Free Energy Landscapes of Encounter Complexes in Protein-Protein Association

Carlos J. Camacho, Zhiping Weng, Sandor Vajda, and Charles DeLisi
Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215 USA

ABSTRACT We report the computer generation of a high-density map of the thermodynamic properties of the *diffusion-accessible* encounter conformations of four receptor-ligand protein pairs, and use it to study the electrostatic and desolvation components of the free energy of association. Encounter complex conformations are generated by sampling the translational/rotational space of the ligand around the receptor, both at 5-Å and zero surface-to-surface separations. We find that partial desolvation is always an important effect, and it becomes dominant for complexes in which one of the reactants is neutral or weakly charged. The interaction provides a slowly varying attractive force over a small but significant region of the molecular surface. In complexes with no strong charge complementarity this region surrounds the binding site, and the orientation of the ligand in the encounter conformation with the lowest desolvation free energy is similar to the one observed in the fully formed complex. Complexes with strong opposite charges exhibit two types of behavior. In the first group, represented by barnase/barstar, electrostatics exerts strong orientational steering toward the binding site, and desolvation provides some added adhesion within the *local* region of low electrostatic energy. In the second group, represented by the complex of kallikrein and pancreatic trypsin inhibitor, the overall stability results from the rather nonspecific electrostatic attraction, whereas the affinity toward the binding region is determined by desolvation interactions.

INTRODUCTION

Most macromolecular processes require rapid and highly specific macromolecular association, and their rates are limited by the rate at which diffusion can bring the reactants together. The maximum rate constant is given by the Smoluchowski equation, $k_{coll} = 4\pi Da$, where D is the relative translational diffusion coefficient and a is the sum of the atomic radii of the two molecules (Noyes, 1961; DeLisi, 1980). In the size range of proteins, k_{coll} is calculated to be 10^9–10^{10} M^{-1} s^{-1}. However, in most cases the rate of diffusion-limited macromolecular association is well below this value, and can be described by a modified Smoluchowski equation $k_{assoc} \simeq 4\pi Da\kappa f$, where κ is a dimensionless interaction parameter, and f is a dimensionless factor that reflects the increase or decrease in the diffusional collision rate due to electrostatic steering (von Hippel and Berg, 1989). The value of f does not exceed 10 even for very favorable interactions (Noyes, 1961). Furthermore, if κ accounts for the small fraction of the surfaces that are reactive (von Hippel and Berg, 1989; Janin, 1997), then for typical proteins one might expect k_{assoc} values not exceeding 10^4–10^5 M^{-1} s^{-1} (Schreiber and Fersht, 1996).

Association rate constants frequently exceed the modified Smoluchowski limit calculated for the given complex on the basis of the receptor/ligand geometry. A model that can explain this apparent contradiction and has been extensively discussed in the literature (DeLisi, 1980; Berg and von Hippel, 1985; von Hippel and Berg, 1989; Schreiber and Fersht, 1996) regards macromolecular association as a stepwise process in which translational diffusion brings the proteins to a "macrocollision" in an orientationally nonspecific or weakly specific fashion, forming an encounter complex (also referred to as a transition state, see Berg and von Hippel, 1985). The encounter complex can be thought of as an ensemble of conformations in which the molecules can rotationally diffuse along each other, or participate in a series of "microcollisions" that properly align the reactive groups. This model applies without assuming any attractive force between the proteins as Brownian motion predicts a certain diffusion entrapment, i.e., macromolecules in aqueous solution undergo several microcollisions before diffusing apart.

The association rates are further increased in the presence of nonspecific attractive interactions that held the two molecules together long enough to increase their chances of finding a mutually reactive configuration (Sommer et al., 1982; Berg and von Hippel, 1985; Schreiber and Fersht, 1996). Various interactions have been proposed (Sommer et al., 1982; Berg and von Hippel, 1985; Schreiber and Fersht, 1996) as being responsible for the increased stability of the encounter complex, including van der Waals, hydrogen bonding, electrostatic, and hydrophobic effects. However, van der Waals forces are unlikely to play a major role, since the interactions between the receptor and the ligand in the encounter complex are nearly compensated by the interactions between the reactants and the solvent in the free state (Berg and von Hippel, 1985). Similar compensation applies to hydrogen bonding, because the hydrogen bonds buried in the encounter complex involve polar groups that tend to form hydrogen bonds with the solvent in the free state (Dill, 1990). Thus, the most likely sources of nonspecific adhe-

Received for publication 30 June 1998 and in final form 30 October 1998.
Address reprint requests to Dr. Carlos J. Camacho, Dept. of Biomedical Engineering, Boston University, 44 Cummington St., Boston, MA 02215. Tel.: 617-353-4842; Fax: 617-353-6766; E-mail: ccamacho@bu.edu.

0006-3495/99/03/1166/13 $2.00

sion in the transition state are electrostatic and hydrophobic interactions.

Electrostatics has been unambiguously shown to substantially enhance the association rate in a number of systems. Reactions of this type include those of proteins with DNA (von Hippel and Berg, 1986), proteins with highly charged small molecules (Sharp et al., 1987), and proteins with oppositely charged protein substrates. In a particularly well-characterized protein-protein complex, barnase-barstar, the basal rate constant of $10^5\ M^{-1}\ s^{-1}$, observed at high ionic strength, is increased to over $5 \times 10^9\ M^{-1}\ s^{-1}$ by electrostatic forces (Schreiber and Fersht, 1996). However, in most cases it is not clear if the rate is increased by long-range and specific electrostatic steering or by nonspecific interactions that stabilize the transition state. Furthermore, the high rates observed in certain systems cannot be explained in terms of electrostatic interactions alone (Sommer et al., 1982).

Because the rapidly exchanging encounter conformations are not accessible to most experimental techniques, the nature of the transition state is not fully established. For example, Northrup and Erickson (1992) reject the concept of an encounter complex stabilized by nonspecific interactions on the grounds that proteins in solution do not exhibit strong enough nonspecific association, even at high concentrations. In their proposed model the partially formed complexes are stabilized by a specific, short-range potential, amounting to a fraction of the forces in the final complex. The physical origin of this locking potential, leading to amplified Brownian entrapment, was not discussed, and the model has been supported only by Brownian dynamic simulations of spherical proteins with reactive patches.

This paper presents direct calculations of the interactions that may contribute to the stability of the transition state in the reaction of protein-protein association. The analysis is based on determining the free energy landscape, as well as its electrostatic and desolvation components, over the configurational space of encounter complexes. Starting from receptor and ligand structures, encounter complex conformations are generated by systematically sampling subsets of the six-dimensional space, defined by translations and rotations of the ligand around the receptor. In the analysis of long-range electrostatic interactions this subset is defined by a fixed surface-to-surface distance, while short-range interactions are studied by restricting the surfaces to close proximity. The analysis is based on well-established free energy evaluation models (Vajda et al., 1994; Zhang et al., 1997a) in which the free energy of association is decomposed into electrostatics and desolvation terms, the latter also accounting for the loss of side chain conformational entropy. Although the two terms are calculated using simplified models, the free energy potential has been carefully tested for consistency with thermodynamic and structural data (Zhang et al., 1997b; Weng et al., 1997).

It is important to note that searching for diffusion-accessible encounter complexes or transition states is very different from docking, which attempts to find the conformation of the fully formed complex (Goodsell and Olson, 1990; Shoichet and Kuntz, 1991; Rosenfeld et al., 1995; Jackson and Sternberg, 1995). The low energy encounter conformations show some similarity to the final complex, but they have much smaller contact surface areas, and the two conformations can have up to 10-Å root mean square deviation (RMSD). Proceeding from the encounter complex to the bound state is a nontrivial computational problem that will not be discussed here.

The present work focuses on elucidating the interactions that may contribute to the stability of encounter complex, including the analysis of the relationship between long-range electrostatic steering and short-range adhesion. In particular, we show that partial desolvation is always an important effect, and it becomes dominant for complexes in which one of the reactants is neutral. Desolvation provides enough stability to rationalize a surface-on-surface, diffusion-mediated search for the combining sites. However, the interaction is not fully nonspecific, but provides a slowly varying attractive force over a small but significant region of the molecular surface; though, in average, the free energy is weakly repulsive. We also find that in complexes with weak electrostatic interactions, the final conformation of the complex is within the region of strong attractive desolvation. In contrast, when both reactants have a net charge, the binding sites are closely identified either by the minimum of the free energy of association (desolvation plus electrostatic energy), or by electrostatics alone. The latter, i.e., dominant electrostatic steering, is found for complexes such as barnase-barstar with strong long-range electrostatic interaction. However, this case appears to be far from typical, and weakly specific partial desolvation may play a major role in driving the molecules to the transition complexes, enhancing the association rates.

MATERIALS AND METHODS

Receptor-ligand complexes

Long-range electrostatic interaction maps will be constructed for the first seven protein pairs listed in Table 1. The desolvation and electrostatic components of the association free energy will be mapped over the configurational space of the encounter complexes for the first four pairs: α-chymotrypsin with turkey ovomucoid third domain (1CHO), human leukocyte elastase with turkey ovomucoid third domain (1PPF), kallikrein with pancreatic trypsin inhibitor (2KAI), and barnase with barstar (1BRS). As shown in Table 1, turkey ovomucoid third domain, also referred to as OMTKY, is neutral, and hence no strong long-range electrostatic interactions are expected in the first two complexes. By contrast, kallikrein and pancreatic trypsin inhibitor (PTI), as well as barnase and barstar are oppositely charged, resulting in strong electrostatic steering. In particular, barnase with barstar is a well-studied example of fast, electrostatically assisted, protein association (Schreiber and Fersht, 1996).

The complex structures are taken from the Protein Data Bank (PDB). To refine the structures, the receptor and the ligand are separately minimized over 200 steps using Version 19 of the CHARMM potential and assuming harmonic constraints on the atomic positions (Brooks et al., 1983).

TABLE 1 Comparison of desolvation energies for 12 protein complexes

PDB Code	Receptor	Charge e.u.	Ligand	Charge e.u.	ΔE_C kcal/mol	ΔG_{ACE} kcal/mol
1CHO	α-chymotrypsin	3	OMTKY	0	−14.03	−15.72
1PPF	Leuk. elastase	11	OMTKY	0	−19.07	−17.46
2KAI	Kallikrein A	−17	PTI	5	−9.71	−10.81
1BRS	Barnase	2	Barstar	−5	10.01	10.97
2SNI	Subtilisin novo	−1	CI-2	−1	−16.34	−14.59
1CSE	Subtilisin carl.	−1	Eglin-c	1	−11.45	−10.29
2PTC	Trypsin	6	BPTI	6	−5.47	−5.56
3SGB	Proteinase B	0	OMTKY	0	−12.27	−11.06
4SGB	Proteinase B	0	PCI-1	3	−19.95	−17.59
2SEC	Subtilisin carl.	−1	N-acetyl eglin-c	1	−10.95	−10.05
1TEC	Thermitase	0	Eglin-c	0	−13.33	−13.36
2TGP	Trypsinogen	3	BPTI	6	−6.39	−5.07

ΔE_C is computed using original ACE potential (see Table 4 in Zhang et al., 1997a) with a 6-Å cutoff, and ΔG_{ACE} is calculated using the modified potential described in the text, including a 9-Å cutoff which accounts for the partial desolvation of diffusion-accessible encounter pairs.

Free energy evaluation

We use the free energy potential of the form (Novotny et al., 1989; Vajda et al., 1994)

$$\Delta G = \Delta E_{coul} + \Delta G_{des} - T\Delta S_{sc} + \Delta G_{rot/trans}, \qquad (1)$$

where ΔE_{coul} is the electrostatic interaction energy between the receptor and the ligand, ΔG_{des} is the desolvation free energy, i.e., the free energy of transferring the buried atoms of the protein from the solvent into a protein environment, and ΔS_{sc} is the side-chain entropy loss. The last term, $\Delta G_{rot/trans}$, is the free energy change associated with the loss of six rotational-translational degrees of freedom, and is considered to be a constant for protein-protein complexes; i.e., it is a weak function of the size and shape (Novotny et al., 1989; Horton and Lewis, 1992; Nauchitel et al., 1995). We assume van der Waals (vdW) compensation, i.e., that the intermolecular vdW interactions are balanced by interactions with the solvent in the free state (Vajda et al., 1994).

Another frequently used decomposition of the free energy involves the full electrostatic free energy contribution which, in addition to the electrostatic interaction energy ΔE_{coul}, includes the self-energy change upon desolvating the charges of polar atoms (Honig and Nicholls, 1995). The electrostatic free energy is calculated by solving the Poisson-Boltzmann equation, and is defined by $\Delta G_{PB} = G^{rl}_{PB} - G^{r}_{PB} - G^{l}_{PB}$, where G^{rl}_{PB}, G^{r}_{PB}, and G^{l}_{PB} denote the electrostatic free energies of the intermediate, the receptor, and the ligand. In terms of ΔG_{PB} the binding free energy is given by

$$\Delta G = \Delta G_{PB} + \Delta G^{np}_{des} - T\Delta S_{sc} + \Delta G_{rot/trans}. \qquad (2)$$

In this equation ΔG^{np}_{des} denotes the nonpolar part of the desolvation free energy, usually calculated by $\Delta G^{np}_{des} = \lambda \Delta A$, where ΔA is the change in solvent-accessible surface area, and λ is a parameter derived from the desolvation free energy of nonpolar molecules (Honig and Nicholls, 1995). We note again that the desolvation contribution from polar groups is part of the desolvation term ΔG_{des} in Eq. 1, whereas in Eq. 2 this contribution is part of the electrostatic free energy ΔG_{PB}.

Desolvation free energy

We calculate the sum $\Delta G_{des} - T\Delta S_{sc}$ in Eq. 1 as a free energy term ΔG_{ACE} based on the atomic contact energy (ACE) developed by Zhang et al. (1997a). The ACE function is an extension of the residue-residue quasichemical potential of Miyazawa and Jernigan (1985). Although the atomic contact energies were estimated by a statistical analysis of atom-pairing frequencies in a set of high resolution protein structures, it has been shown that the function can be used to calculate solvation and entropic contributions to the binding free energy in intermolecular applications. In particular, for nine protease-inhibitor complexes the calculated binding free energies were typically within 10% of the experimentally measured values (Zhang et al., 1997a). The function was also used to calculate the free energy changes associated with the binding of peptides to a major hysto-compatibility complex molecule, and the deviations from experimental data were within 1 kcal/mol. In addition, it was shown that peptide recognition and protein folding can be treated by the same ACE potential (Zhang et al., 1997b).

Zhang et al. (1997a) restricted consideration to atoms within 6 Å when constructing the ACE function. For the analysis of encounter complexes we increase the distance cutoff to 9 Å. The reason is that desolvation interactions are not restricted to the atomic surface, but extend to a few water layers (Israelachvili and Wennerstrom, 1996). The minimum distance between two atomic centers on separately solvated proteins is ~9 Å, i.e., two atomic radii plus the diameter of two water molecules. Indeed, a distance smaller than 9 Å can be attained only when at least one of the protein surfaces is desolvated. Summing up the interactions up to a 9-Å cutoff rather than to 6 Å increases the magnitude of the free energy, but this increase can be removed by an appropriate normalization factor of 0.33 (see Eq. 2b in Zhang et al., 1997a). The new cutoff of 9 Å is actually the midpoint of a switching function which smoothly decays to zero between 8 Å and 10 Å. Results are not significantly affected by shifting the midpoint between 8.5 Å and 10.5 Å. Furthermore, to dismiss contacts between atoms in the interior, and at the same time to make the potential less sensitive to surface fluctuations, *one* of the atoms in each pair is required to have a solvent contact area in the free state of 1 $Å^2$ or more. All surface areas are computed using the method of Lee and Richards (1971), with a water radius of 1.4 Å. As we will show, the modified contact potential and the original version due to Zhang et al. (1997a) provide very similar desolvation free energy profiles.

Electrostatic energy

A commonly used method to estimate electrostatic interactions is solving the linear Poisson-Boltzmann (PB) equation. In combination with finite difference or boundary element methods that allow for incorporation of detailed atomic level structural information, this model has been applied to virtually every problem in macromolecular electrostatics (Sharp and Honig, 1990). We use a finite difference (FD) method as implemented in CONGEN, which features adjustable rectangular grids, a uniform charging scheme that decreases the unfavorable grid energies, and smoothing algorithms that alleviate problems associated with discretization (Bruccoleri et al., 1997). The calculations were carried out using a 0.8-Å grid, with uniform charging, antialiasing, and 15-point harmonic smoothing. An 8-Å grid margin was maintained around the molecules. The dielectric constant of the solvent was set to 78, and unless otherwise specified, that of the protein to 2. The ionic strength was set to 0.05 M.

In this paper we employ PB calculations in two different ways. In the analysis of long-range electrostatic interactions we consider a restricted set of intermediate receptor-ligand configurations in which the shortest distance between any atom of the receptor and any atom of the ligand is 5 Å. The electrostatic free energy change ΔG_{PB} is calculated using the FDPB algorithm. We recall that ΔG_{PB} includes both the direct (Coulombic) receptor-ligand interaction energy at the 5 Å separation, and a smaller effect due to the long-range desolvation of polar groups.

The analysis of short-range interactions involves the mapping of the binding free energy and its electrostatic component over the entire configurational space of the encounter complex. As we will describe, this is equivalent to calculating the thermodynamic quantities at the points of a five-dimensional grid. The large number of required free energy evaluations prevents the use of the full Poisson-Boltzmann calculation just described. In fact, solving the linear PB equation for a single receptor-ligand encounter pair requires ~20 min of CPU on a R10000 computer, and constructing the map of electrostatic energy for the entire set of encounter conformations would take several years for a single complex. Moreover, the PB electrostatic binding energy ΔG_{PB} includes the desolvation of polar groups already accounted for by the desolvation potential ΔG_{ACE}.

To avoid double-counting the desolvation term we calculate the direct (Coulombic) component of the electrostatic energy defined by $\Delta E_{coul} = \sum_{i=1}^{n} \Phi_i q_i$, where q_i is the charge of the atom i of the ligand, Φ_i is the electrostatic potential of the receptor at the position of the same atom, and n denotes the number of ligand atoms. The expression is exact if the potential Φ is calculated by solving the Poisson-Boltzmann equation for the encounter complex in which only the receptor is charged, i.e., the ligand is replaced by a low dielectric ($\epsilon = 2$) cavity. However, the potential must be recalculated following each move of the ligand. To render the mapping computationally feasible we used a semi-Coulombic approximation in which the potential is calculated for the isolated receptor, i.e., the effect of the low dielectric cavity of the ligand on the potential of the receptor is neglected (Northrup et al., 1984).

Gabdoulline and Wade (1996) have optimized the semi-Coulombic approach by fitting effective ligand charges in the presence of the receptor's field against ΔG_{PB}. This procedure reduces the effect of using an approximate potential by increasing the atomic charges of the ligand by factors between 1 and 2, and was shown to provide good approximations of the electrostatic forces unless the separation of protein surfaces was less than twice the water radius. In the analysis of encounter complexes we use a similar approximation that seems to be reasonably accurate for the purposes of the present paper. The charges and partial charges of the ligand are not changed, but the solvent dielectric constant is set to 40, effectively multiplying all charges by a factor of 2. The effective solvent dielectric of 40 accounts for the long-range PB electrostatic energy, smoothly extrapolating this energy to within the partially desolvated interface.

Sampling

As described in the previous section, long-range and short-range interactions are studied separately. First we perform a limited mapping of the electrostatic energy surface at 5 Å separation in order to determine whether long-range electrostatic steering is important for the given complex. The receptor's center of mass is placed at the origin of a coordinate system such that the binding site faces toward the positive x axis. The ligand is then assumed orbiting around the corresponding receptor in the xz plane such that the P1 residue of the ligand always points toward the center of mass of the receptor. The orbits are not circular, and the distance to the ligand's center is set by the constraint that the separation between proteins be 5 Å, defined as the shortest distance between any atom of the receptor and any atom of the ligand. Orienting the ligand toward the receptor eliminates three degrees of freedom. Since the receptor-ligand distance is fixed and the search is restricted to the xz plane, the mapping is along a single coordinate.

By contrast, for the analysis of short-range interactions we perform a search along five degrees of freedom. As in the previous case, the receptor's center of mass is placed at the origin of a coordinate system. The ligand's center of mass in these coordinates is specified by two Euler angles (θ, ϕ), and the minimum distance between the surfaces of the two molecules. Since we are interested in sampling diffusion-accessible states, this distance is set to zero. In practical terms we require that the two molecules do not overlap and that the minimum surface-to-surface atom distance (defined in terms of the vdW radii) be at most 0.25 Å. Three more degrees of freedom (i.e., three Euler angles) are used to describe the orientation of the ligand around its own center of mass. This five-dimensional space is systematically sampled on the grid defined by $\theta = 0°, 20°, \ldots, 180°$ and $\phi = 0°, 20°, \ldots, 360°$. For each grid point we calculate the desolvation, electrostatic, and total free energy. Moreover, at each position of the ligand center of mass on the receptor's grid defined by (θ_i, ϕ_j) we find the ligand orientation that minimizes the particular energy function. Once the minimum is found for a particular target function, we further search for the local minimum using a finer grid within $\pm 5°$ on the receptor surface, and $\pm 5°$ and $\pm 10°$ in the space of ligand's orientations. Notice that for a single ligand-receptor pair the sampling in the five-dimensional conformational space just described requires a total of 272,874 evaluations of the energy functions.

For visualization the results of the exhaustive thermodynamic evaluation are reduced to two-dimensional free energy (or electrostatic or solvation energy) landscapes, i.e., for each (θ, ϕ) pair we retain only the minimum function value found when searching the rotational space of the ligand.

RESULTS

Validation of free energy estimate

While the free energy evaluation model given by Eq. 1 has been extensively used in a variety of applications (Zhang et al., 1997a, b), and its terms have been carefully tested for consistency with thermodynamic data, the function we use here is slightly different and hence requires further validation. First, to account for early desolvation phenomena when two solvated proteins approach each other, the range of the interaction has been increased to an atom-to-atom distance of 9 Å in the atomic contact energy (ACE). Second, in the past we have used the simple Coulombic formula with the distance-dependent dielectric of $\epsilon = 4r$ in electrostatics calculations; here this is replaced by an FDPB electrostatic energy calculation in the analysis of long-range interactions, and by the semi-Coulombic approximation of the PB electrostatic energy in the analysis of short-range phenomena. The effects of these modifications will be discussed in turn.

As we already mentioned, the effect of the longer range of the ACE potential is compensated by an extra normalization factor of 0.33 in the contact potential. Indeed, the resulting potential agrees with the original ACE, introduced by Zhang et al. (1997a). This is clearly shown in Table 1, where we compare binding free energy data for several complexes, including those in Table 4 of Zhang et al. (1997a). The good agreement suggests that the extension of range and the change in normalization alter the atomic contact energy term by <1 kcal/mol.

The advantages of FDPB calculations over the use of a distance-dependent dielectric has been extensively discussed in the literature (Honig and Nicholls, 1995) and hence we examine only the validity of the quasi-Coulombic approximation used for the calculation of the direct electro-

static component ΔE_{coul} (see Methods). Fig. 1 compares this term with the full Poisson-Boltzmann electrostatic energy ΔG_{PB} (*solid lines*) as a function of the minimum distance between the molecular surfaces of barnase and barstar, d_{s-s}, along two arbitrarily chosen trajectories. In (*A*) the molecules are oriented as in their complex structure, and in (*B*) as in the encounter complex conformation on which the minimum of the direct electrostatic energy is attained (see the description of the mapping procedure in Methods). The receptor and the ligand are placed in the conformations just described at the surface-to-surface distance of 14 Å, and then translated along the path joining their centers. As shown in Fig. 1, for molecular separations larger than two water layers the difference between ΔE_{coul} and ΔG_{PB} is <0.25 kcal/mol. Thus, not only the error due to the quasi-Coulombic approximation is negligible, but the entire contribution due to the desolvation of polar atoms, included in ΔG_{PB} but not in ΔE_{coul}, is small. Consistent with the extended range of interaction imposed in our desolvation energy, ΔG_{PB} starts to account for desolvation effects when molecules are less than two water layers apart (see the change in curvature in ΔG_{PB}). As expected, ΔE_{coul} does not show the same change in curvature, and the two quantities substantially differ.

It is important to stress that our method was developed to calculate the electrostatic energy of encounter conformations rather than that of the fully formed complex. In the latter, the use of the effective dielectric constant of $\epsilon = 40$ for the solvent in the region occupied by the ligand would not have physical motivation, since the snugly fit protein-protein complex essentially forms a single low dielectric cavity ($\epsilon \simeq 2$). However, in typical encounter conformations there is a substantial amount of water between the two proteins even at zero surface-to-surface separation. For example, along path (*A*) the molecular surfaces are in contact (i.e., $d_{s-s} = 0$) when the centers of the two molecules are 31.2 Å apart, whereas in the crystal structure the centers are only 23.6 Å apart.

Further validation of the desolvation and direct electrostatic energy is obtained by comparing two different estimates of the association free energy. As we mentioned, in terms of ΔE_{coul} the free energy of association is given by Eq. 1, whereas with ΔG_{PB} the appropriate decomposition is given by Eq. 2. Since the term $\Delta G_{rot/trans}$ is present in both decompositions of the association free energy, in Fig. 1 we compare the expressions $\Delta E_{coul} + \Delta G_{ACE}$ and $\Delta G_{PB} + \lambda \Delta A - T\Delta S_{sc}$. A reasonable estimate of $\lambda = 38$ cal/mol/Å^2 yields an excellent agreement between the two free energies. It is noteworthy that with only one free parameter, the two curves agree within 1 kcal/mol over the whole range of molecular distances.

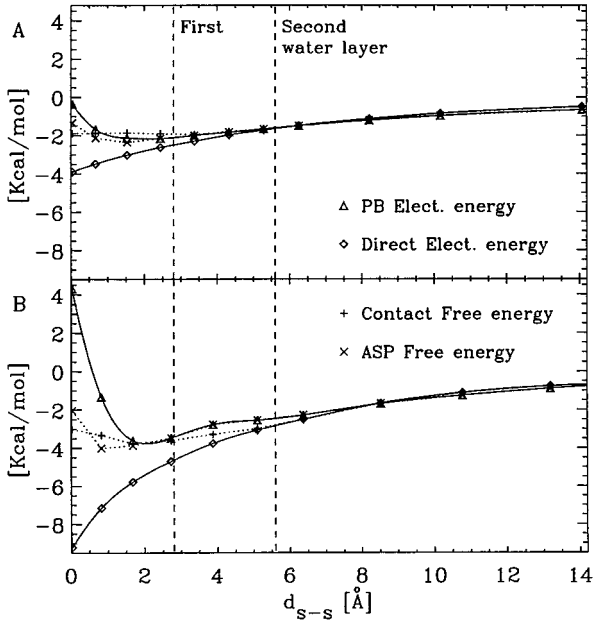

FIGURE 1 Comparisons of Poisson-Boltzmann electrostatic energy ΔG_{PB} and direct electrostatic estimate ΔE_{coul} (*solid lines*). Dotted lines correspond to the association free energy ΔG as a function of the minimum distance between the molecular surfaces of barnase and barstar d_{s-s}, along two arbitrarily chosen trajectories. ΔG is computed using both the contact potential in Eq. 1 and the atomic solvation parameter (ASP) in Eq. 2 (without $\Delta G_{rot/trans}$). For (*A*) the molecules are oriented as in their complex structure, and then translated along the path joining their centers. (*B*) The molecules are oriented as in the minimum direct electrostatic conformation in Fig. 6 *D*, and then translated along the path joining their centers.

Electrostatic steering

The first stage in the binding process corresponds to the diffusion of proteins into close proximity. Experimental evidence has shown that the association rate constant of some oppositely charged complexes such as barnase and barstar (Schreiber and Fersht, 1996), hirudin and thrombin (Stones et al., 1989), and colicin E9 and its cognate immunity protein Im9 (Wallis et al., 1995), are greatly enhanced by electrostatics. However, a simple mapping of the electrostatic interactions ΔG_{PB} between some receptors and their ligands at 5 Å surface-to-surface distance shows that the role of electrostatic forces seen in these systems cannot be extended to all protein-protein complexes. Fig. 2 shows the contour plots of the Poisson-Boltzmann electrostatic energy, ΔG_{PB}, in the *xz* plane for the first seven protein-protein pairs listed in Table 1 as each ligand orbits around its corresponding receptor. As described in the Methods, for each complex the receptor's center of mass is at the origin, the receptor is fixed, and its binding site is facing the positive *x* axis. As indicated by the C-α traces also shown in Fig. 2, the P1 residue of the ligand always points toward the center of mass of the receptor. Thus, the receptor-ligand pairs are perfectly aligned along the positive *x* axis. The ligand's orbits are not circular, but are selected to maintain the constant 5-Å separation. The dotted lines in Fig. 2 correspond to constant energies of -1 kcal/mol (closer to the center), 0 kcal/mol, and 1 kcal/mol (away from the center). The solid line is the spline interpolation of the calculated ΔG_{PB} values (*open circles*).

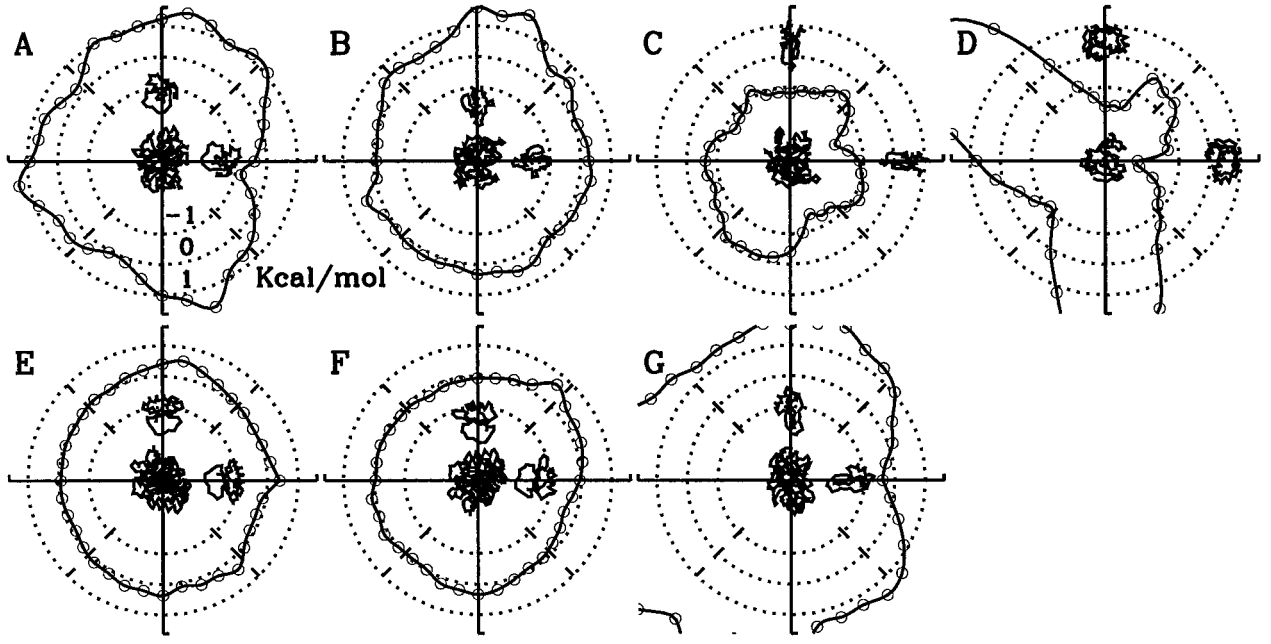

FIGURE 2 Contour plots of the Poisson-Boltzmann electrostatic energy of six ligands orbiting their corresponding receptors on the *xz* plane. Unless otherwise indicated the ionic strength is set to 0.05 M, and the dielectrics are 2 for the protein interior and 78 for the solvent. (*A*) α-Chymotrypsin with turkey ovomucoid third domain (1CHO); (*B*) human leukocyte elastase with turkey ovomucoid third domain (1PPF), ionic strength 0.15 M and protein dielectric 4; (*C*) kallikrein A and pancreatic trypsin inhibitor (2KAI), ionic strength 0.15 M and protein dielectric 4; (*D*) barnase and barstar (1BGS); (*E*) subtilisin and chymotrypsin inhibitor (2SNI); (*F*) subtilisin and eglin-c (1CSE), ionic strength 0.15 M and protein dielectric 4; (*G*) trypsin and bovine pancreatic trypsin inhibitor (2PTC).

The rather large electrostatic attraction of barnase and barstar (Fig. 2 *D*) at optimal orientation contrast sharply with the much weaker electrostatic complementarity of α-chymotrypsin with OMTKY (Fig. 2 *A*), human leukocyte elastase with OMTKY (Fig. 2 *B*), subtilisin and chymotrypsin inhibitor (Fig. 2 *E*), and subtilisin with eglin-c (Fig. 2 *F*). Notice that for some of these complexes the electrostatic interactions are even slightly repulsive at the optimal orientation. In kallikrein A with pancreatic trypsin inhibitor (PTI) the reactants are oppositely charged, and the complex exhibits an overall attraction (Fig. 2 *C*). However, unlike in barnase/barstar, the attraction is virtually independent of the orientation. An interesting behavior is seen for trypsin and PTI (Fig. 2 *G*) in which the two proteins show a small attraction at the binding site despite their like charges. Based on these results, the complexes in Table 1 exhibit three different types of behavior. Barnase and barstar is the classical case of strong and specific electrostatic attraction that is likely to orient the encounter complex toward the bound conformation; kallikrein A with pancreatic trypsin inhibitor that are oppositely charged resulting in overall attraction but no specific steering toward the binding site; and finally the remaining five complexes in which the reactants either have like charges, or one of them is neutral, resulting in a weak or nonexistent electrostatic term. We first describe the analysis of short-range interactions for two complexes from this group, α-chymotrypsin and human leukocyte elastase, both forming complexes with OMTKY.

Encounter complexes without strong electrostatic interactions

The maps in Fig. 3 describe the thermodynamics of diffusion-accessible encounter complexes of α-chymotrypsin with OMTKY. Fig. 3, *A–C* show the desolvation term ΔG_{ACE}, the total free energy of association $\Delta E_{\mathrm{coul}} + \Delta G_{\mathrm{ACE}}$ (without the constant $\Delta G_{\mathrm{rot/trans}}$), and the electrostatic energy ΔE_{coul}, respectively. The small square symbol on all maps indicates the crystal structure position of the center of the ligand on the receptor's surface, in this case located at parallel $\theta = 90°$ and meridian $\phi = 90°$. Dotted lines are drawn along parallels and meridians sparsed every 30° and 45°, respectively.

Fig. 3 *D* shows the C-α trace of the receptor (*black*), and the C-α traces of the ligand corresponding to the encounter complexes with minimum desolvation or electrostatic energies (*blue*). To avoid an overlap of the two latter traces, the one with the lowest electrostatic energy was shifted to the right. The position of the P1 site on the ligand is indicated by a red circle. For this complex the encounter conformations with minimum desolvation free energy and with minimum free energy coincide, and hence the latter is not shown.

The same maps and traces are shown in Fig. 4 for the encounter conformations of human leukocyte elastase with OMTKY. The only difference is that in this case the conformations with minimum desolvation free energy and with

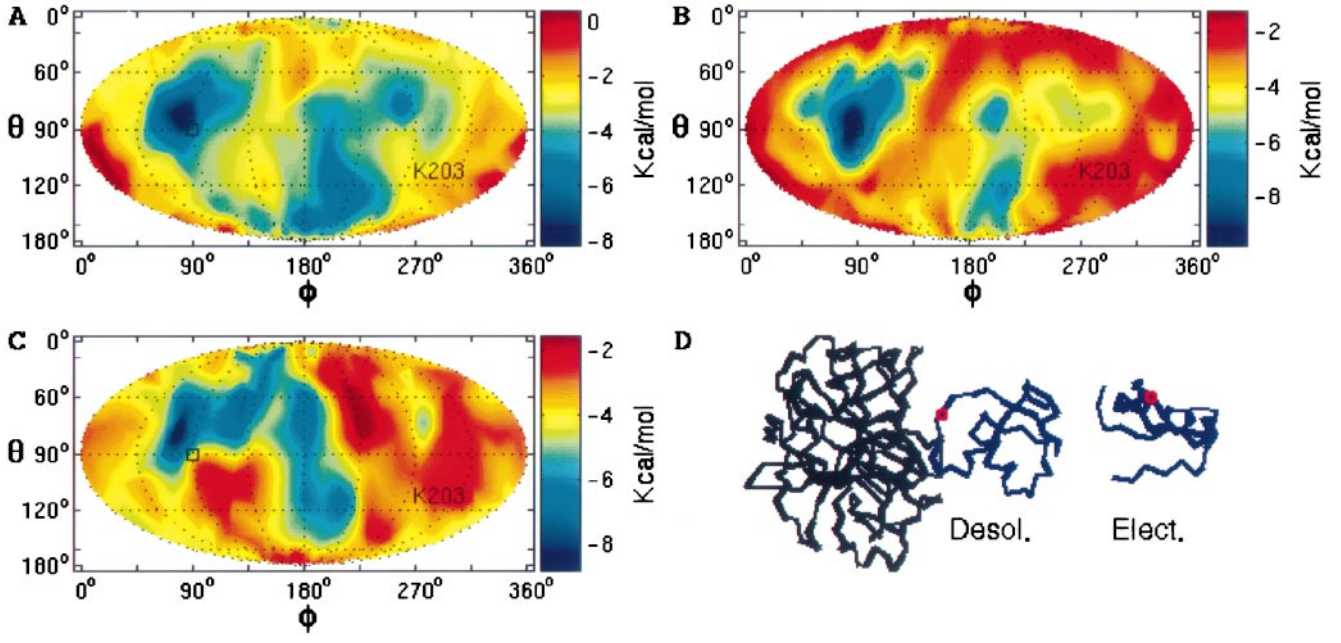

FIGURE 3 Lowest values of interactions in the diffusion-accessible encounter complexes of α-chymotrypsin with turkey ovomucoid third domain (OMTK). (*A*) desolvation energy; (*B*) total free energy of association; (*C*) electrostatic energy. On all maps, the center of the ligand in the x-ray structure is at position (90°, 90°), and the position of a receptor's residue is indicated. (*D*) C-α trace of the receptor (*black*) and traces of the ligand conformations in encounter complexes on which the minimum is reached in (*A*) and (*C*). To assist visualization, the latter trace was shifted to the right. The native complex conformation is not shown, but in *D* it faces the receptor along the positive *x* axis.

minimum free energy slightly differ, and their traces are shown separately. According to Figs. 3 *A* and 4 *A*, for each of the two complexes the minimum of the desolvation free energy is at the grid point closest to the receptor's binding site. Moreover, as shown in Figs. 3 *D* and 4 *D*, the orientation of the ligand in the encounter conformation with the

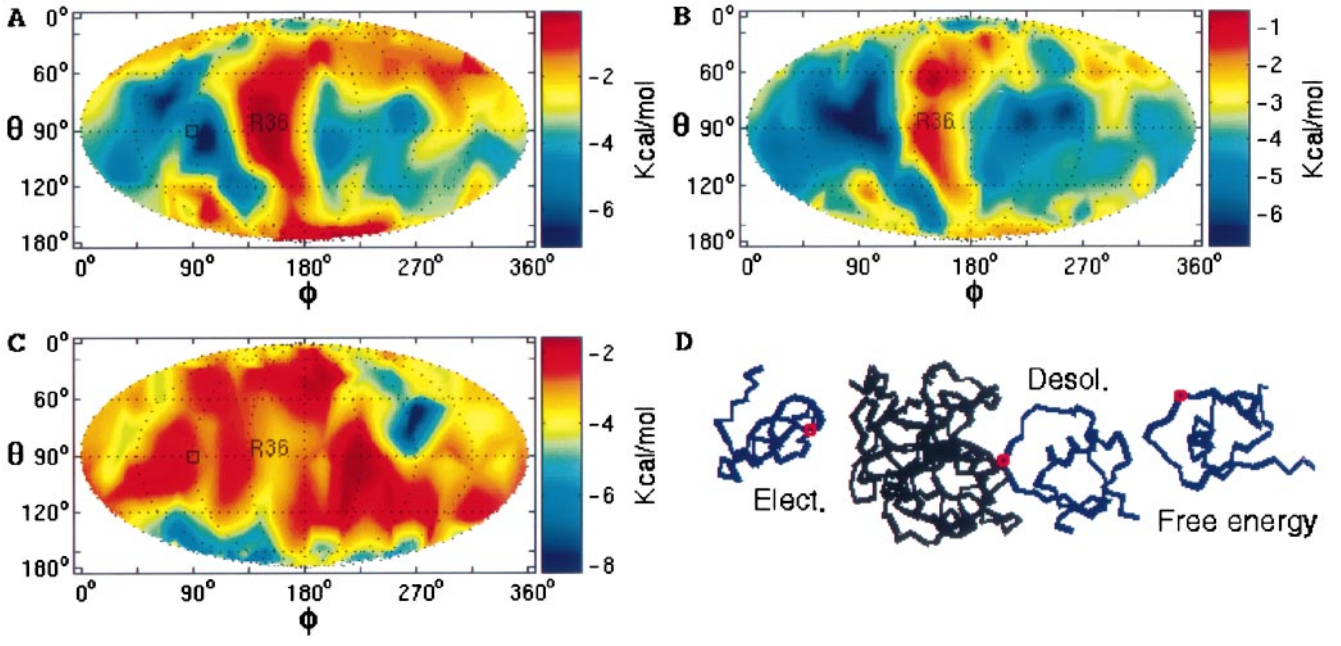

FIGURE 4 Lowest values of interactions in the diffusion-accessible encounter complexes of human leukocyte elastase with turkey ovomucoid third domain (OMTK). (*A*) desolvation energy; (*B*) total free energy of association; (*C*) electrostatic energy. On all maps, the center of the ligand in the x-ray structure is at position (90°, 90°), and the position of a receptor's residue is indicated. (*D*) C-α trace of the receptor (*black*) and traces of the ligand conformations in encounter complexes on which the minimum is reached in (*A*), (*B*), and (*C*). To assist visualization, the trace of ligand configuration with the minimum free energy was shifted to the right. The native complex conformation is not shown, but in *D* it faces the receptor along the positive *x* axis.

lowest value of the desolvation free energy is similar to the orientation observed in the crystal structure, which would position along the positive x axis. In these minima, the stability provided by desolvation alone is -8.16 kcal/mol at position ($\theta = 85°$, $\phi = 85°$) and -7.07 kcal/mol at ($\theta = 95°$, $\phi = 100°$) for Figs. 3 and 4, respectively.

Since the electrostatic contributions are relatively small, the total binding free energy maps in Figs. 3 *B* and 4 *B* closely follow the corresponding maps of the desolvation free energy. As shown in Fig. 3 for the first complex, adding electrostatics makes the region of the minimum desolvation energy smoother and better defined, but both landscapes (desolvation and total free energy) are dominated by the same encounter pair. For the second complex in Fig. 4, the addition of electrostatics slightly shifts the global minimum from ($\theta = 95°$, $\phi = 100°$) to ($\theta = 80°$, $\phi = 95°$). The free energies of these two states are -6.3 kcal/mol and -6.8 kcal/mol, respectively. The minimum free energy conformation shown in Fig. 4 *D* correctly identifies the binding loop in OMTKY, but the minimum desolvation energy provides a slightly better overall orientation.

For each of the two complexes, the landscape of the direct electrostatic energy ΔE_{coul} (Figs. 3 *C* and 4 *C*) shows just one small encounter region with low electrostatic energy (~ -8 kcal/mol). For α-chymotrypsin this region is close to its active site, but the ligand orientation is completely wrong, involving its terminal residues (Fig. 3 *D*). Similarly, Fig. 4 *D* shows that the strong electrostatic attraction of elastase and OMTKY involves residues close to the C-terminals, D214 and K55, respectively. Because of the overall unfavorable desolvation energy of charged residues, we find that these regions tend to contain conformations that have relatively high desolvation energy. In particular, at the global minima of the electrostatic energy in the two complexes the desolvation free energies are positive, 3.3 kcal/mol and 7.5 kcal/mol, respectively. Conversely, at the global minima of the desolvation free energy, the electrostatic contributions are small, -1.59 kcal/mol and 0.81 kcal/mol, respectively.

The typical solvent-accessible areas buried in the encounter conformations are much smaller than those in the final complexes. For chymotrypsin and elastase with OMTKY the buried areas at the minimum desolvation energy are 518 Å^2 and 617 Å^2, whereas in the complex they are 1499 Å^2 and 1357 Å^2, respectively. The distances between the centers of the receptor and the center of the ligand are more than 4 Å larger than in the final complexes. However, the rotational differences between the encounter complex with the lowest desolvation free energy, and the corresponding bound conformation in the complex, are relatively small. In terms of the ligand Euler angles these differences are given by (25°, 29°, 3°) and (3°, 4°, 50°), respectively, for Figs. 3 and 4. The RMSD from the x-ray structures of the two complexes are 8 Å and 10 Å. The minimum free energy ligand conformation in Fig. 4 *D* is rotated by (61°, $-10°$, $-145°$) from the orientation in the complex.

Encounter complexes with strong electrostatic interactions

Kallikrein with pancreatic trypsin inhibitor (PTI) and barnase with barstar are two complexes in which the reactants are oppositely charged (Table 1). As shown in Fig. 2, this leads to long-range electrostatic attraction for both, but only yields strong orientational steering for barnase with barstar. We study these complexes in turn. Fig. 5 shows the maps of the desolvation energy, the total free energy of association, and the electrostatic energy in the diffusion-accessible encounter complexes for kallikrein with PTI. The position of the center of the ligand in the x-ray structure of the complex is at the position ($\theta = 107°$, $\phi = 10.7°$), shown as a small square on all maps. Contrary to the first two cases studied, the desolvation energy landscape shown in Fig. 5 *A* does not delimit the binding region. However, with added electrostatics it does, and the global minimum of the association free energy is found at (95°, 0°), very close to the binding site. At this point the free energy (without $\Delta G_{rot/trans}$) is -15.6 kcal/mol, and only -2.5 kcal/mol comes from desolvation and entropic factors. The second lowest free energy is -15.4 kcal/mol at ($\theta = 115°$, $\phi = 0°$), and the desolvation at this point is positive, 0.6 kcal/mol. In both encounter complexes the ligand's orientation is close to the one in the x-ray structure of the complex, with Euler angle differences (26°, $-14°$, $-27°$) and (18°, 0°, $-60°$), respectively.

As shown in Fig. 5 *C*, direct electrostatics alone yields a strong attractive pocket around the positions (65°, 45°) and (45°, 40°), with the minimum value of -17.0 kcal/mol and orientation shown in Fig. 5 *D*. However, in this region the desolvation term ΔG_{ACE} is over 6 kcal/mol, and thus the total free energy of association is not the most favorable. There is a second pocket with favorable electrostatics and unfavorable desolvation at positions ($\theta = 120°$, $\phi = 20°$) and ($\theta = 80°$, $\phi = 20°$). These are close to the binding site, but the ligand is oriented transversely from its correct position. Since we use a rigid body approximation when generating conformations for the encounter complex, the large exposed side chains of some charged residues prevent the close association of the reactants in the encounter complex, resulting in the burial of relatively small surface areas. For the two conformations with the lowest and second-lowest free energy, the buried areas are 394 Å^2 (for Fig. 5 *D*) and 465 Å^2, whereas in the complex the buried area is 1440 Å^2. The RMSDs of these intermediates from the x-ray structures of the corresponding complexes are 11 Å and 12 Å, respectively.

Fig. 6 shows the maps of the desolvation energy, the total free energy of association, and the electrostatic energy in the diffusion-accessible encounter complexes for barnase with barstar. The center of the ligand in the x-ray structure is at ($\theta = 99°$, $\phi = 11.6°$). For this complex the desolvation free energy does not have a global minimum near the binding region (Fig. 6 *A*). While the region around the binding site has a relatively low free energy, even the closest local minima are far apart (Fig. 6 *B*).

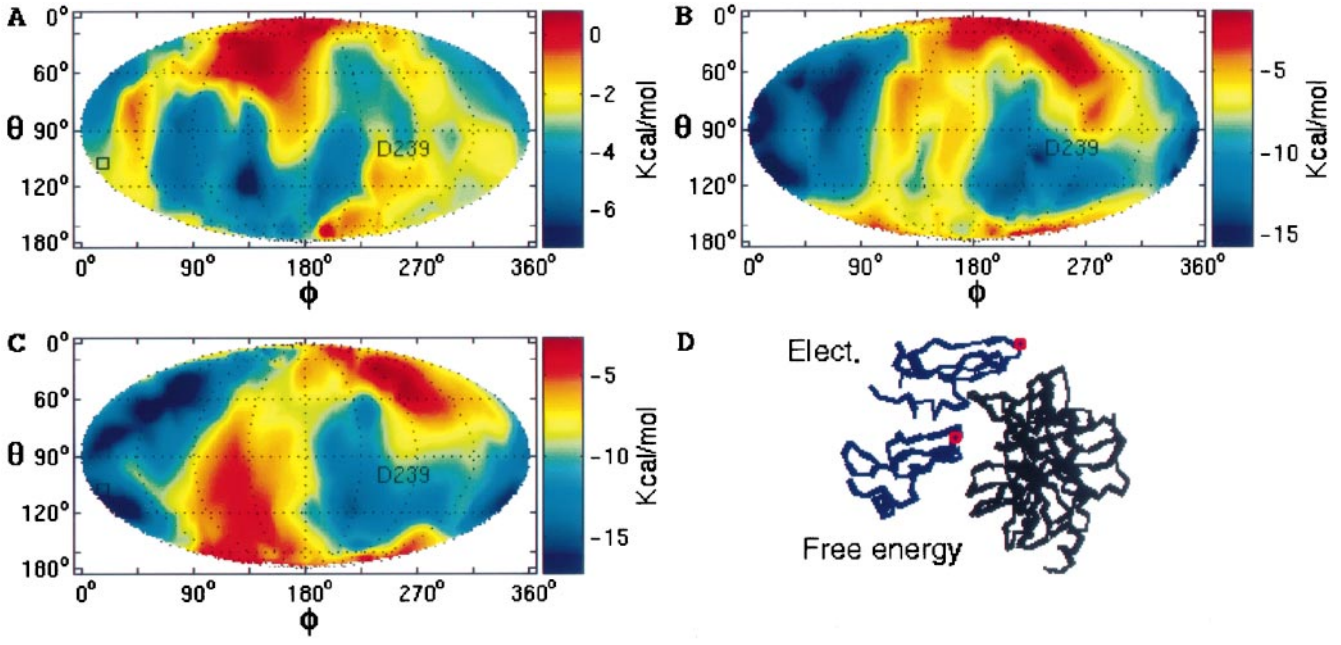

FIGURE 5 Lowest values of interactions in the diffusion-accessible encounter complexes of kallikrein A with pancreatic trypsin inhibitor (PTI). (*A*) desolvation energy; (*B*) total free energy of association; (*C*) electrostatic energy. On all maps, the center of the ligand in the x-ray structure is at position (107°, 10.7°). (*D*) C-α trace of the receptor (*black*) and traces of the ligand conformations in encounter complexes on which the minimum is reached in (*B*) and (*C*). The starting orientation for this complex is the same as in the x-ray structure of the complex, i.e., at position (107°, 10.7°) on the receptor's surface. The native complex conformation is not shown, but in *D* it faces the receptor along the negative *x* axis.

For barnase/barstar the analysis of electrostatic interactions at 5 Å separation showed strong orientational steering. The global minimum of the electrostatic energy of the encounter complexes is −9.2 kcal/mol at (115°, 15°), very close to the binding site. The corresponding orientation is shown in Fig. 6 *D*. The second-lowest electrostatic energy is −9.0 kcal/mol at (100°, 20°). The orientational mismatches of these two configurations, with respect to the x-ray structure, are also relatively small, (5.2°, 19°, −27°) and (17°, 62°, 0°), respectively. Consistent with the diffusional encounter complex found in a Brownian dynamics simulation (Gabdoulline and Wade, 1997), both electrostatic intermediates have the barstar's helix (residues 66–77), tilted toward the guanine binding loop (residues 57–60) of the barnase (see Fig. 6 *D*), supporting the role of this loop as the recognition site for barstar.

Although these conformations are relatively well oriented, their desolvation energy is high, 6.2 and 6.6 kcal/mol, respectively. As we will discuss, the resulting free energies of −3.0 kcal/mol and −2.4 kcal/mol provide only a weak affinity. Since barnase/barstar long-range electrostatics is an important factor, it is possible that once electrostatics steer the ligand into its attractive pocket, desolvation forces adjust the encounter complex by moving the ligand toward the *local* free energy minimum. In agreement with this model, in barnase/barstar the grid point closest to the binding site, (θ = 95°, ϕ = 20°), is in the region of low electrostatics, and is a local minimum of the free energy, with the function value $\Delta E_{coul} + \Delta G_{PB} = -5.5$ kcal/mol. This conformation has an orientation very similar to that of the bound ligand, with Euler angle differences of (−4°, −21°, 14°). Shifting to this conformation from the position with the lowest electrostatics moves barstar away from the guanine binding loop (Fig. 6 *D*). This observation highlights the intricate nature of the binding pathways, since the charges on this loop substantially contribute to long-range electrostatic steering, but the loop is eventually discarded as a good interaction site due to its unfavorable short-range desolvation.

The buried surface areas of intermediates with the lowest and the second-lowest electrostatic energies are 329 Å^2 (Fig. 6 *D*) and 517 Å^2, whereas the complex buried surface is 1585 Å^2. The corresponding RMSDs from the x-ray structure are 10 Å and 11 Å, respectively.

DISCUSSION

Analysis of free energy landscapes

In this paper we calculated the interactions that can stabilize diffusion-accessible encounter pairs in three distinct classes of protein-protein complexes. The first class includes α-chymotrypsin and human leukocyte elastase, both interacting with OMTKY, a neutral ligand. The second is represented by kallikrein A and PTI, a complex which has strong charge complementarity and long-range electrostatic interactions, but apparently no strong steering toward a particular orientation of the encounter complex (Fig. 2). The third category is represented by barnase and barstar, probably the best studied example of rapid, electrostatically assisted protein association for which the analysis of long-

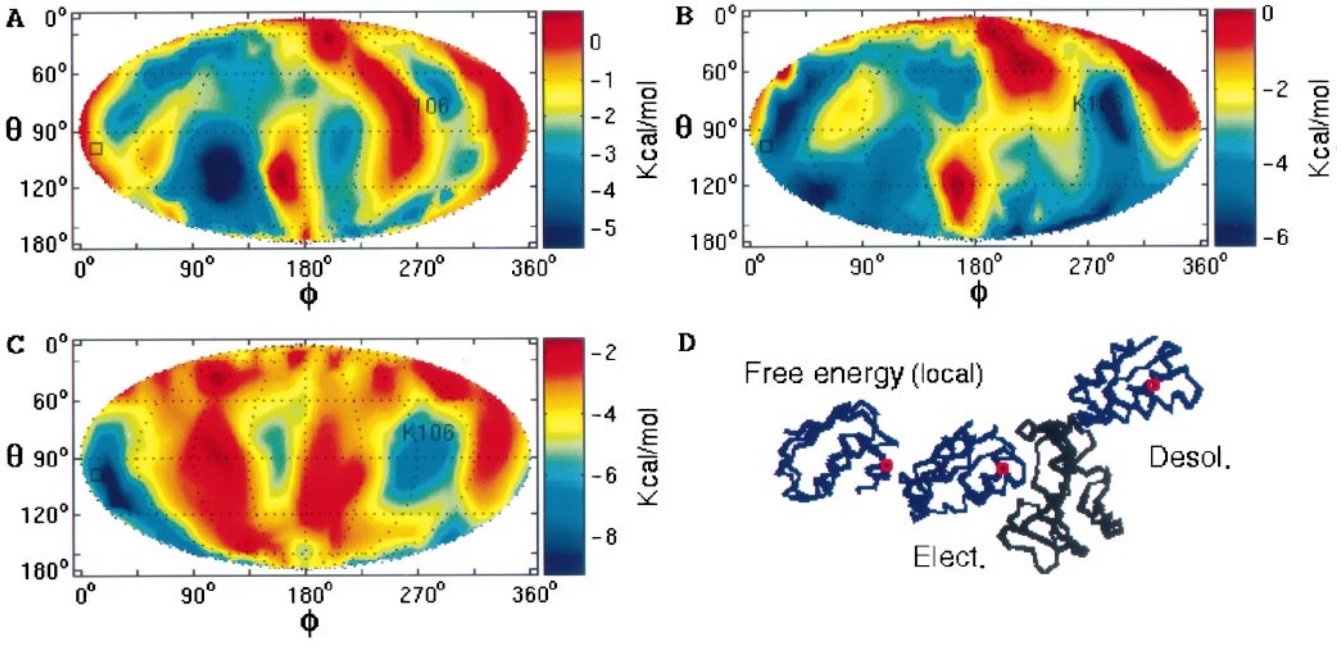

FIGURE 6 Lowest values of interactions in the diffusion-accessible encounter complexes of barnase with barstar. (*A*) desolvation energy; (*B*) total free energy of association; (*C*) electrostatic energy. On all maps, the center of the ligand in the x-ray structure is at position (99°, 11.6°). (*D*) C-α trace of the receptor (*black*) and traces of the ligand conformations in encounter complexes on which the minimum is reached in (*A*) and (*B*). The trace of the local free energy minimum at (95°, 20°) is also shown, shifted to the left to assist visualization. It is noteworthy that the main difference between the electrostatic and local free energy minima is a rotation removing residues 66–77 away from the surface of barnase (near residues 57–60). For this complex the coordinates of the receptor were reoriented, shifting the locus of the ligand to (99°, 11.6°) on the receptor's surface. The native complex conformation is not shown, but in *D* it faces the receptor along the negative *x* axis.

range electrostatic interactions reveals a strong orientational steering.

For each of these receptor-ligand systems we mapped the desolvation free energy ΔG_{ACE}, the direct electrostatic energy ΔE_{coul}, and the sum $\Delta G_{ACE} + \Delta E_{coul}$ over the conformational space of diffusion-accessible encounter pairs. The stability of the encounter pair is governed by the free energy of association calculated as $\Delta G = \Delta G_{ACE} + \Delta E_{coul} + \Delta G_{rot/trans}$, where $\Delta G_{rot/trans}$ is the free energy change associated with the loss of six rotational-translational degrees of freedom. While it is generally accepted that for protein-protein complexes $\Delta G_{rot/trans}$ can be considered constant, the values reported in the literature vary between 6.0 kcal/mol (Janin, 1997) and 15 kcal/mol (Janin, 1995). The particular value of $\Delta G_{rot/trans}$ is not critical, since the desolvation and electrostatic interactions reported here should nevertheless increase the lifetime of the encounter complexes, thus enhancing the association rate whenever $\Delta G_{ACE} + \Delta E_{coul} < 0$. However, we will say that the interactions in an encounter complex are attractive if the electrostatic and desolvation energies can compensate for the loss of translational/rotational entropy. Using the most conservative estimate, an attractive free energy interaction will be such that $\Delta G_{ACE} + \Delta E_{coul} + \Delta G_{rot/trans} < 0$, where $\Delta G_{rot/trans} \simeq 6$ kcal/mol.

It is important to stress that the energy landscapes shown in Figs. 3–6 correspond to the optimal orientation of ligands for every position on the receptor's surface. However, the average interaction over all possible conformations is repulsive. Indeed, for chymotrypsin and elastase in association with OMTKY the average desolvation $\langle \Delta G_{ACE} \rangle$ is 1.1 ± 0.1 and 1.2 ± 0.1 kcal/mol, respectively, whereas the electrostatic interactions $\langle \Delta E_{coul} \rangle$ average out to zero. Complexes with strong charge complementarity have a potentially critical problem, namely the higher likelihood of nonspecific aggregation with other charged proteins. For these complexes we find $\langle \Delta G_{ACE} \rangle = 5.7 \pm 0.3$ kcal/mol and $\langle \Delta E_{coul} \rangle = -1.8 \pm 0.2$ kcal/mol for kallikrein/PTI, and $\langle \Delta G_{ACE} \rangle = 5.9 \pm 0.3$ kcal/mol and $\langle \Delta E_{coul} \rangle = -0.21 \pm 0.03$ kcal/mol for barnase/barstar. Hence, nonspecific aggregation is avoided by a strong short-range repulsion. The overall repulsion of desolvation interactions for all four systems is consistent with the negligible nonspecific protein aggregation found in equilibrium.

Desolvation-mediated specificity and stability

As shown by Figs. 3 and 4, for α-chymotrypsin and human leukocyte elastase with OMTKY the partial desolvation of the encounter pairs is the dominant driving force in binding, and it provides both the stability and the specificity of the encounter complexes. Indeed, it can compensate for $\Delta G_{rot/trans}$. The specificity follows from the fact that, for both complexes, the minimum of the desolvation energy is found at the grid point closest to the receptor's binding site.

Moreover, as shown in Figs. 3 *D* and 4 *D*, the desolvation free energy have the ligands properly oriented with respect to the complex crystal structure.

We have shown that hydrophobicity can serve as the attractive potential contributing to the stability of the encounter conformations. However, as we will discuss, the best-studied examples of rate enhancement exhibit strong electrostatic interactions, as in the association of barnase and barstar, and long-range electrostatics is frequently regarded as the typical driving force in protein-protein recognition (Janin, 1997). Our results show that the mechanism elucidated from barnase and barstar cannot be trivially generalized to all complexes.

Models of desolvation-mediated recognition

In the introduction we reviewed two models that have been used to explain the large association rate constant observed for a number of protein-protein complexes. The classical model, originally developed to describe DNA-protein association (with electrostatics as the adhesive force) assumes the formation of a low-affinity encounter complex stabilized by an attractive potential which held the molecules together long enough to increase the chance of aligning the reactive groups. The potential must be nonspecific and relatively weak to allow for the ligand to slide along the receptor (Berg and von Hippel, 1985; von Hippel and Berg, 1989). The second model, proposed by Northrup and Erickson (1992) and supported by Brownian dynamic simulations, assumes that the partially folded complexes are stabilized by a short-range, specific potential that locks the encounter complex in particular conformations, thereby restricting the configurational space to be searched for the bound state.

We argue that the behavior of encounter complexes for α-chymotrypsin and human leukocyte elastase with OMTKY can be best described by a combination of the two models. First, for every point of the receptor surface, there exist ligand orientations such that the combination of desolvation and electrostatic interactions provide additional stability in a nonspecific fashion, i.e., $\Delta G_{ACE} + \Delta E_{coul} < 0$, contributing to the diffusion entrapment. This is in agreement with the classical model. In a region around the binding site, however, the strong partial desolvation yields $\Delta G_{ACE} + \Delta E_{coul} < -6.0$ kcal/mol, and thus results in a net attraction even when accounting for the loss of rotational/translational entropy. Since this interaction is short-range and specific to a relatively small fraction of the conformational space, it tends to "lock" encounter complexes into this region. Therefore, the model proposed by Northrup and Erickson (1992) seems to be more appropriate. However, within the region of strong desolvation the free energy function is relatively flat, suggesting a weakly specific reaction complex. This complex is stable enough to allow a local surface-on-surface diffusion-mediated search for the combining site, thus within this region the classical model may be again useful for describing the transition state.

Hierarchy of electrostatics and desolvation effects

While electrostatic interactions play an important role in the association of kallikrein with PTI and barnase with barstar, both with oppositely charged reactants, their detailed analysis uncovered substantial differences. In barnase-barstar, long-range electrostatics has a strong orientational steering toward the region of the binding site, thereby increasing the probability of favorable encounter complex conformations even at the first collision. The optimal association free energy $\Delta E_{coul} + \Delta G_{ACE}$ is attractive over a relatively large region, and the attraction there seems to be nonspecific enough to allow for a vigorous rearrangement of encounter complex conformations in the transition state. We have even argued that once electrostatic forces steer the ligand into this attractive pocket, short-range electrostatics plus desolvation further steer the ligand toward the local free energy minima.

These findings are in good agreement with the experimental results of Schreiber and Fersht (1996). In the absence of long-range electrostatic forces, the association rate of barnase and barstar drops by five orders of magnitude, emphasizing the role of long-range electrostatic steering. However, even in this regime, partially desolvated intermediates would provide the local adhesion needed to overcome the loss of rotational and translational entropy. This indicates that the low-affinity state before forming the complex is steered and held together by mostly electrostatic forces. However, desolvation interactions filter out conformations with unfavorable desolvation energies, and they may even play a crucial kinetic role, preventing the guanine binding loop from interfering with the productive binding pathways.

The kallikrein-PTI complex provides a good example of what should be expected in heavily charged complexes in which the overall stability may be due to electrostatics, but the affinity toward the binding region is also affected by desolvation. While none of these forces confine the binding region on their own, the total association free energy does have its minimum in the right place. Through the strong and rather nonspecific electrostatic attraction might unfavorably affect the association rate by slowing down the search for the mutually reactive orientations.

Is desolvation fast enough?

We have shown that without accounting for desolvation it would be difficult to rationalize the specificity and stability of intermediate states in three of the four complexes studied. However, desolvation of mobile intermediates can be a factor only if it is fast enough compared to the diffusion limited lifetime of the transition state. The time scale of the microcollisions between proteins is $\bar{t} \sim R^2/D$, where R is a length scale on the order of the radius of one protein and D is the relative translational diffusion constant of the receptor and ligand. Typical values of $R = 30$ Å and $D = 20$ Å^2/ns (Creighton, 1993) yield $\bar{t} \approx 5 \times 10^{-8}$ s. Information on the residence time of water molecules near protein surfaces can

be obtained by NMR techniques (Karplus and Faerman, 1994). According to such data, a handful of buried water molecules have very long residence times ranging from 10^{-8} to 10^{-2} s (Levitt and Park, 1993). However, the surface waters of hydration are in rapid exchange with bulk water and have residence times below 500 ps (Otting and Wuthrich, 1989), which is two orders of magnitude faster than the process of diffusion within the encounter complex.

The length scale associated with the manifestation of desolvation forces must be on the order of 3- to 6-Å separation between atom surfaces (i.e., 6- to 9-Å separation between atomic centers). At this distance the steric hindrance of the first layers of water molecules becomes relevant, and water molecules start to move out of the entropically unfavorable interface faster than those that move in. This is also the length scale at which the Poisson-Boltzmann electrostatic energy is affected by desolvation effects (see Fig. 2).

Protein binding and protein folding: a common framework

It has long been recognized that protein binding and folding respond to a similar set of principles (Creighton, 1993). However, little concrete evidence has been presented on whether the kinetic and thermodynamic (Zhang et al., 1997b) properties of these two apparently related problems have anything in common.

The results of this paper suggest interesting similarities between protein-protein association, and protein folding for which a similar multistage mechanism has been proposed (Camacho and Thirumalai, 1993). The model offered here, i.e., the selection of the binding region by means of partially desolvated intermediates prior to the formation of the fully desolvated interface, is certainly consistent with the general view of the driving forces in protein folding (Dill, 1990). It is also consistent with a recent prediction of a theoretical model (Camacho, 1996) of protein folding, where it was found that the limiting step is the accessibility of partially folded intermediates before the late transition to the native state, a possibility first envisaged in Camacho and Thirumalai (1993). Another similarity unveiled by the energy landscapes is that the main barriers in both binding and folding appear to be entropic rather than enthalpic. In binding, the barrier is due to the loss of rotational and translational entropy, whereas in folding it is due to the loss of configurational (mostly backbone) entropy (Schellman, 1955; Nemethy and Scheraga, 1965; Camacho, 1996; Zhang et al., 1997b). Finally, our conclusion that the origin of the specificity in at least some protein complexes is given by desolvation rather than by direct electrostatics is consistent with a prediction by Hendsch and Tidor (1994) who suggested that salt bridges are not that important in protein folding.

Contrary to geometric complementarity methods (Wodak and Janin, 1978; Shoichet and Kuntz, 1991) of assessing protein-protein recognition, we provide perhaps a first glimpse of how protein interactions could lead to specific conformations as suggested by the simple locking model of Northrup and Erickson (1992), without an excruciating search over an astronomically large number of possible states. The apparent consistency between binding and folding confirms that a similar set of principles governs these processes.

FINAL REMARKS

According to our results, the role played by long-range electrostatics in enhancing the association of barnase and barstar cannot be trivially generalized to other complexes. However, the formation of a low-affinity, weakly specific complex, held together by *both* electrostatic and/or desolvation forces, appears to be the key step in protein binding, preceding the transition to the docked conformation. The rather broad free energy bottlenecks of these well-oriented intermediates efficiently traps the receptor-ligand pair, allowing for the more microscopic search of the pathways leading to the high affinity complex.

The structural characteristics of the encounter pairs at the onset of a productive association are a root-mean-square-deviation of 10 ± 1 Å from the complex structure (including a 4- to 5-Å translation), and solvent-accessible areas buried on the order of 400 ± 100 Å^2. It should be pointed out that a less coarse-grained sampling of the interfacial free energy would only result in the enhancement of short-range electrostatic contacts (~ -1 kcal/mol) of the artificially rigid encounter pairs. Since the entropy of these conformations is very small, their overall impact on the binding pathways is expected to be very limited. The above notwithstanding, the true short-range electrostatic interactions generated by contacts of charges and partial charges at the flexible interface should play a critical role on the side-chain rearrangements that take place in the late transition leading to the complex structure. Indeed, these short-range electrostatic contacts have an important contribution to the overall stability of the complex of both charge and neutral proteins (see, e.g., Zhang et al., 1997a).

Desolvation is the dominant driving force in binding for proteins with weak charge complementarity. The energy landscapes suggest two mechanisms for protein binding, i.e., electrostatically driven (Schreiber and Fersht, 1996) and desolvation-mediated. The desolvation and electrostatic stability threshold required by the partially desolvated intermediates to identify the binding region is found to be ~ -6 kcal/mol, whereas the overall repulsion that should prevent aggregation is found to be on the order of 1 kcal/mol for the association of neutral proteins and 3–4 kcal/mol for alike charged proteins. At this point a word of caution is in order since ultimately the true kinetic implications of these energies and length scales should be corroborated by dynamic simulations of the receptor and ligand system.

C. J. C. is indebted to C. Zhang for the stimulating role he played in this project, and for sharing his deep insights on the forces in proteins.

This research was supported by Grant DBI-9630188 from the National Science Foundation, and by the Donors of The Petroleum Research Fund.

REFERENCES

Berg, O. G., and P. H. von Hippel. 1985. Diffusion-controlled macromolecular interactions. *Ann. Rev. Biophys. Biophys. Chem.* 14:131–160.

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 dynamic calculations. *J. Comput. Chem.* 4:187–217.

Bruccoleri, R. E., J. Novotny, M. Davis, and K. A. Sharp. 1997. Finite difference Poisson-Boltzmann electrostatic calculations: increased accuracy achieved by harmonic dielectric smoothing and charge antialiasing. *J. Comp. Chem.* 18:268–276.

Camacho, C. J. 1996. Entropic barriers, frustration, and order: basic ingredients in protein folding. *Phys. Rev. Lett.* 77:2324–2327.

Camacho, C. J., and D. Thirumalai. 1993. Kinetics and thermodynamics of folding in model proteins. *Proc. Natl. Acad. Sci. USA.* 90:6369–6372.

Creighton, T. E. 1993. Proteins. W. H. Freeman and Co., New York.

DeLisi, C. 1980. The biophysics of ligand-receptor interactions. *Q. Rev. Biophys.* 13:201–230.

Dill, K. 1990. Dominant forces in protein folding. *Biochemistry.* 29: 7133–7155.

Gabdoulline, R. R., and R. C. Wade. 1996. Effective charges for macromolecules in solvent. *J. Phys. Chem.* 100:3868–3878.

Gabdoulline, R. R., and R. C. Wade. 1997. Simulation of the diffusional association of barnase and barstar. *Biophys. J.* 72:1917–1929.

Goodsell, D. S., and A. J. Olson. 1990. Automated docking of substrates to proteins by simulated annealing. *Proteins.* 8:195–202.

Hendsch, Z. S., and B. Tidor. 1994. Do salt bridges stabilize proteins? *Protein Sci.* 3:211–226.

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

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

Israelachvili, J., and H. Wennerstrom. 1996. Role of hydration and water structure in biological and colloidal interactions. *Nature.* 379:219–225.

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.

Janin, J. 1995. Elusive affinities. *Proteins: Struct. Funct. Genet.* 21:30–39.

Janin, J. 1997. The kinetics of protein-protein recognition. *Proteins: Struct. Funct. Genet.* 28:153–161.

Karplus, P. A., and C. Faerman. 1994. Ordered water in macromolecular structure. *Curr. Opinion Struct. Biol.* 4:770–776.

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

Levitt, M., and B. H. Park. 1993. Water: Now you see it, now you don't. *Structure.* 1:223–226.

Miyazawa, S., and R. L. Jernigan. 1985. Estimation of effective interresidue contact energies from protein crystal structures: quasi-chemical approximation. *Macromolecules.* 18:534–552.

Nauchitel, V., M. C. Villaverde, and F. Sussman. 1995. Solvent accessibility as a predictive tool for the free energy of inhibitor binding to the HIV-1 protease. *Protein Sci.* 4:1356–1364.

Nemethy, G., and H. A. Scheraga. 1965. Theoretical determination of sterically allowed conformations of a polypeptide chain by a computer method. *Biopolymers.* 3:155–184.

Northrup, S. H., S. A. Allison, and J. A. McCammon. 1984. Brownian dynamics simulation of diffusion-influenced bimolecular reactions. *J. Chem. Phys.* 80:1517–1524.

Northrup, S. H., and H. P. Erickson. 1992. Kinetics of protein-protein association explained by Brownian dynamics computer simulations. *Proc. Natl. Acad. Sci. USA.* 89:3338–3342.

Novotny, J., R. E. Bruccoleri, and F. A. Saul. 1989. On the attribution of binding energy in the antigen-antibody complexes mcpc 603, d1.3, and hyhel-5. *Biochemistry.* 28:4735–4749.

Noyes, R. M. 1961. Effects of diffusion rates on chemical kinetics. *Prog. React. Kinet.* 1:129–160.

Otting, G., and K. Wuthrich. 1989. Studies of protein hydration in aqueous solution by direct NMR observation of individual protein-bound water molecules. *J. Am. Chem. Soc.* 111:1871–1975.

Rosenfeld, R., S. Vajda, and C. DeLisi. 1995. Flexible docking and design. *Annu. Rev. Biophys. Biomol. Struct.* 24:677–700.

Schellman, J. 1955. The stability of hydrogen-bonded peptide structures in aqueous solution. *Trav. Lab. Carlsberg Ser. Chim.* 29:230–259.

Schreiber, G., and A. R. Fersht. 1996. Rapid, electrostatically assisted association of proteins. *Nature Struct. Biol.* 3:427–431.

Sharp, K., R. Fine, and B. Honig. 1987. Computer simulations of the diffusion of a substrate to an active site of an enzyme. *Science.* 236: 1460–1463.

Sharp, K., and B. Honig. 1990. Electrostatic interactions in macromolecules: theory and applications. *Annu. Rev. Biophys. Biophys. Chem.* 19:301–332.

Shoichet, B. K., and I. D. Kuntz. 1991. Protein docking and complementarity. *J. Mol. Biol.* 221:327–346.

Sommer, J., C. Jonah, R. Fukuda, and R. Bersohn. 1982. Production and subsequent second-order decomposition of protein disulfide anions. *J. Mol. Biol.* 159:721–744.

Stones, R. S., S. Dennis, and J. Hofsteenge. 1989. Quantitative evaluation of the contribution of ionic interactions to the formation of thrombin-hirudin complex. *Biochemistry.* 28:6857–6863.

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.

von Hippel, P. H., and O. G. Berg. 1986. On the specificity of DNA-protein interactions. *Proc. Natl. Acad. Sci. USA.* 83:1608–1612.

von Hippel, P. H., and O. G. Berg. 1989. Facilitated target location in biological systems. *J. Biol. Chem.* 264:675–678.

Wallis, R., G. R. Moore, R. James, and C. Kleanthous. 1995. Protein-protein interactions in colicin E9 DNase-immunity protein complexes. 1. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry.* 34:13743–13750.

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

Wodak, S. J., and J. Janin. 1978. Computer analysis of protein-protein interaction. *J. Mol. Biol.* 124:323–342.

Zhang, C., G. Vasmatzis, J. L. Cornette, and C. DeLisi. 1997a. Determination of atomic desolvation energies from the structures of crystallized proteins. *J. Mol. Biol.* 267:707–726.

Zhang, C., J. L. Cornette, and C. DeLisi. 1997b. Consistency in structural energetics of protein folding and peptide recognition. *Protein Sci.* 6:1057–1064.