$$

The normal mode frequencies ν_j were computed using the CHARMM program (Brooks et al., 1983) with the version 22 all-hydrogen forcefield (MacKerell et al., 1998) using an iterative diagonalization scheme (Janezic and Brooks, 1995), a dielectric constant of one, and a cutoff of 12 Å to truncate the nonbonded forces. For each structure, there were six zero-frequency modes corresponding to overall translational and rotational motions and no imaginary frequencies.

The conformational entropy change, ΔS^{conf}, was approximated by the loss of side-chain conformational entropy upon binding, which, in turn, was estimated using the empirical scale of Pickett and Sternberg (1993, 1994). This model assumes that solvent-accessible side chains (with relative SASA > 60%) populate different rotamers, whereas buried side chains (with relative SASA ≤ 60%) are restricted to one rotamer. The conformation entropy of a given side chain r($S^{conf,r}$) is given by,

$$S^{conf,r} = -R\sum_{i=1}^{N} p_i \ln(p_i), \tag{16}$$

where p_i is the probability of the side chain in rotameric state i. The values of p_i were obtained from the observed distribution of side-chain rotamers in 50 nonhomologous protein crystal structures, taking into account the effects of symmetry and free rotation (Pickett and Sternberg, 1993).

RESULTS

In computing ΔG^0 according to Schemes 1–3, two assumptions were made: the free energy components in Eq. 1 were assumed to be additive; and they were based on a single time-averaged x-ray structure rather than ensemble averaged. Hence, the present scheme is best suited for the noncovalent binding of two proteins that undergo only slight conformational changes to form a high affinity complex, as in the case here (see Table 1). However, for the binding of flexible peptides to proteins, free energy estimates based on a single conformation of the unbound and bound species may not be appropriate.

Tables 3 and 4 summarize the results for D1.3·Hel and trypsin·BPTI binding, respectively. In each case, the binding free energy was computed using three different sets of structures for the free proteins (see Methods): the relaxed all-hydrogen x-ray structures (Scheme 1); rigid $[r]_{sln}$ =

TABLE 3 Free energy and its components for the binding of the Fv region of mouse monoclonal antibody D1.3 to hen egg-white lysozyme (Hel)*

	This Work				
Scheme§	Scheme 1	Scheme 2	Scheme 3	Rigid†	Rigid‡
Interface area¶	1448	1479	1470	1640	
Total binding free energy‖,**					
ΔG°	**−11.4 (−9.4)**	**38.4 (44.2)**	**5.4 (7.4)**	**−9 ± 3**	
Solvation contributions					
$\Delta\Delta G_{solv}^{cav}$	−66.6	−68.0	−67.6		
$\Delta\Delta G_{solv}^{vdW}$	56.2	57.4	57.0		
$\Delta\Delta G_{solv}^{elec}$**	98.1 (100.1)	118.7 (124.5)	118.1 (120.2)		139.6
$\Delta\Delta G_{solv}$	**87.7 (89.7)**	**108.1 (113.9)**	**107.6 (109.6)**		
Gas phase contributions					
$\Delta\Delta E_{isom}^{vdW}$††	−39.4	31.1	−26.3		
$\Delta\Delta E_{isom}^{elec}$ ††	71.0	148.7	88.7		
ΔE_{gas}^{vdW}	−67.8	−62.1	−52.0		
ΔE_{gas}^{elec}	−106.1	−242.1	−163.0		−104.8
$-T\Delta S^{trans}$	13.4	13.4	13.4		
$-T\Delta S^{rot}$	14.1	14.1	14.1		
$-T\Delta S^{vib}$‡‡	7.0	18.5	14.2		
$-T\Delta S^{conf}$	9.3	9.3	9.3	42 ± 3	
$\Delta G_{gas+isom}$§§	**−99.1**	**−69.7**	**−102.2**		
Net component contributions					
$\Delta\Delta G_{solv}^{nonel}$¶¶	−10.4	−10.6	−10.6	−41	
ΔE^{vdW}‖‖	−107.2	−31.0	−78.3		
ΔG^{elec}***	63.0 (65.0)	25.3 (31.1)	43.9 (45.9)	−21	34.8
$-T\Delta S^{trans+rot+vib}$†††	34.5	46.0	41.7	9	

*All energies are in kcal/mol; a blank means that the value is not available.
†From Novotny et al. (1989). The authors also computed a cratic entropy term, $-T\Delta S_{cratic}$, which was set to 2 kcal/mol.
‡Data for homologous HyHel10-HEL system from Novotny et al. (1997). Note that an internal dielectric constant $\varepsilon_{int} = 4$ was used to solve the PB equation, whereas an $\varepsilon_{int} = 1$ was used in this work.
§See Methods and Table 2; "rigid" implies that the rigid-binding approximation was used, see text.
¶The difference between the solvent-accessible surface areas of bound and free proteins in Å^2.
‖$\Delta G^\circ = \Delta\Delta G_{solv,PB/GB} + \Delta G_{gas+isom}$; see Methods.
**The numbers without and with parentheses were computed using the GB model and finite difference PB methods, respectively.
††ΔE_{isom} is the energy required to isomerize the local energy minimum in the gas phase to the respective relaxed all-hydrogen X-ray structure.
‡‡The TS values of the individual species are taken from the last column of Table 6.
§§$\Delta G_{gas+isom} = \Delta\Delta E_{isom} + \Delta G_{gas}$ using Eqs. 10–12.
¶¶$\Delta\Delta G_{solv}^{nonel} = \Delta\Delta G_{solv}^{cav} + \Delta\Delta G_{solv}^{vdW}$ using Eq. 5.
‖‖$\Delta E^{vdW} = \Delta\Delta E_{isom}^{vdW} + \Delta E_{gas}^{vdW}$.
***$\Delta G^{elec} = \Delta\Delta G_{solv}^{elec} + \Delta\Delta E_{isom}^{elec} + \Delta E_{gas}^{elec}$.
†††$T\Delta S^{trans+rot+vib} = T\Delta S^{trans} + T\Delta S^{rot} + T\Delta S^{vib}$.

$[R]_{sln}$ and $[l]_{sln} = [L]_{sln}$ structures (Scheme 2); and $[r]_{sln} = [R^{min+cons}]_{sln}$ and $[l]_{sln} = [L^{min+cons}]_{sln}$ (Scheme 3). In what follows, we will first compare the free energies computed using the relaxed all-hydrogen x-ray structures with experiment. Agreement with experiment then allows us to assess the relative contributions from protein–protein versus protein–solvent versus solvent–solvent interactions and electrostatic versus vdW forces to the binding free energy. We then evaluate the accuracy of binding free energies computed using rigid as well as $[R^{min+cons}]_{sln}$ and $[L^{min+cons}]_{sln}$ free protein structures.

Comparison between computed and experimental ΔG_{exp}: GB versus PB

Tables 3 and 4 show that the ΔG_{GB}^0 (−11.4 kcal/mol for D1.3·Hel and −18.6 kcal/mol for trypsin·BPTI) based on $[r]_{sln}$, $[l]_{sln}$, and $[R \cdot L]_{sln}$ relaxed x-ray structures and $\Delta\Delta G_{solv,GB}^{elec}$ are in excellent agreement with the experimental values (−11.4 and −18.1 kcal/mol). In contrast, the ΔG_{PB}^0 based on $[r]_{sln}$, $[l]_{sln}$, and $[R \cdot L]_{sln}$ relaxed x-ray structures and $\Delta\Delta G_{solv,PB}^{elec}$ exhibit larger deviations from the experimental numbers than ΔG_{GB}^0, whereas the ΔG_{GB}^0 and ΔG_{PB}^0 differ by roughly 20%. The remarkable agreement between ΔG_{GB}^0 and experiment indicates that systematic errors involved in computing free energies/energies in the right and left legs of Scheme 1 have largely cancelled. Because the $\Delta G_{solv,GB}^{elec}$ based on x-ray structures are more negative than the respective $\Delta G_{solv,PB}^{elec}$ by $\leq$7.6%, but the CPU time needed to compute $\Delta G_{solv,GB}^{elec}$ is 5 to 6 times less than that for $\Delta G_{solv,PB}^{elec}$, the GB formulation together with Scheme 1 appears to be an efficient and reliable way of computing binding free energies and obtaining trends (see below).

TABLE 4 Free energy and its components for the binding of trypsin to BPTI*

	This Work					
Free Structures	Scheme 1	Scheme 2	Scheme 3	Rigid†	Rigid‡	Rigid§
Interface area	1667	1542	1538	1500	1573	1231
Total binding free energy						
ΔG°	**−18.6 (−23.1)**	**−7.5 (−4.4)**	**−34.1 (−31.1)**	**−21.7**	**−10.7**	**−15.7**
Solvation contributions						
$\Delta\Delta G_{solv}^{cav}$	−76.7	−70.9	−70.7		−61.8	−73.3
$\Delta\Delta G_{solv}^{vdW}$	64.7	59.8	59.7			
$\Delta\Delta G_{solv}^{elec}$	84.6 (80.1)	198.0 (201.1)	180.5 (183.5)		119.0	−101.9
$\Delta\Delta G_{solv}$	**72.6 (68.1)**	**186.9 (190.0)**	**169.4 (172.4)**			
Gas phase contributions						
$\Delta\Delta E_{isom}^{vdW}$	44.3	42.1	21.7			
$\Delta\Delta E_{isom}^{elec}$	166.1	132.0	41.1			
ΔE_{gas}^{vdW}	−105.8	−86.0	−80.4			
ΔE_{gas}^{elec}	−228.3	−314.7	−213.1			
$-T\Delta S^{trans}$	12.6	12.6	12.6			
$-T\Delta S^{rot}$	13.6	13.4	13.4			
$-T\Delta S^{vib}$	−5.0	−2.7	−7.6			
$-T\Delta S^{conf}$	11.9	9.4	9.4	36.0	9.6	9.8
$\Delta G_{gas+isom}$	**−91.3**	**−194.4**	**−203.5**			
Net component contributions						
$\Delta\Delta G_{solv}^{nonel}$	−12.0	−11.1	−11.1	−37.5		
ΔE^{vdW}	−61.6	−43.8	−58.7			
ΔG^{elec}	22.5 (18.0)	15.3 (18.4)	8.5 (11.5)	−30.0	55.4	47.8
$-T\Delta S^{trans+rot+vib}$	21.2	23.4	18.5	9.0		

*See footnotes under Table 3.
†From Krystek et al. (1993).
‡From Jackson and Sternberg (1995).
§From Polticelli et al. (1999).

Solvation versus gas-phase contributions to binding affinity

Scheme 1 enables the net binding free energy to be dissected into solvation (protein–solvent and solvent–solvent) versus gas-phase (protein–protein) contributions. The latter, which is a sum of ΔG_{gas} and $\Delta\Delta G_{isom}$ terms, is negative, thus favoring complexation, whereas $\Delta\Delta G_{solv}$ is positive and opposes binding (Tables 3 and 4). For both systems, gas-phase vdW and electrostatic protein–protein interactions favor binding (ΔE_{gas}^{vdW} and ΔE_{gas}^{elec} negative). Note that ΔE_{gas}^{elec} is negative even though the receptor and ligand have net positive charges (Table 1). The solvent–solvent cavitation term (Eq. 5) also favors protein–protein complexation ($\Delta\Delta G_{solv}^{cav}$ negative) as less work is required to create the complex cavity than the free protein cavities in solution. In contrast to the favorable protein–protein and solvent–solvent interactions, vdW and electrostatic protein–solvent interactions generally oppose binding ($\Delta\Delta G_{solv}^{vdW}$ and $\Delta\Delta G_{solv}^{elec}$ positive) due to the cost of desolvating the unbound proteins.

Decomposition of ΔG^0 into component energies

Tables 3 and 4 show that, for both systems, $\Delta\Delta G_{solv}^{nonel}$ (−10.4 kcal/mol for D1.3·Hel and −12.0 kcal/mol for trypsin·BPTI) roughly cancels the gas-phase conformational entropy term, $-T\Delta S^{conf}$ (9.3 kcal/mol for D1.3·Hel and 11.9 kcal/mol for trypsin·BPTI). This finding can be rationalized if $\Delta\Delta G_{solv}^{nonel} \approx -T\Delta\Delta S_{solvent}$, based on the fact that $\Delta\Delta G_{solv}^{nonel}$ is related to the hydrophobic effect, which, at room temperature, is dominated by the solvent entropy term. The observed cancellation can then be attributed to the increase in the solvent entropy upon complexation (as protein side chains at the interface release water molecules), and the concomitant decrease in the solute entropy (as protein side chains at the interface lose torsional degrees of freedom upon interacting with other residues) (Novotny et al., 1989). To verify whether the observed cancellation of $\Delta\Delta G_{solv}^{nonel}$ and $T\Delta S^{conf}$ is general, these two quantities were also computed for other systems for which x-ray structures are available for both the free and bound proteins (although the structures for these systems may not be as accurate because they do not satisfy all the criteria specified in the Introduction, $\Delta\Delta G_{solv}^{nonel}$ and $T\Delta S^{conf}$ are not very sensitive to the structures because they depend on the SASA, see Tables 3 and 4). The results in Table 5 show that $\Delta\Delta G_{solv}^{nonel}$ and $-T\Delta S^{conf}$ oppose each other and their difference is less than 1.5 kcal/mol, indicating that $\Delta\Delta G_{solv}^{nonel}$ and $-T\Delta S^{conf}$ generally offset one another. It should be emphasized that the observed $\Delta\Delta G_{solv}^{nonel}$ and $T\Delta S^{conf}$ compensation rests on the magnitude of the coefficient γ, 7.2 cal/mol/Å^2, used in this work (see below).

TABLE 5 Changes in the side-chain conformational entropy ($T\Delta S^{conf}$) and nonelectrostatic solvation free energy ($\Delta\Delta G^{nonel}_{solv}$) in various protein–protein associations

Receptor*	Proteinase B	D1.3	α-Chymotrypsin	Subtilisin Carlsberg	Trypsin	Thermitase	Subtilisin Novo
	(1sgb)	(1vfa)	(2cha)	(1sbc)	(2ptn)	(1thm)	(1sup)
Ligand*	OMTKY3	HEL	OMTKY3	Eglin C	BPTI	Eglin C	CI-2
	(1omt)	(1dpx)	(1omt)	(1egl)	(1bpi)	(1egl)	(2ci2)
Complex	3sgb	1vfb	1cho	1cse	2ptc	1tec	2sni
Interface area†	1447	1448	1632	1640	1667	1719	1795
$\Delta\Delta G^{nonel}_{solv}$‡	−10.4	−10.4	−11.8	−11.8	−12.0	−12.3	−12.9
$-T\Delta S^{conf}$‡	8.9	9.3	10.4	10.7	11.9	11.2	12.1
$\Delta\Delta G^{nonel}_{solv}$ $-T\Delta S^{conf}$	−1.5	−1.1	−1.4	−1.1	−0.1	−1.1	−0.8

*PDB entry in brackets.
†The difference between the solvent-accessible surface areas of bound and free proteins in Å^2.
‡In kcal/mol.

Because the $\Delta\Delta G^{nonel}_{solv}$ and $-T\Delta S^{conf}$ terms offset one another in Tables 3 and 4, the remaining components of the binding affinity can be partitioned into net gas-phase protein–protein vdW effects (ΔE^{vdw}), net protein–protein and protein–solvent electrostatic effects (ΔG^{elec}), and the sum of translational, rotational, and vibrational entropy changes ($-T\Delta S^{trans+rot+vib}$). The magnitudes of ΔE^{vdW}, ΔG^{elec}, and $T\Delta S^{trans+rot+vib}$ in Tables 3 and 4 are sizable, indicating that all three components contribute to the overall binding affinity, so that

$$\Delta G^{\circ} \sim \Delta E^{vdw} + \Delta G^{elec} - T\Delta S^{trans+rot+vib}. \quad (17)$$

The ΔG^{0}_{GB} computed using Eq. 17 are −9.7 kcal/mol for D1.3·Hel and −17.9 kcal/mol for trypsin·BPTI, which underestimate the experimental numbers by 1.7 and 0.2 kcal/mol, respectively. Of the three terms in Eq. 17, only the gas-phase protein–protein vdW interactions favor binding (ΔE^{vdw} negative). Furthermore, the magnitude of ΔE^{vdw} is larger than that of ΔG^{elec} or $T\Delta S^{trans+rot+vib}$ for both D1.3·Hel and trypsin·BPTI complexation. These findings are consistent with the experimental observation that enthalpy drives D1.3·Hel binding from measurements of enthalpy and entropy changes by titration calorimetry (Bhat et al., 1994). Hence, close packing and shape complementarity play important roles in noncovalent protein association. The net electrostatic interactions generally oppose formation of complexes with an interface area between 1400 and 1700 Å^2 (ΔG^{elec} positive). This finding does not mean that electrostatic interactions are unimportant for protein binding, but shows that the gain in hydrogen bonding and charge–charge protein–protein interactions across the interface is offset by the loss of favorable electrostatic protein–solvent interactions. Although translational and rotational entropy loss inevitably oppose complexation ($\Delta S^{trans+rot}$ negative), vibrational entropy can favor or disfavor protein–protein association: ΔS_{vib} is negative for D1.3·Hel complexation, but is positive for trypsin–BPTI binding (see also Discussion).

Because the magnitude of the $T\Delta S^{trans+rot}$ term is similar for two ligands of roughly equal volumes, l and l', binding to a common receptor r, their binding free energy difference will be given by

$$\Delta\Delta G^{\circ} \sim \Delta E^{vdw} - \Delta E'^{vdw} + \Delta G^{elec} - \Delta G'^{elec} + T\Delta S^{vib} - T\Delta S'^{vib}, \quad (18)$$

where the prime denotes the thermodynamic change for binding of l'. Hence, the discrimination between two similarly shaped ligands, which defines specificity, is governed by close packing of the protein surfaces so that the proper hydrogen bonds can be formed (McCammon, 1998). Eq 18 shows that the affinity of a drug ligand l' for its receptor protein with an interface area between 1400 and 1700 Å^2 can be improved by mutations that increase the magnitude of $\Delta E'^{vdw}$, but decrease the magnitude of $\Delta G'^{elec}$.

Evaluation of the rigid-binding approximation

The RMSDs of the heavy atoms in the free x-ray structure from the corresponding complex crystal structure is <1.3 Å for both systems studied (Table 1). Therefore, the structural changes upon complexation are relatively small and it seems reasonable, to a first approximation, to neglect any conformational changes upon protein–protein association; i.e., to assume $[r]_{sln} = [R]_{sln}$ and $[l]_{sln} = [L]_{sln}$. In fact, previous studies used this rigid-binding approximation to compute the free energy for D1.3·Hel and trypsin·BPTI binding (Tables 3 and 4). The binding free energy computed using Scheme 2 does not agree with the experimental value (Tables 3 and 4, column 3), and is even quite positive for the binding of D1.3 antibody to Hel. The discrepancy between theory and experiment arises from both protein–protein as well as protein–solvent electrostatic interactions. The rigid free receptor and free ligand conformations isomerize to local minima, which have less favorable electrostatic interactions than the local minima derived from the relaxed x-ray structures of the free proteins. Consequently,

the ΔE^{elec}_{gas} values computed using the rigid free receptor and free ligand structures (−242 and −315 kcal/mol) are more negative than the respective numbers in the second column of Tables 3 and 4 (−106 and −228 kcal/mol). In contrast, the $\Delta\Delta G^{elec}_{solv}$ values computed using the rigid free structures (119 and 198 kcal/mol) are more positive than the respective numbers in column 2 of Tables 3 and 4 (98 and 85 kcal/mol).

In the case of D1.3 antibody binding to Hel, the relative component contributions to the binding free energy computed using Scheme 2 do not agree with those obtained using Scheme 1. In particular, Scheme 2 predicts that the dominant contribution to the net binding free energy for D1.3·Hel complexation is not ΔE^{vdW}, but the entropy term, $T\Delta S^{trans+rot+vib}$, in contrast to Scheme 1 and experimental observations (see above). These results show that, even if the structural changes upon complexation are known to be small, the rigid-binding approximation (Scheme 2) will generally not yield accurate binding free energy, and may also not reveal the dominant contribution to the binding free energy.

Evaluation of free $R^{min+cons}$ receptor and free $L^{min+cons}$ ligand structures

In cases where the free receptor or free ligand structures have not been experimentally solved but their complex crystal structure is known, the former can be predicted from the latter by minimizing the receptor and ligand conformations in the complex (see Structures section above) provided that binding results in minimal conformational changes. Like the rigid structures, the $[r]_{sln} = [R^{min+cons}]_{sln}$ and $[l]_{sln} = [L^{min+cons}]_{sln}$ structures predict a positive binding free energy for D1.3·Hel complexation, whereas they severely overestimate the magnitude of the free energy for trypsin·BPTI binding. However, they yield trends in the relative component contributions to the binding free energy that are similar to those based on the relaxed x-ray structures (Tables 3 and 4). The magnitude of the component contributions based on the $[R^{min+cons}]_{sln}$ and $[L^{min+cons}]_{sln}$ structures are in-between the respective numbers based on the x-ray and rigid structures (Tables 3 and 4). This suggests that approximating $[r]_{sln}$ and $[l]_{sln}$ by $[R^{min+cons}]_{sln}$ and $[L^{min+cons}]_{sln}$, respectively, in cases where slight conformational rearrangements accompany complexation, as in the cases studied here, can reveal qualitative features in the binding free energies.

Dependence of component free energies on free and complex structures

The results using Schemes 1–3 show that accurate structures are needed to obtain reliable absolute binding free energies (Sharp, 1998). This sensitivity of the atomic coordinates is expected because electrostatic and vdW interactions depend on the interatomic distances. Hence, the ΔE^{vdw} and ΔG^{elec} terms computed using different free structures vary significantly (see Tables 3 and 4). The vibrational entropy term is also sensitive to the local minimum structure used to compute the frequencies (see Table 6). In contrast, the $\Delta\Delta G^{nonel}_{solv}$ and $T\Delta S^{conf}$ terms computed using different free structures do not vary as much as they depend on the solvent-accessible solvent surface area. These findings are in accord with previous works (Froloff et al., 1997; Jackson and Sternberg, 1995; Novotny et al., 1997).

TABLE 6 Lowest four nonzero vibrational frequencies (cm^{-1}) and vibrational entropies (*TS* in kcal/mol) from normal mode calculations for the fully minimized structures*

Protein	Mode 1	Mode 2	Mode 3	Mode 4	*TS*
D1.3 (X-ray)	3.95	4.70	5.48	5.87	1569.2
D1.3 (rigid)	3.42	4.77	5.41	6.17	1562.1
D1.3 (predicted)	3.65	4.82	5.43	6.09	1564.2
HEL (X-ray)	4.35	5.01	5.36	5.97	718.9
	3.72†	5.29†	6.17†	6.41†	
HEL (rigid)	3.48	5.01	5.12	6.39	737.4
HEL (predicted)	3.96	4.96	5.89	6.46	731.2
D1.3·HEL	2.17	4.62	4.97	5.44	2281.1
Trypsin (X-ray)	4.44	5.91	6.39	7.14	1476.0
Trypsin (rigid)	6.21	6.83	7.14	7.62	1478.3
Trypsin (predicted)	6.15	6.52	6.81	7.49	1473.9
BPTI (X-ray)	6.31	8.24	8.58	9.80	401.2
	6.19‡	7.70‡	8.70‡	9.56‡	
BPTI (rigid)	7.87	8.64	9.98	10.78	401.1
BPTI (predicted)	7.86	9.05	10.67	11.95	400.7
BPTI·Trypsin	2.40	3.61	4.93	5.45	1882.1

*The notations X-ray, rigid, and predicted refer to the structures in the first three rows of Table 2, which have been fully minimized as described in Methods; none of the structures resulted in an imaginary frequency.
†Gibrat and Go (1990).
‡Janezic and Brooks (1995).

DISCUSSION

Below, we compare our above findings with results obtained in previous studies for the same or homologous systems and for other protein–protein complexes.

Comparison of the various free energy decomposition schemes

The scheme used here to compute the free energy for protein–protein complexation is closest in spirit to that used by Jayaram et al. (1999) for protein–DNA complexation. The key difference lies in the isomerization step in Scheme 1, which was introduced to obtain local-minimum structures (with no imaginary frequencies) for normal mode frequency calculations. In contrast, Jayaram et al. computed first the isomerization/adaptation of the protein structure in the free state to the conformation in the complex state (the adaptation energy), and, subsequently, their desolvation (see Scheme 4). As a result, the gas-phase complexation step

corresponds to species that are not in energy minima, thus their normal mode frequencies cannot be computed.

$$
\begin{array}{ccc}
[R]_{\mathrm{gas}} & + [L]_{\mathrm{gas}} & \xrightarrow{\Delta G_{\mathrm{gas}}} [(R \cdot L)]_{\mathrm{gas}} \\
\uparrow {\scriptstyle -\Delta G_{\mathrm{solv}}(r)} & \uparrow {\scriptstyle -\Delta G_{\mathrm{solv}}(l)} & \\
[R]_{\mathrm{sln}} & + [L]_{\mathrm{sin}} & \downarrow {\scriptstyle \Delta G_{\mathrm{solv}}(R \cdot L)} \\
\uparrow {\scriptstyle -\Delta G_{\mathrm{isom}}(r)} & \uparrow {\scriptstyle -\Delta G_{\mathrm{isom}}(l)} & \\
[r]_{\mathrm{sln}} & + [l]_{\mathrm{sln}} & \xrightarrow[\Delta G]{} [R \cdot\ L]_{\mathrm{sln}}
\end{array}
$$

In computing the absolute free energies for D1.3·HEL (Novotny et al., 1989) and trypsin·BPTI binding (Krystek et al., 1993) in Tables 3 and 4, Novotny and coworkers used rigid $[r]_{\mathrm{sln}} = [R]_{\mathrm{sln}}$ and $[l]_{\mathrm{sln}} = [L]_{\mathrm{sln}}$ structures and neglected all vdW interactions, protein–solvent electrostatic interactions, and the vibrational entropy change. Furthermore, they computed the free energy components directly in solution via the expression,

$$\Delta G^{\circ} \approx \Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}} + \Delta G^{\mathrm{elec}} - T\Delta S^{\mathrm{conf}} - T\Delta S^{\mathrm{trans+rot}} - T\Delta S^{\mathrm{cratic}}. \quad (19)$$

$\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$ was computed using Eq. 5 with γ set to 25 cal/mol/Å^2, whereas ΔG^{elec} was computed from a solvent-screened Coulomb potential with $\epsilon = 4r_{\mathrm{ij}}$,

$$\Delta G^{\mathrm{elec}} = \sum_{\mathrm{i}=1}^{\mathrm{n}-1} \sum_{\mathrm{j}=\mathrm{i}+1}^{\mathrm{n}} \frac{q_{\mathrm{i}} q_{\mathrm{j}}}{16\pi r_{\mathrm{ij}}^2}. \quad (20)$$

The $-T\Delta S^{\mathrm{conf}}$ term was estimated by $0.6N$, where N is the number of immobilized side-chain torsions, and the $-T\Delta S^{\mathrm{trans+rot}}$ and $-T\Delta S^{\mathrm{cratic}}$ terms were set to 9 and 2 kcal/mol, respectively.

Like Novotny and coworkers, Jackson and Sternberg (1995) and Polticelli et al. (1999) used the rigid-binding approximation and computed the free energy components for trypsin·BPTI binding directly in solution. Both groups omitted the entropy terms ($T\Delta S^{\mathrm{trans+rot+vib}}$) because their focus was on relative rather than absolute binding affinities. In addition, Jackson and Sternberg (1995) neglected protein–protein and protein–solvent vdW interactions assuming that atoms making interactions at the interface will also make contacts with water in the unbound state. Hence, they assumed that $-T\Delta S^{\mathrm{conf}} \approx \Delta E^{\mathrm{vdw}} + \Delta\Delta G^{\mathrm{vdW}}_{\mathrm{solv}}$ and estimated the binding free energy by

$$\Delta G^{\circ} \approx \Delta\Delta G^{\mathrm{cav}}_{\mathrm{solv}} + \Delta\Delta G^{\mathrm{elec}}_{\mathrm{solv,PB}} + \Delta E^{\mathrm{elec}}_{\mathrm{sln}}. \quad (21)$$

Polticelli et al. (1999) also did not consider protein–protein and protein–solvent vdW interactions explicitly, but assumed that any difference will be implicitly taken into account by γ (see below) (Honig et al., 1993),

$$\Delta G^{\circ} \approx \Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}} + \Delta\Delta G^{\mathrm{elec}}_{\mathrm{solv,PB}} + \Delta E^{\mathrm{elec}}_{\mathrm{sln}} - T\Delta S^{\mathrm{conf}}. \quad (22)$$

The $\Delta\Delta G^{\mathrm{cav}}_{\mathrm{solv}}$, $\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$, $\Delta\Delta G^{\mathrm{elec}}_{\mathrm{solv,PB}}$, and $T\Delta S^{\mathrm{conf}}$ terms in Eqs. 21 and 22 were computed as in this work, except that Jackson and Sternberg (1995) used $\gamma^{\mathrm{cav}} = 42.0$ cal/mol/Å^2, whereas Polticelli et al. (1999) used $\gamma = 58.2$ cal/mol/Å^2 with curvature-corrected accessible area. Furthermore, both groups used atomic charges and radii from the PARSE parameter set (Sitkoff et al., 1994), and a dielectric constant of two for the protein in computing $\Delta\Delta G^{\mathrm{elec}}_{\mathrm{solv,PB}}$, but a dielectric constant of one for the protein was used in this work because the CHARMM vdW radii and atomic charges have been parameterized using $\epsilon = 1$. Note that the $\Delta E^{\mathrm{elec}}_{\mathrm{sln}}$ in Eqs. 20 and 21 corresponds to the electrostatic interactions between the two proteins embedded in a medium of dielectric constant two. In general, the internal dielectric constant of a protein is ambiguous in computing free energy components directly in solution using a continuum dielectric model. However, it can be unambiguously set to unity in computing gas-phase ΔE^{elec} using Schemes 1–3.

$\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$ and $T\Delta S^{\mathrm{conf}}$ compensation

Using Scheme 1, $\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$ was found to offset $-T\Delta S^{\mathrm{conf}}$ for seven systems for which x-ray structures are available for both the free and bound proteins (see Table 5). Novotny and coworkers also found $\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$ (using γ in Eq. 5 equal to 25 cal/mol/Å^2, see above) to be roughly equal to $T\Delta S^{\mathrm{conf}}$ for both D1.3·HEL (Novotny et al., 1989) and trypsin·BPTI binding (Krystek et al., 1993) (see Tables 3 and 4) as well as for the homologous system, HyHEL-5·HEL, and McPC603·phosphocholine (Novotny et al., 1989). Likewise, in computing the change in free energy upon binding of HEL to antibody HyHEL-10 arising from point mutations in HEL, Sharp (1998) found that $\Delta\Delta G^{\mathrm{nonel}}_{\mathrm{solv}}$ (using $\gamma = 25$ cal/mol/Å^2) and $T\Delta S^{\mathrm{conf}}$ contributions compensate to some extent. If this feature is indeed a general one, then the hydrophobic effect and the loss in side-chain conformational entropy do not dictate high binding affinity. Support for this comes from the observed poor correlation between the change in free energy on mutating a side chain to alanine and the change in total (or hydrophobic) side-chain SASA on complex formation (Bogan and Thorn, 1998). Further corroboration comes from the large and favorable (negative) enthalpy changes observed for several protein–protein association reactions (Bhat et al., 1994; Pearce et al., 1996). For example, the binding between barnase and barstar is enthalpically driven with negligible entropy change (Frisch et al., 1997). In contrast, previous works have attributed hydrophobicity to be the major factor stabilizing protein–protein association (Kauzmann, 1959; Chothia and Janin, 1975; Horton and Lewis, 1992; Froloff et al., 1997).

vdW interactions (ΔE^{vdw})

Tables 3 and 4 show that the single largest attractive contribution to the binding free energy in the two systems studied here is ΔE^{vdW}. This is in accord with the free energy perturbation results of Miyamoto and Kollman (1993) for the high-affinity interaction between biotin and streptavidin, whose binding free energy ($\Delta G_{exp} = -18.3$ cal/mol) is similar to that for trypsin and BPTI. In contrast, Novotny and coworkers neglected ΔE^{vdW}, assuming that this term contributes little to the overall free energy for D1.3·HEL and trypsin·BPTI binding. As well, Jackson and Sternberg (1995) neglected protein–protein and protein–solvent vdW interactions, assuming that their sum, $\Delta E^{vdW} + \Delta\Delta G^{vdW}_{solv} \approx -T\Delta S^{conf}$ (see above). However, using Scheme 1, $\Delta E^{vdW} + \Delta\Delta G^{vdW}_{solv}$ for D1.3·HEL binding (−51 kcal/mol) opposes, but does not cancel, the corresponding $-T\Delta S^{conf}$ term (9 kcal/mol), and $\Delta E^{vdW} + \Delta\Delta G^{vdW}_{solv}$ for trypsin·BPTI complexation is positive (3 kcal/mol), and does not even oppose the respective $-T\Delta S^{conf}$ term, which is also positive. In other studies (Horton and Lewis, 1992; Nicholls et al., 1991; Shen, 1997; Vajda et al., 1994; Williams et al., 1991; Wilson et al., 1991) protein–protein vdW interactions at the interface were thought to be balanced by protein–water vdW interactions in the unbound state. Although this may seem to be the case for trypsin·BPTI binding because ΔE^{vdW} is largely offset by $\Delta\Delta G^{vdW}_{solv}$ (−62 versus 65 kcal/mol, Table 4), the former is almost twice the latter for D1.3·HEL complexation (−107 versus 56 kcal/mol, Table 3).

Translational and rotational entropy ($\Delta S^{trans+rot}$)

Using Scheme 1, the loss of translational and rotational degrees of freedom ($-T\Delta S^{trans+rot}$) was found to oppose D1.3·HEL and trypsin·BPTI binding by 26 to 28 kcal/mol. The difference between this value and that (9 kcal/mol, see above) used by Novotny and coworkers (Krystek et al., 1993) is because the latter is an estimate in solution (Page and Jencks, 1971), whereas the $T\Delta S^{trans+rot}$ obtained in this work was computed from Eq. 14 in the gas phase. However, the $T\Delta S^{trans+rot}$ obtained for D1.3·HEL and trypsin·BPTI binding is similar to that computed for insulin dimerization (−27 kcal/mol) (Tidor and Karplus, 1994) and for DNA-EcoRi complexation (−32 kcal/mol) (Jayaram et al., 1999), which are also based on ideal gas partition functions. The similarity of the magnitude of $T\Delta S^{trans+rot}$ computed for the noncovalent association of molecules with varying masses suggests a near cancellation of the mass-dependent terms in the translational and rotational entropy (Eq. 14). Hence, Scheme 1 leads to a binding free energy that is nearly independent of molecular mass (Gilson et al., 1997). Because the magnitude of $T\Delta S^{trans+rot}$ is comparable to that of ΔE^{vdW} and ΔG^{elec}, it should not be neglected in computing absolute binding free energies.

Vibrational entropy (ΔS^{vib})

The four lowest nonzero vibrational frequencies obtained here for HEL and BPTI are in good agreement with those reported in previous studies (Gibrat and Go, 1990; Janezic and Brooks, 1995) (see Table 6). Furthermore, the mean square fluctuations of the backbone atoms obtained from normal mode analyses of HEL and BPTI were found to be in good agreement with those deduced from B-factors (Gibrat and Go, 1990; Janezic and Brooks, 1995). The four lowest frequencies in the D1.3·HEL and trypsin·BPTI complexes were found to be significantly lower than those in the free structures. However, the overall vibrational entropy change opposes D1.3·HEL binding but favors trypsin·BPTI complexation. The finding that vibrational entropy can make opposite contributions to protein–protein association is consistent with previous normal mode analysis studies, which found that vibrational entropy opposes subtilisin-eglin C complexation (Ishida et al., 1998), but favors insulin dimerization (Tidor and Karplus, 1994). Furthermore, the $-T\Delta S^{vib}$ values for D1.3·HEL (7.0 kcal/mol) and trypsin·BPTI (−5.0) binding are similar in magnitude to those for subtilisin-eglin C complexation (9.0 kcal/mol) (Ishida et al., 1998) and insulin dimerization (−7.2 kcal/mol), respectively. Errors in the normal mode vibrational frequencies due to truncation and anharmonic effects are present in both the free and bound states and are expected to cancel, thus they are unlikely to change the key findings of this work.

Net electrostatic interactions (ΔG^{elec})

Using Scheme 1, the net electrostatic effect is to oppose D1.3·HEL and trypsin·BPTI binding (ΔG^{elec} positive). In contrast, Novotny and coworkers found the net ΔG^{elec} to be negative for D1.3·HEL (Novotny et al., 1989) and trypsin·BPTI binding (Krystek et al., 1993) (Tables 3 and 4). This is because they computed the ΔG^{elec} term from a solvent-screened Coulomb potential (Eq. 20) that neglects charge solvation/desolvation effects. In later works, Novotny et al. (1997) computed ΔG^{elec} using PB methods and an internal dielectric constant of 4, which yielded a positive value for the homologous HyHEL-10·HEL system (see Table 3, *last column*). The same (positive ΔG^{elec}) trend was also found by Jackson and Sternberg (1995) and Polticelli et al. (1999) for trypsin·BPTI binding. This feature is observed in the binding of other proteins; e.g., the binding of a series of peptides to MHC class I compounds (Froloff et al., 1997), the binding of DNA to λcI repressor (Misra et al., 1998) and EcoRi endonuclease (Jayaram et al., 1999), and dimer formation by the capsids of three icosahedral viruses (Reddy et al., 1998) and by the GCN4 leucine zipper (Hendsch and Tidor, 1999). Thus, it seems that the net electrostatic interactions generally oppose formation of complexes with an interface area between 1400 and 1700 Å^2, but it is not clear if this is the case for near neutral proteins that form smaller interface areas, which are likely to form weaker complexes.

As mentioned above, electrostatic protein–solvent interactions are found to oppose binding (positive $\Delta\Delta G^{elec}_{solv}$), whereas hydrogen bonding or charge–charge protein–protein interactions drive complexation (negative $\Delta E^{elec}_{gas} + \Delta\Delta E^{elec}_{isom}$). This was also found by Novotny et al. (1997) for D1.3·HEL binding and by Jackson and Sternberg (1995) for trypsin·BPTI binding (see Tables 3 and 4). Xu et al. (1997) found that salt bridges across the binding interface can significantly stabilize complexes. In sharp contrast, Polticelli et al. (1999) found the opposite for trypsin·BPTI binding: the solvation term was found to favor binding ($\Delta\Delta G^{elec}_{solv} = -101.9$ kcal/mol), whereas the Coulomb energy was found to oppose binding ($\Delta E^{elec}_{gas} = 149.7$ kcal/mol). They attributed the latter to Coulomb repulsion between a +8e trypsin and +6e BPTI, and the former to the more favorable solvation of the +14e complex compared to the free proteins. Because the aforementioned groups computed the solvation term using the same continuum dielectric method (PB), the observed discrepancy may be due to the higher net charge on trypsin of +8e used by Polticelli et al. (1999) (as opposed to +6e in this work, see Table 1), which probably resulted in significant conformational changes, as evidenced by the much smaller trypsin·BPTI interface area of 1231 Å^2 (see Table 4).

CONCLUSIONS

1. For proteins that form high-affinity complexes, relatively accurate absolute binding free energies can be obtained using Scheme 1 and the GB model to compute the electrostatic solvation free energy, provided accurate x-ray structures for the free proteins and respective complex are available. The latter is required, not only to account for conformational changes upon binding, but because ΔE^{vdw}, ΔG^{elec} and ΔS^{vib} are sensitive to the 3-dimensional structure (see Tables 3 and 4).
2. There are three major advantages in using Scheme 1. First, each of the free energy components in Eq. 1 corresponds to a well-defined physical process. Second, by separating the gas-phase protein–protein interactions from protein–solvent and solvent–solvent interactions, the electrostatic components do not depend on the internal dielectric constant of the protein, whose value is uncertain in computing the electrostatic components of the binding free energy directly in solution. Third, the translational, rotational, and vibrational entropy changes can be computed in a straightforward manner (using Eqs. 14 and 15) in the gas-phase, but not in solution.
3. In cases where an accurate complex crystal structure is known but not its unbound counterparts, approximating $[r]_{sln}$ by $[R^{min+cons}]_{sln}$ and $[I]_{sln}$ by $[L^{min+cons}]_{sln}$ can yield trends in the relative contributions of the component terms to the binding free energy.
4. Neglecting vdW interactions from the free energy function can be a source of large errors because the gain in protein–protein vdW interaction energy at the interface may not be totally offset by the loss of protein-water vdW energy in the unbound state.
5. The following findings apply to protease—inhibitor and antibody–antigen complexes with an interface area in the range of 1400–1700 Å^2.
   a. Due to the observed $\Delta\Delta G^{nonel}_{solv}$ and $-T\Delta S^{conf}$ compensation, the magnitude of the binding affinity can be estimated from three components using Scheme 1 (see Eq. 17); viz., net gas-phase, protein–protein vdW effects (ΔE^{vdw}); net protein–protein and protein–solvent electrostatic effects (ΔG^{elec}); and the sum of gas-phase translational, rotational, and vibrational entropy changes ($-T\Delta S^{trans+rot+vib}$).
   b. The forces driving antibody–antigen and protease–inhibitor noncovalent association stem from vdW and electrostatic (including salt bridges and hydrogen bonds) protein–protein interactions between complementary surfaces over a relatively large area (1400–1700 Å^2) (Pauling, 1974; Pauling and Pressman, 1945). In contrast, vdW and electrostatic protein–solvent interactions oppose antibody–antigen and protease–inhibitor binding. These two opposite trends may be a general feature of highly charged complexes with an interface area in the range of 1400–1700 Å^2.
   c. As a consequence of b., the affinity of an antigen (inhibitor) for an antibody (protease) may be improved by increasing the close packing and hydrogen bonding (as opposed to charge–charge) interactions between the two proteins, and reducing the electrostatic cost of desolvating the free proteins (e.g., using water molecules to mediate charge–charge interactions at the protein interface) (Bhat et al., 1994).

We are grateful to Professor M. Karplus for the CHARMM program. S.N. is supported by a postdoctoral fellowship from Academia Sinica.

This work is supported by the Institute of Biomedical Sciences at Academia Sinica, the National Center for High Performance Computing, and the National Science Council, Republic of China.

REFERENCES

Andrews, P. R., D. J. Craik, and J. L. Martin. 1984. Functional group contributions to drug–receptor interactions. *J. Med. Chem.* 27: 1648–1657.

Babu, S., and C. Lim. 1999. Theory of ionic hydration: insights from molecular dynamics simulations and experiment. *J. Phys. Chem. B.* 103: 7958–7968.

Babu, S., and C. Lim. 2001. Incorporating nonlinear solvent response in continuum dielectric models using a two-sphere description of the born radius. *J. Phys. Chem. A.* 105:5030–5036.

Bernstein, F. C., T. F. Koetzle, G. J. B. Williams, E. F. Meyer, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and M. Tasumi. 1977. The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* 122:535–542.

Bhat, N. T., G. A. Bentley, G. Boulot, M. I. Greene, D. Tello, W. Dall'acqua, H. Souchon, F. P. Schwarz, R. A. Mariuzza, R. J. and Poljak. 1994. Bound water molecules and conformational stabilization help mediate an antigen–antibody association. *Proc. Natl. Acad. Sci. U.S.A.* 91:1089–1093.

Bogan, A. A., and K. S. Thorn. 1998. Anatomy of hot spots in protein interfaces. *J. Mol. Biol.* 280:1–9.

Bohm, H. J. 1994. The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput. Aided Mol. Des.* 8:243–256.

Born, M. 1920. *Z. Phys.* 1:45–48.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMm: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187–217.

Chothia, C., and J. Janin. 1975. Principles of protein–protein recognition. *Nature.* 256:705–708.

Dominy, B. N., and C. L. Brooks, III. 1999. Development of a generalized Born model parametrization for proteins and nucleic acids. *J. Phys. Chem. B.* 103:3765–3773.

Friedman, R. A., and B. Honig. 1995. A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.* 69:1528–1535.

Frisch, C., G. Schreiber, J. C. M., and, F. A. R. 1997. Thermodynamics of the interaction of barnase and barstar: changes in the free energy vs. changes in enthalpy on mutation. *J. Mol. Biol.* 267:696–706.

Froloff, N., A. Windemuth, and B. Honig. 1997. On the calculation of binding free energies using continuum methods: application to MHC class I protein–peptide interactions. *Prot. Sci.* 6:1293–1301.

Gibas, C. J. 1997. pH dependence of antibody/lysozyme complexation. *Biochemistry.* 36:15599–15614.

Gibrat, J.-F., and N. Go. 1990. Normal mode analysis of human lysozyme: study of the relative motion of the two domains and characterization of the harmonic motion. *Proteins: Struct. Func. Genet.* 8:258–279.

Gilson, M. K., J. A. Given, B. A. Bush, and J. A. McCammon. 1997. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* 72:1047–1069.

Gilson, M. K., and B. Honig. 1988. Calculation of the total electrostatic energy of macromolecular system: solvation energy, binding energies and conformational analysis. *Proteins: Struct. Func. Genet.* 4:7–18.

Hendsch, Z., and B. Tidor. 1999. Electrostatic interaction in the GCN4 leucine zipper: substantial contribution arise from intramolecular interaction enhanced on binding. *Protein Sci.* 8:1381–1392.

Honig, B., and A. Nicholls. 1995. Classical electrostatics in biology and chemistry. *Science.* 268:1144–1149.

Honig, B., K. Sharp, and A. Yang. 1993. Macroscopic models of aqueous solutions: biological and chemical application. *J. Phys. Chem.* 97: 1101–1109.

Horton, N., and M. Lewis. 1992. Calculation of free energy of association for protein complexes. *Protein Sci.* 1:169–181.

Ishida, H., Y. Jochi, and A. Kidera. 1998. Dynamic structure of subtilisin-eglin c complex studied by normal mode analysis. *Proteins: Struct. Func. Genet.* 32:324–333.

Jackson, R. M., and M. J. E. Sternberg. 1995. A continuum model for protein–protein interactions: application to the docking problem. *J. Mol. Biol.* 250:258–275.

Janezic, D., and B. Brooks. 1995. Harmonic analysis of large systems. II. Comparison of different protein models. *J. Comp. Chem.* 16:1543–1553.

Jayaram, B., K. J. McConnell, B. D. Surjit, and D. L. Beveridge. 1999. Free energy analysis of protein-DNA binding: The *Eco*RI Endonuclease-DNA complex. *J. Comp. Phys.* 151, 333–357.

Jorgensen, W. L., and J. Tirado-Rives. 1995. Free energies of hydration for organic molecules from Monte Carlo simulations. *In* Perspectives in Drug Discovery and Design, Vol. 3, K. Muller (editor). ESCOM Science Publishers, Leiden, The Netherlands. 123–138.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv. Prot. Chem.* 14:1–63.

Kollman, P. A. 1996. Advances and continuing challenges in achieving realistic and predictive simulations of the properties of organic and biological molecules. *Acc. Chem. Res.* 29:461–469.

Kollman, P. A., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. A. Case, and T. E. Cheatham, III. 2000. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33:889–897.

Krystek, K., T. Stouch, and J. Novotny. 1993. Affinity and specificity of serine endopeptidase–protein inhibitor interactions: empirical free energy calculations based on x-ray crystallographic structures. *J. Mol. Biol.* 234:661–679.

Lee, B., and F. M. Richards. 1971. The interpretation protein structures: estimation of static accessibility. *J. Mol. Biol.* 55:379–400.

Luzzati, V. 1952. Traitement statistique des erreurs dans la determination des structures cristallines. *Acta Cryst.* 5:802–810.

MacKerell, J. A. D., D. Bashford, M. Bellott, R. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. I. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus, M. 1998. All-hydrogen empirical potential for molecular modeling and dynamics studies of proteins using the CHARMM22 force field. *J. Phys. Chem. B.* 102:3586–3616.

McCammon, J. A. 1998. Theory of biomolecular recognition. *Curr. Opin. Struct. Biol.* 8:245–249.

McQuarrie, D. A. 1976. *Statistical Mechanics*, Harper and Row, New York.

Misra, V. K., J. L. Hecht, A.-S. Yang, and B. Honig. 1998. Electrostatic contributions to the binding free energy of the λcl repressor to DNA. *Biophys. J.* 75:2262–2273.

Miyamoto, S., and P. A. Kollman. 1993. Absolute and relative binding free energy calculations of the interaction of biotin and its analogs with streptavidin using molecular dynamics/free energy perturbations approaches. *Proteins: Struct. Funct. Genet.* 16:226–245.

Nicholls, A., K. A. Sharp, and B. Honig. 1991. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Func. Genet.* 11:281–296.

Novotny, J., R. Bruccoleri, M. Davis, and K. A. Sharp. 1997. Empirical free energy calculations: a blind test and further improvements to the method. *J. Mol. Biol.* 268:401–414.

Novotny, J., R. Bruccoleri, and F. A. Saul. 1989. On the attribution of binding energy in antigen–antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry.* 28:4735–4746.

Ohtaki, H., and N. Fukushima. 1992. A structural study of saturated aqueous solutions of some alkali halides by x-ray diffraction. *J. Sol. Chem.* 21:23–25.

Olson, M. A. 1999. Mean-field analysis of protein–protein interaction. *Biophys. Chem.* 75:115–128.

Page, M. I., and W. P. Jencks. 1971. Entropic contributions to rate accelerations in enzymic and intramolecular reactions and the chelate effect. *Proc. Natl. Acad. Sci. U.S.A.* 68:1678–1683.

Pascual-Ahuir, J. L., and E. Silla. 1990. GEPOL: an improved description of molecular surfaces. I. Building the spherical surface set. *J. Comp. Chem.* 11:1047–1060.

Pauling, L. 1974. Molecular basis of biological specificity. *Nature.* 248: 769–771.

Pauling, L., and D. Pressman. 1945. *J. Am. Chem. Soc.* 67:1003–1012.

Pearce, K. H., M. H. Ultsch, K. R. H., V. A. d. M., and J. A. Wells. 1996. Structural and mutational analysis of affinity-inert contact residues at the growth hormone–receptor interface. *Biochemistry.* 35:10300–10307.

Philippopoulos, M., and C. Lim. 1995. Molecular dynamics simulation of *E. Coli* ribonuclease H1 in solution: correlation with NMR and x-ray data and insights into biological function. *J. Mol. Biol.* 254:771–792.

Pickett, S. D., and M. J. Sternberg. 1994. Protein side-chain conformational entropy derived from fusion data—comparison with other empirical scales. *Protein Eng.* 7:149–155.

Pickett, S. D., and M. J. E. Sternberg. 1993. Empirical scale of side-chain conformational entropy in protein folding. *J. Mol. Biol.* 231:825–839.

Polticelli, F., P. Ascenzi, M. Bolognesi, and B. Honig. 1999. Structural determinants of trypsin affinity and specificity for cationic inhibitors. *Prot. Sci.* 8:2621–2629.

Qiu, D., P. S. Shenkin, F. P. Hollinger, and W. C. Still. 1997. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J. Phys. Chem.* 101:3005–3014.

Reddy, V. S., H. A. Giesing, R. Morton, A. Kumar, C. Post, C. L. I. Brooks, and J. E. Johnson. 1998. Energetics of quasiequivalence: computational analysis of protein–protein interactions in icosahedral viruses. *Biophys. J.* 74:546–558.

Sharp, K. A. 1998. Calculation of HyHel10-Lysozyme binding free energy changes: effect of ten point mutations. *Proteins: Struct. Func. Genet.* 33:39–48.

Sharp, K. A., A. Nicholls, R. F. Fine, and B. Honig. 1991. Reconciling the magnitude of the macroscopic and microscopic hydrophobic effects. *Science.* 252:106–109.

Shen, J. 1997. A theoretical investigation of tight-binding thermolysin inhibitors. *J. Med. Chem.* 40:2953–2958.

Sitkoff, D., K. A. Sharp, and B. Honig. 1994. Accurate calculation of hydration free energies using macroscopic solvent models. *J. Phys. Chem.* 98:1978–1988.

Still, W. C., A. Tempczyk, R. C. Hawley, and T. Hendrickson. 1990. Semianalytical treatment of solvation for molecular mechanics and dynamics. *J. Am. Chem. Soc.* 112:6127–6129.

Tanokura, M. 1983. ^{1}H-NMR study on the tautomerism of the imidazole ring of histidine residues. I. Microscopic pK values and molar ratios of tautomers in histidine-containing epptides. *Biochim. Biophys. Acta.* 742: 576–585.

Tembe, B. L., and J. A. McCammon. 1984. Ligand–receptor interactions. *Comput. Chem.* 8:281–283.

Tidor, B., and M. Karplus. 1994. The contribution of vibrational entropy to molecular association. The dimerization of insulin. *J. Mol. Biol.* 238: 405–414.

Vajda, S., Z. Weng, R. Rosenfeld, and C. DeLisi. 1994. Effect of conformational flexibility and solvation on receptor–ligand binding free energies. *Biochemistry.* 33:13977–13988.

Verhoeyen, M., C. Milstein, and G. Winter. 1988. Reshaping human antibodies: grafting an antilysozyme activity. *Science.* 239:1534–1535.

Vincent J. P., and M. Lazdunski. 1972. Trypsin-pancreatic trypsin inhibitor association. Dynamics of the interaction and role of disulfide bridges. *Biochemistry.* 11:2967–2979.

Weng, Z., C. Delisi, and S. Vajda. 1997. Empirical free energy calculation: comparison to calorimetric data. *Protein Sci.* 6:1976–1984.

Williams, D. H., J. P. Cox, A. J. Doig, M. Gardner, I. A. Nicholls, C. J. Salter, and R. C. Mitchell. 1991. Toward the semiquantitative estimation of binding constants: guides for peptide–peptide binding in aqueous solution. *J. Am. Chem. Soc.* 113:7020–7030.

Williams, J. P., M. S. Searle, J. P. Mackay, U. Gerhard, and R. A. Maplestone. 1993. Toward an estimation of binding constants in aqueous solution. *Proc. Natl. Acad. Sci. U.S.A.* 90:1172–1178.

Wilson, C., J. E. Mace, and D. A. Agard. 1991. Computational method for the design of enzymes with altered substrate specificity. *J. Mol. Biol.* 220:495–506.

Xavier, K. A., and R. C. Wilson. 1998. Association and dissociation kinetics of anti-hen egg lysozyme monoclonal antibodies HyHEL-5 and HyHEL-10. *Biophys. J.* 74:2036–2045.

Xu, D., S. L. Lin, and R. Nussinov. 1997. Protein binding vs. protein folding: the role of hydrophilic bridges in protein associations. *J. Mol. Biol.* 265:68–84.

Zou, X., Y. Sun, and I. D. Kuntz. 1999. Inclusion of solvation in ligand binding free energy calculations using the generalized-Born model. *J. Am. Chem. Soc.* 121:8033–8043.