Protein Science (1996), 5:614-626. Cambridge University Press. Printed in the USA.

Prediction of protein complexes using empirical free energy functions

ZHIPING WENG, SANDOR VAJDA, AND CHARLES DELISI
Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215

(RECEIVED November 14, 1995; ACCEPTED January 11, 1996)

Abstract

A long sought goal in the physical chemistry of macromolecular structure, and one directly relevant to understanding the molecular basis of biological recognition, is predicting the geometry of bimolecular complexes from the geometries of their free monomers. Even when the monomers remain relatively unchanged by complex formation, prediction has been difficult because the free energies of alternative conformations of the complex have been difficult to evaluate quickly and accurately. This has forced the use of incomplete target functions, which typically do no better than to provide tens of possible complexes with no way of choosing between them. Here we present a general framework for empirical free energy evaluation and report calculations, based on a relatively complete and easily executable free energy function, that indicate that the structures of complexes can be predicted accurately from the structures of monomers, including close sequence homologues. The calculations also suggest that the binding free energies themselves may be predicted with reasonable accuracy. The method is compared to an alternative formulation that has also been applied recently to the same data set. Both approaches promise to open new opportunities in macromolecular design and specificity modification.

Keywords: desolvation free energy; electrostatic interaction energy; rigid body docking; side-chain conformational search

Understanding biological function at its most fundamental level, and manipulating it for therapeutic purposes, requires understanding the molecular basis of specificity. The critical determinant of true understanding is the ability to make quantitative predictions. In the case of molecular recognition, this means predicting the geometry of a complex from the free structures of its constituents.

Even when conformations remain relatively unperturbed by reaction, thereby restricting the search for potential molecular complexes to a manageable size, achieving the goal of true prediction has proved elusive. Rigid body docking algorithms (Cherfils et al., 1991; Jiang & Kim, 1991; Shoichet & Kuntz, 1991; Bacon & Moult, 1992; Stoddard & Koshland, 1992; Walls & Sternberg, 1992; Helmer-Citterich & Tramontano, 1994; Judson et al., 1995) permit exploration of the orientations of a ligand in a receptor combining site, and effectively select those relatively few that meet prespecified criteria—including surface complementarity, interaction energy, and hydrophobic surface burial.

As shown by Shoichet and Kuntz (1991), the methods go far toward reducing hundreds of thousands of possibilities to tens of possibilities, one of which is usually within 1-2 Å of the observed structure. However, even when the free molecules are chemically identical to those in the complex and their conformations differ only slightly (for example, independently crystallized structures), the differences within the final set of indistinguishable structures sometimes exceed 10 Å, and almost always exceed 5 Å (Cherfils et al., 1991; Jiang & Kim, 1991; Shoichet & Kuntz, 1991; Bacon & Moult, 1992). The problem becomes even more severe for the more general and more frequently arising situation of predicting a complex when only the homologues of the reactants are available crystallographically, rather than the reactants themselves.

The difficulty reflects the limitations of incomplete target functions. However, evaluating the ideal target function, viz the binding free energy, using standard methods such as free energy perturbation or thermodynamic integration, is computationally far too demanding for docking.

An alternative is to evaluate the binding free energy empirically. Depending on the method of treating solvation, there are two main approaches to empirical free energy evaluation. The most direct way to obtain the electrostatic contributions to the solvation free energy is by solving Poisson's equation (or the Poisson-Boltzmann equation if free ions are included). The total free energy change is then obtained by evaluating the desolvation entropy gain empirically. This electrostatic free energy

Reprint requests to: Charles DeLisi, Department of Biomedical Engineering, Boston University, Boston, Massachusetts 02215; e-mail: delisi@buenga.bu.edu.

(EFE) approach, introduced by Smith and Honig (1994), has been used by Zachary and Tidor (1994) for calculating the electrostatic contribution of salt bridges to the free energy of folding, and has been extended recently to the analysis of protein-protein interactions by Jackson and Sternberg (1995), who applied the method to the docked conformations generated by Shoichet and Kuntz (1991) using three different protease-protein inhibitor systems.

An alternative, which is rapid enough to be generally useful, relates solvation free energy to solvent-accessible surface areas or hydration shell volumes through atomic solvation parameters (Chothia & Janin, 1975; Eisenberg & Mclachlan, 1986; Ooi et al., 1987; Colonna-Cesari & Sander, 1990; Cramer & Truhlar, 1992; Wesson & Eisenberg, 1992; Stouten et al., 1993). In this case, the empirical procedure takes account of not just surface area (entropy), but of the types of atomic groups (enthalpy) on the surface. Atomic solvation parameter (ASP) models have been incorporated into various empirical free energy functions (Novotny et al., 1989; Wilson et al., 1991; Horton & Lewis, 1992; Wesson & Eisenberg, 1992; Murphy et al., 1993; Vajda et al., 1994). We recently evaluated this method on several data sets, including endopeptidase-inhibitor complexes, and found it to be reasonably accurate (Vajda et al., 1994).

In this paper, we apply an ASP-based binding free energy function to the three protease-protein inhibitor complexes studied by Shoichet and Kuntz (1991) and Jackson and Sternberg (1995). Both target functions, their formal differences not withstanding, successfully discriminate between correct and incorrect complex conformations. However, the ASP method is substantially faster, and yields better agreement with the experimentally determined binding free energies. We also show that a conformational search applied to the important residues in the binding region further improves the results. If the free monomers are used in docking, the lowest free energy is attained on complexes that are within 1–2 Å all-atom RMS deviation (RMSD) from the crystallographically observed complex. This finding suggests that an effective docking algorithm can be obtained by rapid rigid body docking with a conventional scoring function such as surface complementarity (Shoichet & Kuntz, 1991) to generate a small number of candidates, followed by a side-chain search and free energy evaluations to distinguish between them.

Empirical free energy evaluation principles

The free energy change accompanying complex formation is conveniently divided into two components: one dependent on environment and conformation (ΔG_{var}); the other—to a good first approximation—independent of both (ΔG_{const}). ΔG_{const} consists of rigid body translational and rotational free energies, including the necessary dilution of the number of independent molecules (the so-called cratic term) resulting from complex formation. It is roughly a constant for different protein complexes, with estimates in the range of 7–15 kcal/mol (Page & Jenks, 1971; Erickson, 1989; Finkelstein & Janin, 1989; Tidor & Karplus, 1994). We have set its value at 9 kcal/mol (Novotny et al., 1989; Vajda et al., 1994), regardless of data set, although the precise value is inconsequential for docking. ΔG_{var}, referred to as the potential of mean force by Jackson and Sternberg (1995), contains everything else, including desolvation and conformational free energy changes (only side-chain entropies if, apart from possible side-chain movements, the molecules are considered rigid).

The desolvation component of ΔG_{var} is composed of the energy change associated with changes in the environment of atoms in the vicinity of the contact area of the complex, and the entropy change in water. The first of these generally has both van der Waals and electrostatic components. However, the receptor and ligand are assumed to bind and adjust to one another in a way that keeps van der Waals contacts relatively invariant; i.e., van der Waals contacts lost to water are supposed to be compensated by van der Waals contacts between the proteins (Adamson, 1976; Novotny et al., 1989; Nicholls et al., 1991; Krystek et al., 1993). The energetic part of ΔG_{var} is then entirely electrostatic. If we denote by ΔS_s and ΔS_c as the entropy changes due to desolvation (favorable) and side-chain restrictions (unfavorable) accompanying binding, respectively, ΔG_{var} is given by

$$\Delta G_{var} = E^{rl} + \Delta E^{r} + \Delta E^{l} - T\Delta S_s - T\Delta S_c, \tag{1}$$

where all energy terms have only electrostatic contributions; E^{rl} is the interaction energy between receptor and ligand; and ΔE^{r} and ΔE^{l} denote the internal energy changes of receptor and ligand upon complex formation.

The electrostatic contribution of the free ligand can be written as the sum of two components: a *direct* term, which is the sum of Coulombic interactions between each charge pair; and *indirect* terms arising from a reaction field caused by polarization of water near the surface. As an example, consider two opposite charges at fixed position, separated by r, interior to a sphere (e.g., the ligand) of radius a and dielectric ϵ_p, embedded in a dielectric, ϵ_w (water). Let θ be the angle subtended by the charges at the center of the sphere, and let r_1 and r_2 be the distances from the center. Then (Tanford & Kirkwood, 1957):

$$E = -\frac{e^2}{a\epsilon_p}\left[\frac{a}{r} - \sum_{n=0}^{\infty}\frac{(n+1)(\epsilon_w - \epsilon_p)}{(n+1)\epsilon_w - n\epsilon_p}(r_1 r_2/a^2)^n P_n(\cos\theta)\right]. \tag{2}$$

The second term (the infinite sum) captures the effect of environmental heterogeneity (the dielectric difference between protein and water). It is generally dependent on geometry and is the physical basis for the electrostatic contribution to desolvation.

If the ligand and receptor do not change conformation on binding, their direct electrostatic terms cancel. Equation 1 then becomes

$$\Delta G_{var} = E_d^{rl} + E_i^{rl} + \Delta E_i^{r} + \Delta E_i^{l} - T\Delta S_s - T\Delta S_c, \tag{3}$$

where the subscripts i refer to the energies associated with the induced (i.e., reaction) field, and d denotes the direct Coulomb interaction. The four central terms on the right represent the free energy change due to desolvation, ΔG_s:

$$\Delta G_s = E_i^{rl} + \Delta E_i^{r} + \Delta E_i^{l} - T\Delta S_s. \tag{4}$$

Equation 3 can then be written as

$$\Delta G_{var} = E_d^{rl} + \Delta G_s - T\Delta S_c. \tag{5}$$

The EFE approach calculates all the energy terms in Equation 3 by solving Poisson's equation (Jackson & Sternberg, 1995) for a protein treated as a low dielectric continuum ($\epsilon_p = 2$) with discrete atomic charges, embedded in a high dielectric solvent ($\epsilon_w = 80$). In effect, it is a more realistic version of the Kirkwood model (Equation 2), where the protein sphere is replaced by it actual surface, a change that necessitates abandoning an analytic solution. Recently, Abagyan and Totrov (1994) reported a fast method to obtain the approximate solution of the electrostatic free energy.

A term proportional to the difference between the solvent-accessible surface areas (Smith & Honig, 1994), or the molecular surface areas (Jackson & Sternberg, 1995), of the complex and the free molecules is used to approximate $T\Delta S_s$. Jackson and Sternberg (1995) use an empirical scale for the calculation of side-chain entropy (Pickett & Sternberg, 1993), but they assume that the side-chain entropy loss upon association is offset by more favorable van der Waals interactions, and do not include this term in ΔG_{var}.

The ASP method calculates the desolvation free energy ΔG_s in Equation 5 from the difference of the transfer free energies of complex (ΔG_{tr}^{rl}), receptor (ΔG_{tr}^{r}), and ligand (ΔG_{tr}^{l}):

$$\Delta G_s = \Delta G_{tr}^{rl} - \Delta G_{tr}^{l} - \Delta G_{tr}^{r}. \quad (6)$$

The transfer free energies are obtained by a linear combination of the solvent-accessible areas weighted by ASPs, and the ASPs are derived from the experimental free energies of transferring individual amino acids from octanol or hydrocarbon to water (Vajda et al., 1995). In this way, the ASP method incorporates both energy and entropy contributions of ΔG_s into a single surface area term.

Results

In order to be as explicit as possible about the connection with previous methods, we will focus on the three systems (Table 1A) analyzed by Shoichet and Kuntz (1991) and Jackson and Sternberg (1995): subtilisin complexed with a chymotrypsin inhibitor; chymotrypsin complexed with ovomucoid third domain; and trypsin complexed with bovine pancreatic trypsin inhibitor. The binding free energies (Vincent & Lazdunski, 1972; Svendsen et al., 1980; Bigler et al., 1993) and the crystal structures are available for each system. The structure of each complex is calculated either using the complexed conformers (bound conformer docking), or the isolated monomers or their close homologues (free conformer docking).

Recovering the crystal complex using bound conformers

For bound conformer docking, 8 of 15 structures have an all-atom RMSD within 1 Å of the observed structures, and none exceed 3.3 Å (Table 1A). The free energies of the observed subtilisin/chymotrypsin inhibitor complex and the four structures that are within 0.29 Å of it, calculated according to Equation 5, range between −15.3 and −15.8 kcal/mol, with the crystal structure having a free energy of −15.6 kcal/mol (Svendsen et al., 1980). This is to be compared to the free energies of the two structures having RMSDs of 3.1 and 3.3 Å, which are −12.5 and −12.0 kcal/mol, respectively. Similar results are obtained for the chymotrypsin/ovomucoid and the trypsin/BPTI complexes.

Some perspective on the magnitude of the binding free energy differences between various docked conformations can be obtained by comparing the differences of measured and calculated binding free energies. These are −15.8/−15.6 for the subtilisin/chymotrypsin inhibitor; −14.4/−14.5 for chymotrypsin/ovomucoid inhibitor and −18.1/−17.2 for trypsin/trypsin inhibitor. The first two results are atypically accurate; more generally, errors of approximately 5% can be expected (Vajda et al., 1994). Therefore, the 3 kcal/mol difference between the low and high RMSD structures appears to be within the resolving power of the free energy function, whereas the 0.2 kcal/mol difference between the lowest free energy structure and the crystal structure is not.

Although the correlation between free energy and RMSD is expected to be imperfect—in part due to inaccuracies in the function and in part due to the ruggedness of the free energy landscape—there is no difficulty using free energy to select the observed structure, or a structure that is within a few tenths of an Å from it (Figs. 1, 2, 3).

Recovering the crystal complex using free conformers

Side-chain conformations held fixed

The more general and more difficult problem is to predict the structure of a complex, starting with the observed structures of free monomers. Whereas the results in Table 1A were obtained by docking structures identical to those in the crystal complex, we now consider monomeric starting structures that differ from their structure in the complex (Table 1B). The difference is present even when the free monomers are identical chemically (i.e., have identical sequences) to those in the complex, because the structures of their surface side chains in the contact region will generally be perturbed by reaction. The trypsin/BPTI inhibitor system falls into this category, viz the crystallized monomeric molecules are identical chemically to the molecules in the crys-

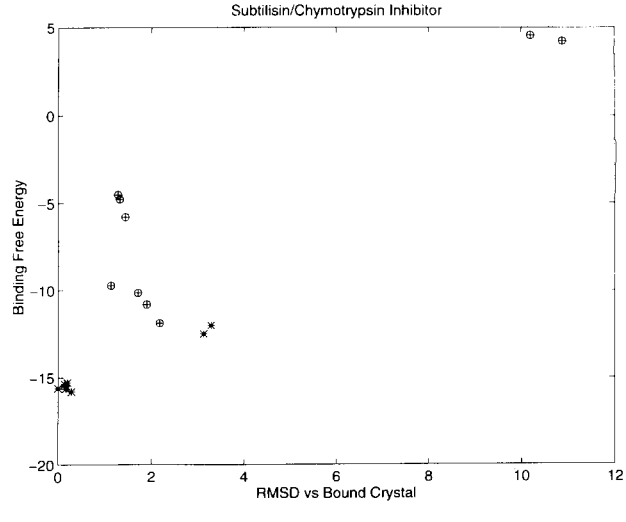

Fig. 1. Binding free energies calculated for the docked conformations of subtilisin/chymotrypsin inhibitor complex, plotted against the RMSD. Free energies are in kcal/mol and RMSDs are in Å and represent the docked conformations using bound and free conformers, respectively. Binding free energies are calculated by Equation 5.

Table 1. *Calculated free energies*[a] *for docked protein complexes*

Structures	RMSD[b] (Å)	E_d^{el}	ΔG_s	$-T\Delta S_c$	ΔG_{cal}
A. Bound					
Substilin/chymotrypsin-inhibitor SNI/CI	0.0	−15.5	−30.6	21.5	−15.6
	0.14	−13.9	−31.3	20.9	−15.4
	0.17	−14.2	−31.6	21.1	−15.6
	0.20	−13.8	−31.3	20.8	−15.3
	0.29	−14.6	−31.4	21.2	−15.8
	3.13	−10.4	−32.6	21.6	−12.5
	3.29	−10.0	−32.9	21.9	−12.0
Chymotrypsin/ovomucoid CHO/OVO	0.0	−17.6	−25.4	19.5	−14.5
	0.72	−14.3	−24.9	18.6	−11.5
	1.17	−14.3	−24.1	18.3	−11.2
	1.18	−13.6	−24.0	18.2	−10.3
	2.44	−13.8	−27.5	20.2	−12.1
Trypsin/BPTI 2PTC/PTI	0.0	−22.1	−19.4	17.1	−17.2
	0.29	−22.4	−19.3	16.9	−15.7
	0.58	−22.7	−19.3	16.9	−16.1
	0.95	−21.5	−18.5	16.5	−14.5
	1.40	−22.2	−18.5	16.5	−15.2
	3.10	−20.7	−18.6	16.3	−14.0
B. Unbound					
Substilin/chymotrypsin-inhibitor 1SBC/2CI2	1.13	−7.7	−27.0	16.0	−9.7
	1.28	−3.2	−25.8	15.5	−4.5
	1.32	−3.4	−26.0	15.6	−4.8
	1.44	−4.3	−26.0	15.5	−5.8
	1.71	−7.8	−26.2	14.9	−10.1
	1.90	−8.0	−26.5	14.6	−10.8
	2.18	−9.4	−27.1	15.7	−11.8
	10.20	3.5	−28.1	20.2	4.5
	10.88	0.3	−28.5	23.4	4.2
Chymotrypsin/ovomucoid 5CHA/2OVO	1.55	−7.7	−26.5	17.9	−7.3
	1.73	−9.9	−27.2	18.6	−9.4
	1.70	−8.2	−26.7	18.4	−7.4
	1.87	−10.3	−25.5	18.2	−8.5
	2.09	−8.8	−25.8	18.3	−7.3
	2.31	−7.3	−27.7	18.9	−7.2
	2.69	−4.0	−29.5	19.9	−4.6
	4.33	−10.4	−27.0	19.9	−8.5
	7.98	−5.4	−24.0	18.5	−1.9
	18.10	5.8	−27.7	19.8	6.9
Trypsin/BPTI 2PTN/4PTI	1.73	−16.7	−15.7	16.1	−7.3
	1.78	−22.0	−17.0	16.3	−13.6
	1.85	−15.4	−16.5	16.3	−6.6
	2.66	−22.8	−17.4	17.1	−14.1
	4.65	−5.3	−15.1	16.3	4.9
	9.30	−18.7	−23.6	21.4	−11.9
	19.31	−18.7	−23.7	21.4	−11.9
	21.09	−13.6	−22.7	19.7	−7.5

[a] In kcal/mol.
[b] RMSD of the docked inhibitor structures from the inhibitor structure in the crystal complex.

tallized complex, but we nevertheless expect structural differences due to reaction. For the chymotrypsin/ovomucoid system, silver pheasant ovomucoid (2OVO) is used as a model for unbound turkey ovomucoid third domain, and residue Leu 18 is replaced by Met; whereas for the subtilisin/chymotrypsin inhibitor system, the monomeric subtilisin Carlsberg (1SBC) differs from the complexed form (subtilisin novo SNI) at six positions (T33S, S99D, S103Q, A129P, S156E, L217Y), all of which are in the binding site.

These caveats not withstanding, among the 27 candidate complexes generated by Shoichet and Kuntz (1991), 13 have all-atom RMSDs within 2 Å of the crystal (Table 1B). Among this same set, however, are seven complexes that have an RMSD in excess of 5 Å, and six that have an RMSD in excess of 10 Å. Free en-

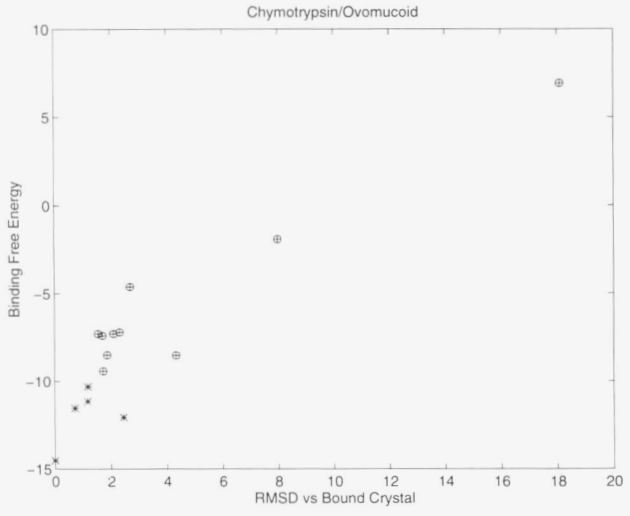

Fig. 2. Same as Figure 1, for chymotrypsin/ovomucoid third domain complex.

ergy assignments differentiate this group with encouraging precision. The two outlying subtilisin/chymotrypsin complexes have free energies that are 16.0 and 16.3 kcal/mol above the minimum free energy structure; the two chymotrypsin/ovomucoid outliers have free energies 7.5 and 16.3 kcal/mol above the minimum, and the three trypsin/BPTI outliers have free energies of 2.2, 2.2, and 6.6 kcal/mol above the minimum (Table 1B; Figs. 1, 2, 3). The lowest free energy structures have all-atom RMSDs of 2.2, 1.7, and 2.7 Å from the observed complex, and their free energies are 3.8, 5.1, and 3.1 kcal/mol higher. These results are obtained without correction for side-chain conformational changes.

Side-chain conformations searched

As a first approach to relaxing the rigid body approximation, we exhaustively searched those inhibitor side chains with a large fraction of their surface areas buried (Table 2) using the CONGEN algorithm (Bruccoleri & Novotny, 1992). The 10 lowest CHARMm energy structures were retained and ranked according to their free energy. For the chymotrypsin/ovomucoid system, we mutated Met 18 of ovomucoid to Leu during the side-chain search, in order to obtain a more reliable comparison with the crystal complex. The six positions of subtilisin were left unchanged because of technical difficulties with a CONGEN side-chain search when docked residues are mutated to larger size. A viable alternative would be to mutate prior to docking. However, in order to compare the results with those of Jackson and Sternberg (1995), apart from adjusting some side chains, we retain the complex conformations generated by Shoichet and Kuntz (1991).

For the subtilisin/chymotrypsin inhibitor (1SBC/2CI2) system, improvements are seen for residues Thr 58 and Met 59 (Table 2; Fig. 4). The average improvement for the RMSD of Met 59, using all but the 10.20-Å and 10.93-Å structures, is from 3.8 Å to 1.3 Å. For the chymotrypsin/ovomucoid (5CHA/2OVO) system, the RMSDs for Asn 36 degrade slightly, whereas for Tyr 20 and Arg 21, they remain about the same. Improvements can be seen for Thr 17 and Met 18 (Fig. 5). For Met 18, the RMSD before the CONGEN search is not shown because the corresponding residue in bound structure is Leu, but it is mutated back to Leu during the side-chain search and the average RMSD of Leu after the search is 1.5 Å for the low RMSD structures.

For trypsin/BPTI (2PTN/4PTI) (the only system for which the bound and free crystal structures are chemically identical), the improvement after the side-chain search is dramatic. For the low RMSD structures (1.73, 1.78, 1.85, and 2.66 Å RMSDs) the RMSD for Lys 15 decreases from an average 2.4 Å to 0.6 Å

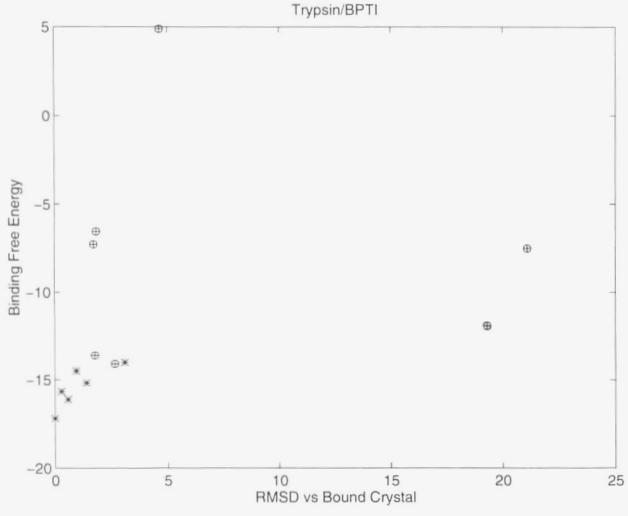

Fig. 3. Same as Figure 1, for trypsin/BPTI complex.

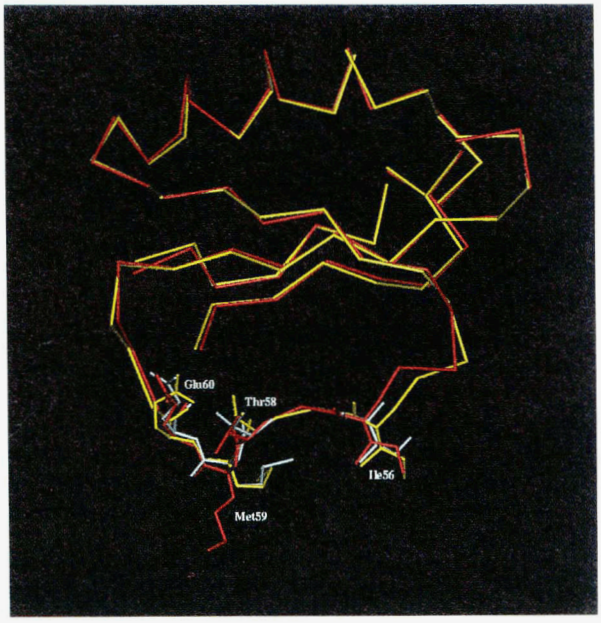

Fig. 4. Results of side-chain conformational search for residues Ile 56, Thr 58, Met 59, and Glu 60 of chymotrypsin inhibitor complexed with subtilisin. Only Cα traces and residues Ile 56, Thr 58, Met 59, Glu 60 are shown. The structures are color-coded as follows: red, lowest binding free energy structure without side-chain search; white, lowest binding free energy structure with side-chain search; yellow, inhibitor in the X-ray structure of the complex.

Table 2. *Residues subjected to side chain search*

Structure[a]	RMSD (Å)	Residue[b]
	1.13	I56(100/0.9/1.4);T58(199/1.8/0.7);M59(100/3.7/1.1);E60(59/1.5/1.4)
	1.28	I56(100/1.3/1.2);T58(96/2.0/0.7);M59(100/4.0/1.1);E60(59/1.7/1.6)
	1.32	T58(96/2.1/0.9);M59(100/4.1/1.2);E60(60/1.7/1.6)
	0.44	I56(100/1.3/1.3);T58(96/2.0/0.7);M59(100/4.2/1.0);E60(59/1.7/1.4)
1SBC	1.71	I56(100/1.1/2.4);T58(100/1.6/1.9);M59(100/3.4/1.5);E60(53/1.6/1.5)
2CI2	1.90	I56(100/1.0/1.6);T58(100/1.6/0.5);M59(100/3.5/1.4);E60(50/1.7/1.7)
	2.18	I56(100/1.1/1.6);T58(100/1.9/0.5);M59(100/3.8/1.4);E60(58/2.0/1.7)
	10.20	M59(88/5.8/5.9);E60(100/5.8/5.2);Y61(82/5.7/5.1);R81(64/7.6/7.8)
	10.93	T58(100/3.4/3.5);M59(58/5.3/5.7);Y61(71/4.3/3.8);R81(64/8.0/5.4)
	1.55	T17(93/2.1/1.2);M18(100/−/1.6);Y20(100/1.1/1.6);R21(54/2.3/2.1);N36(61/1.9/2.5)
	1.73	T17(95/2.3/1.3);M18(100/−/2.0);Y20(62/1.5/1.7);R21(52/1.9/1.8);N36(61/2.1/2.7)
	1.70	T17(93/2.0/1.4);M18(100/−/1.2);Y20(70/1.2/1.6);R21(58/2.5/2.2);N36(59/1.8/2.5)
	1.87	T17(93/2.0/1.2);M18(100/−/0.8);Y20(62/2.0/1.8);R21(54/3.2/3.1);N36(59/2.1/2.7)
5CHA	2.09	T17(92/2.1/1.4);M18(100/−/1.0);Y20(66/1.5/1.7);R21(60/3.2/2.6);N36(55/2.2/2.6)
2OVO	2.31	T17(91/2.0/1.5);M18(100/−/1.3);Y20(75/1.5/1.8);R21(57/2.2/2.3);N36(63/1.7/2.5)
	2.69	T17(92/2.1/1.6);M18(100/−/1.6);Y20(75/1.4/1.7);R21(51/2.6/3.0);N36(60/2.6/2.9)
	4.33	T17(100/1.9/0.9);M18(100/−/1.5);Y20(68/1.2/1.8);R21(71/3.4/3.5);N36(42/4.0/4.0)
	7.98	M18(100/−/2.8);K29(54/8.2/8.1);N36(63/3.6/3.9)
	18.10	T17(71/9.2/9.4);M18(69/−/14.2);E43(64/24.8/24.6);N45(100/31.9/32.0)
	1.73	K15(100/2.4/0.4);R17(73/4.3/1.6);R39(43/2.4/1.2)
	1.78	K15(100/2.2/0.6);R17(79/4.6/2.2);R39(55/2.4/1.5)
	1.85	K15(100/2.8/0.7);R17(91/4.1/2.3);R39(59/2.2/1.6)
2PTN	2.66	K15(100/2.2/0.8);R17(87/4.5/2.3);R39(78/3.7/1,4)
4PTI	4.65	R17(100/9.8/10.1)
	19.30	K26(100/3.1/2.9)[c]
	19.31	K26(100/3.1/2.9)[c]
	21.09	K26(100/2.2/2.7)[c]

[a] PDB code.

[b] Single-letter amino acid codes. The first number in the parenthesis is the percentage of the buried side-chain surface area. The reference side-chain surface areas are: I 104.8 Å^2; T81.3 Å^2; M 119.6 Å^2; E 118.5 Å^2; R 162.4 Å^2; N 91.2 Å; K 130.5 Å^2; Y 141.6 Å^2. When the buried side-chain area is greater than or equal to the reference area, the percentage is considered to be 100 in the Table. The reference side-chain areas are calculated from the average of different side-chain conformations in the crystal structures of proteins. The second and the third numbers in the parentheses are the side-chain RMSD before and after the CONGEN side-chain search.

[c] The RMSD of these three residues are compared to Lys15, as explained in the text.

(Fig. 6); for Arg 39 decreases from an average 2.7 Å to 1.4 Å; and for Arg 17 decreases from an average 4.4 Å to 2.1 Å. For the very distant structures (19.30, 19.31, 21.09 Å RMSDs), a side-chain search is not meaningful; e.g., these structures have Lys 26 placed where Lys 15 should be, in an entirely incorrect location.

The structural improvements are apparent from Figures 4, 5, and 6, and are reflected in the an improved correlation between binding free energy and RMSD (Figs. 7, 8, 9). Taken collectively, the free energies and RMSDs for the lowest free energy structures decrease after side-chain relaxation (compare Tables 1B and 3; Table 4). For the subtilisin/chymotrypsin inhibitor system, the correlation coefficient between free energy and RMSD is improved from 0.84 (Fig. 1) to 0.9 (Fig. 7). For the chymotrypsin/ovomucoid system, the correlation coefficient is improved from 0.96 (Fig. 2) to 0.97 (Fig. 8). For the trypsin/BPTI system, the free energy versus RMSD plot improves substantially with the three most distant structures having higher free energies than any of the four closest structures (compare Figs. 3 and 9). The 4.65-Å RMSD structure, which still has positive binding free energy, is actually a very distinct structure because it fits Arg 17 instead of Lys 15 in the crystal inhibitor into the S1 packet of the receptor. Without taking this structure into account, the correlation coefficient between free energy and RMSD is improved from 0.42 (Fig. 3) to 0.94 (Fig. 9).

Discussion

Docking with an ASP-based free energy target

For the protease-protein inhibitor systems considered in this paper, the rigid body docking procedure of Shoichet and Kuntz (1991) successfully limits the number of potentially acceptable complex conformations to fewer than 10, at least one of which is within 2 Å of the observed crystal structure. However, members of the final set show substantial structural variability, and are indistinguishable by the selection criteria used by Shoichet and Kuntz (1991).

The results presented here indicate that a free energy function, which is evaluated rapidly, splits the degeneracy and can clearly

Table 3. *Calculated free energies[a] for structures after CONGEN side-chain search*

Structures	RMSD[b] (Å)	E_d^{el}	ΔG_s	$-T\Delta S_c$	ΔG_{cal}
	1.13	−9.9	−28.0	16.9	−12.0
	1.28	−7.3	−25.8	15.5	−8.6
Subtilisin/chymotrypsin-inhibitor	1.32	−8.3	−25.4	15.7	−9.0
1SBC/2CI2	1.44	−6.8	−26.5	15.6	−8.7
	1.71	−9.0	−26.4	14.9	−11.5
	1.90	−7.8	−27.2	15.6	−10.4
	2.18	−10.1	−27.0	16.2	−11.9
	10.20	0.3	−28.2	21.2	2.5
	10.88	−4.7	−29.2	23.9	−1.0
	1.55	−12.2	−25.7	17.1	−12.0
	1.73	−12.0	−26.2	18.1	−11.1
Chymotrypsin/ovomucoid	1.70	−12.0	−26.1	17.5	−11.6
5CHA/2OVO	1.87	−11.6	−25.2	17.9	−10.0
	2.09	−10.6	−25.4	17.6	−9.4
	2.31	−11.4	−26.5	18.1	−10.8
	2.69	−8.0	−28.0	18.7	−8.4
	4.33	−6.9	−28.5	18.5	−8.0
	7.98	−6.0	−25.9	16.6	−6.2
	18.10	5.8	−27.5	19.1	6.4
	1.73	−17.8	−20.5	16.9	−12.4
	1.78	−22.1	−18.6	16.7	−15.0
Trypsin/BPTI 2PTN/4PTI	1.85	−22.0	−18.6	16.9	−14.6
	2.66	−20.7	−18.7	17.7	−12.8
	4.65	−8.3	−15.2	16.7	2.2
	19.30	−16.2	−23.4	21.7	−8.9
	19.31	−16.1	−23.4	21.8	−8.7
	21.00	−13.0	−22.9	19.3	−7.6

[a] In kcal/mol.
[b] RMSD of the docked inhibitor structures from the inhibitor structure in the crystal complex.

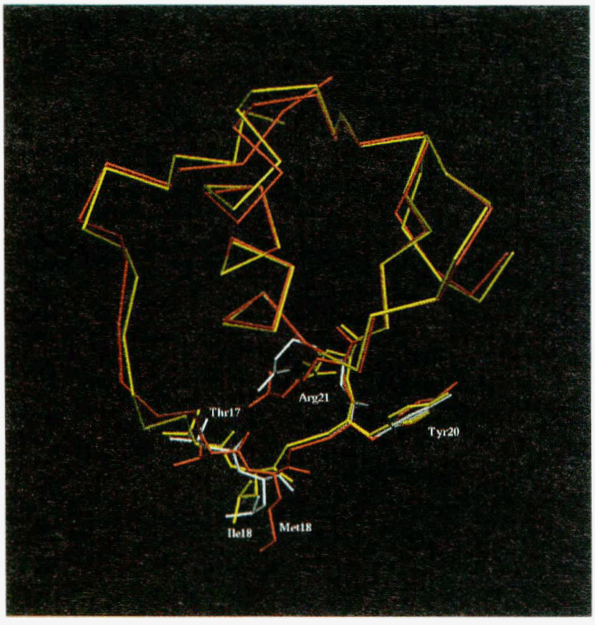

Fig. 5. Results of side-chain conformational search for residues Thr 17, Ile(Met) 18, Tyr 20, and Arg 21 of ovomucoid third domain complexed with chymotrypsin. Only Cα traces and residues Thr 17, Ile(Met) 18, Tyr 20, Arg 21 are shown. Color coding is the same as in Figure 4.

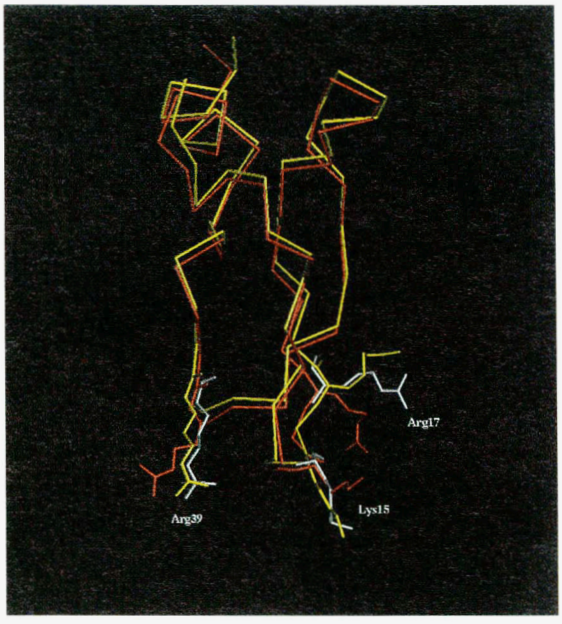

Fig. 6. Results of side-chain conformational search for residues Lys 15, Arg 17, and Ile 18 of BPTI complexed with trypsin. Only Cα traces and residues Lys 15, Arg 17, Ile 18 are shown. Color coding is the same as in Figure 4.

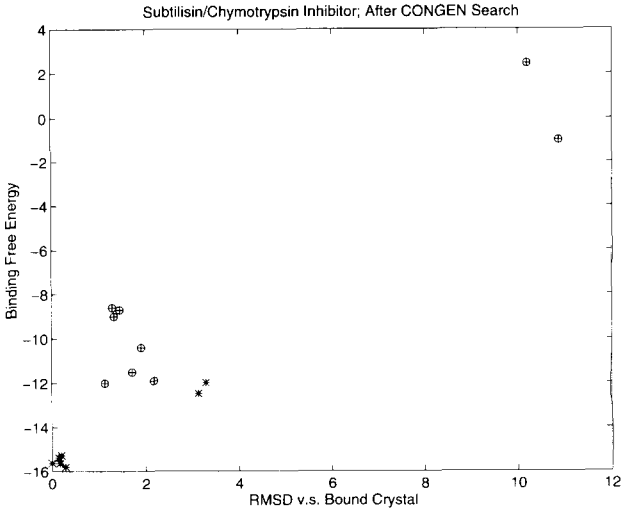

Fig. 7. Same as Figure 1 after side-chain conformational search.

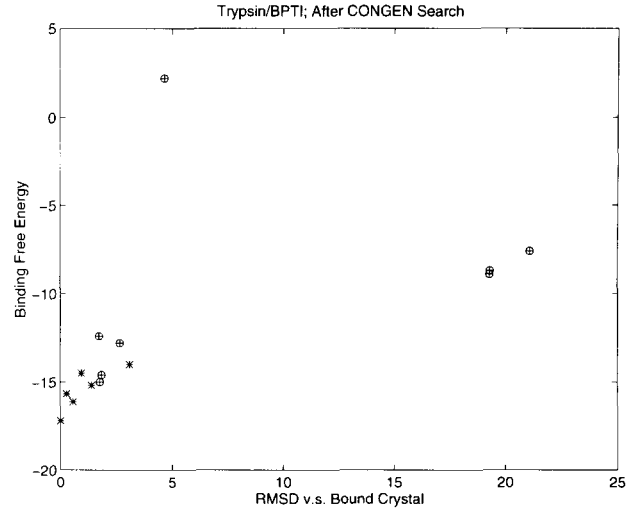

Fig. 9. Same as Figure 3 after side-chain conformational search.

identify the complexes that are reasonably close to the X-ray structure. There is an almost linear relationship between binding free energy and RMSD if the bound conformations of the monomers are used in docking. Although the use of free monomer conformations weakens the free energy–RMSD correlation, by adjusting some side-chains in the contact region, the free energy minima are attained on conformations that deviate from the crystal structures by 1.13, 1.55, and 1.78 Å for subtilisin/chymotrypsin inhibitor, chymotrypsin/ovomucoid, and trypsin/BPTI, respectively. Furthermore, the calculated free energies of the crystal structures are within 5% of the measured free energies. The latter observation indicates that, in the case of rigid proteins, the free energy function is accurate and probably does not limit the reliability of the calculation.

A further indication that the free energy function is not limiting is the relationship between the free energy of the predicted (i.e., lowest free energy) structure and the free energy of the crystal structure. For the docked bound conformers of chymotrypsin/ovomucoid, trypsin/BPTI, these are $-12.1/-14.5$ and $-16.1/-17.2$ (Table 3). The subtilisin/chymotrypsin inhibitor complex is a marginal exception: the numbers are $-15.8/-15.6$, with the lowest free energy structure being 0.3 Å from the crystal structure. Thus, the search algorithm rather than the free energy function appears to limit the accuracy of the computation.

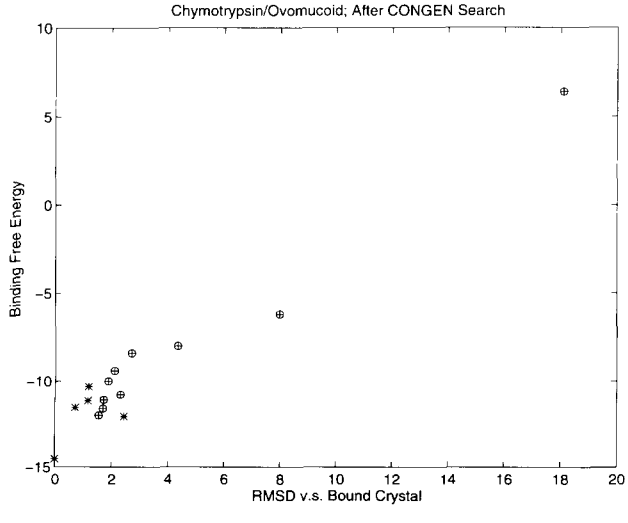

Fig. 8. Same as Figure 2 after side-chain conformational search.

The results in Table 4 are especially promising. They indicate that, given the crystal structures of uncomplexed conformers, the complex that they form or a complex formed by close homologues may be accurately predictable: in the case under consideration, the all-atom RMSDs are 1.13, 1.55, and 1.78 Å. The prescription for obtaining such results is (1) a rapid screen using the DOCK algorithm (Shoichet & Kuntz, 1991) followed by (2) discrimination using Equation 5 as a target function, and an algorithm that includes a surface side-chain conformational search.

By Equation 5, for rigid structures, the variable part of the binding free energy is the sum of the electrostatic interaction energy E_d^{rl}, the desolvation free energy ΔG_s, and the entropic term $-T\Delta S_c$ due to the conformational entropy loss of side-chains in the contact region. It is interesting to note that the first two terms have been used separately by Shoichet and Kuntz (1991) for ranking the docked complexes, but did not perform very well. In fact, without adjusting the side chains, the 7.98-Å RMSD chymotrypsin/ovomucoid structure has lower electrostatic interaction energy than the 2.69-Å RMSD structure, and in the trypsin/BPTI system, both the 19.30-Å and 19.31-Å RMSD structures have lower electrostatic interaction energies than the 1.73-Å and 1.85-Å structures. Hydrophobicity alone is even less suitable for ranking the complexes. Thus, the trade off between desolvation free energy, electrostatic interaction energy, and the side-chain entropy loss is very important to obtain an estimate of the free energy that can distinguish the correct docked conformations from the incorrect ones.

We note that, prior to calculating the energy, only constrained energy minimization is performed. The procedure removes steric clashes and severe deviations from standard geometry, but does not induce substantial structural changes. It is important,

Table 4. *Lowest free energy structures compared to crystal complexes*

Complexes	Crystal free energy[a]	Rigid body docking[b]		Side-chain relaxation[c]	
		Free energy	RMSD	Free energy	RMSD
1SBC/2CI2	−15.6	−11.8	2.18	−12.0	1.13
5CHA/2OVO	−14.5	−9.4	1.73	−12.0	1.55
2PTN/4PTI	−17.2	−14.1	2.66	−15.0	1.78

[a] In kcal/mol; see Table 1A.
[b] From the unbound conformations of monomers; see Table 1B.
[c] See Table 3.

because a deep unconstrained minimization using the CHARMm potential usually gives unrealistically negative electrostatic energies; a direct consequence of the inherent approximation in decoupling the electrostatic interaction calculation from the desolvation process. With limited minimization, the electrostatic energy on its own provides a better ranking than found by Shoichet and Kuntz (1991).

ASP and EFE approaches

Both ASP and EFE models are self-consistent macroscopic solvent models which integrate a molecular mechanics treatment of the solute with a continuum treatment of the solvent, and both rank the docked conformations generated by Shoichet and Kuntz (1991), in good agreement with the RMSD from the X-ray structure of the complexes. In the ASP model, the electrostatic interaction is calculated according to Coulomb's law using a distance-dependent dielectric constant ($\epsilon = 4r$), and the self-energy part of the electrostatic component of solvation free energy is accounted for by ASPs for different atom types. In the EFE model, the electrostatic interaction between ligand and receptor and the electrostatic contributions to desolvation free energies are all obtained by solving Poisson's equation, whereas the favorable entropic component of desolvation is assumed to be proportional to the change in the solvated surface area. The dielectric constant of protein is chosen to be 2, in order to account for electronic polarizability, and solvation screening is taken into account explicitly in solving Poisson's equation for the electrostatic contribution to solvation free energy.

Both methods assume that the change in van der Waals interactions upon binding is relatively small (i.e., both solute-solvent and solute-solute interfaces are well packed). However, Jackson and Sternberg (1995) assume that the van der Waals interaction between receptor and ligand is more favorable than the van der Waals interaction of the proteins with water, and that the difference is offset by the increase in conformational entropy of the side-chains. Consequently, neither the side-chain entropy loss nor the van der Waals interaction term is included in the free energy function.

In a different implementation of the EFE method, Smith and Honig (1994) calculate protein-protein van der Waals interactions explicitly, and use vapor to water transfer free energy coefficients to estimate the protein-solvent van der Waals interactions. Although formally complete, the van der Waals energy has a much larger scale than the solvation term, and even small errors will cause serious inaccuracies. The first-order approximation of van der Waals cancellation is therefore necessary to avoid masking other energy components that control binding.

The ASP method assumes that the van der Waals interaction between receptor and ligand is exactly counterbalanced by the van der Waals interaction with water, therefore the van der Waals contribution to the binding free energy is simply zero (Novotny et al., 1989; Vajda et al., 1994). This assumption is usually valid for the well-packed crystal complexes, as demonstrated by the accurate evaluation of the binding free energy of nine crystal complexes in our previous study (Vajda et al., 1994), but needs to be taken with caution when it is applied to predicted structures.

It is interesting to compare the various contributions to the free energy as calculated by the two methods. Although the total binding free energies correlate poorly (discussed below), the electrostatic interaction energies have a correlation coefficient of 0.77 (Fig. 10). The electrostatic energies from EFE model after close contact correction are used (Jackson & Sternberg, 1995). The close contact correction sets the electrostatic interaction energy between a pair of atoms to a prescribed value whenever these two atoms overlap. The prescribed value is calculated by Coulomb's law with the distance between these two atoms set to the sum of their van der Waals radii (Jackson &

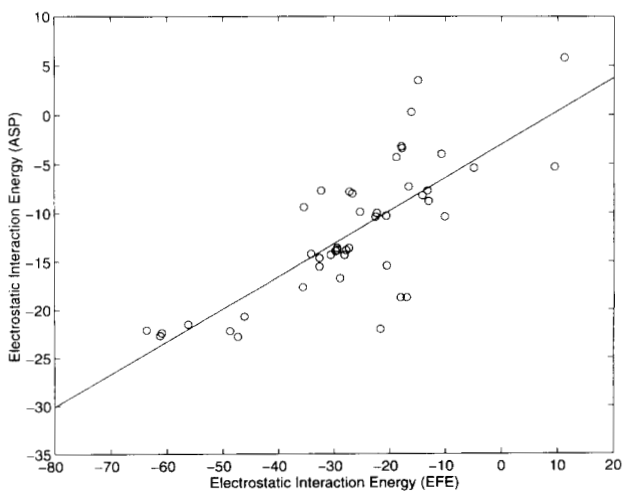

Fig. 10. Electrostatic interaction energies calculated using the ASP method plotted against those calculated using the EFE method. All docked conformations are included. Correlation coefficient is 0.77.

Sternberg, 1995). If only the bound complexes are considered, the correlation between the electrostatic interaction energies given by the two methods is 0.96 (Fig. 11). The reason for this strong correlation is that the charges in the complexes are too far from solvent to establish an induced attenuation field. As a result, the solution to the Poisson's equation will yield a simple Coulomb interaction, which is what is used in the ASP model. The difference in magnitude is the result of a different choice of dielectric, which does not, of course, affect the correlation.

Free energy values by the EFE model

Although both EFE and ASP models give satisfactory results for ranking candidate docked structures, their success in estimating the observed free energies of the complexes differs substantially. For the EFE method, only the chymotrypsin/ovomucoid range (between −10 and −18 kcal/mol) is accurate; the measured free energy of subtilisin/chymotrypsin inhibitor (−15.8 kcal/mol) is 3.8 kcal/mol from the center of an 8 kcal/mol range (between −8 and −16 kcal/mol), and the measured binding free energy of trypsin/BPTI (−18.1) is well outside the predicted range (between +4 and −4 kcal/mol) (Jackson & Sternberg, 1995). As a consequence, the predicted rank ordering of the binding affinities for these three complexes is chymotrypsin/ovomucoid > subtilisin/chymotrypsin inhibitor > trypsin/BPTI, whereas the experimental ranking is the exact reverse. Some insight into these results, and the parameters on which they depend, can be obtained by considering simple analytical models of the dependence of solvation energy on the charge distributions in these three systems.

Analysis of the the binding loops of the three inhibitors (Table 5, positions P4–P4′ as defined in Hubbard et al. [1991]) indicates that chymotrypsin inhibitor (CI) and ovomucoid third domain (OVO) have a negatively charged Glu at P1′ and a positively charged Arg at P3′, whereas BPTI has two positively charged residues: Lys at P1 and Arg at P2′. Because the distance between charges, as well as the distance between the charges and the surfaces, is small compared to the radius of curvature of the proteins, we consider a model in which charges are buried, in a dielectric ϵ_p, slightly below the surface (at a distance d) of an infinite plane, outside of which is a dielectric medium, ϵ_w.

First consider two surfaces far apart, with equal and opposite charges $+q$ and $-q$ (Fig. 12A). In the bound state, the two semi-infinite dielectrics (ϵ_p) are immediately adjacent to one another, and thus create a homogeneous dielectric environment. For reasons that will become apparent, we are interested in the most favorable energy change and, to this end, we take the final distance between the two charges as $2d$. With q in electrostatic units, and d in Å, the desolvation energy E_s, ($\Delta E_i^r + \Delta E_i^l$ in Equation 3), in kcal/mol is (see for example, Jackson [1962]):

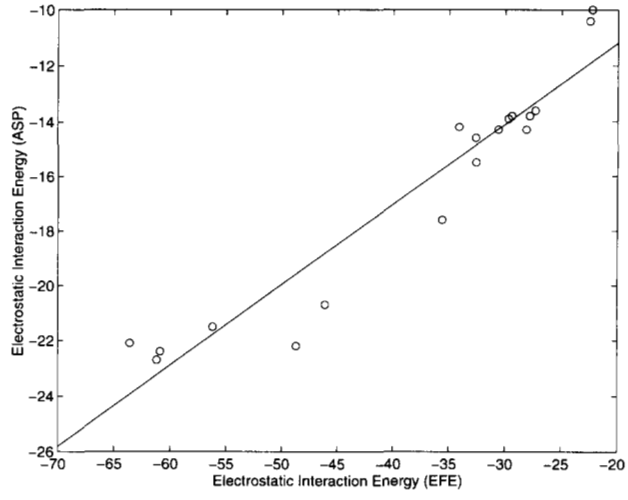

Fig. 11. Electrostatic interaction energies calculated using the ASP method plotted against those calculated using the EFE method. Only the docked conformations using bound conformers are included. Correlation coefficient is 0.96.

$$E_s = -332 \times \frac{qq'}{2d\epsilon_p}, \tag{7}$$

where q' is the image charge of q:

$$q' = -\frac{\epsilon_w - \epsilon_p}{\epsilon_w + \epsilon_p} q. \tag{8}$$

The Coulombic interaction energy E_d, which corresponds to E_d^{rl} in Equation 3, is:

$$E_d = -332 \times \frac{q^2}{2d\epsilon_p}. \tag{9}$$

The sum of E_s and E_d gives us the electrostatic contribution to the binding free energy (E):

$$E = -332 \times \frac{q^2}{2d\,\frac{\epsilon_w + \epsilon_p}{2}}. \tag{10}$$

Table 5. *Amino acid compositions of the binding loops of inhibitors*[a]

Complexes	P4	P3	P2	P1	P1′	P2′	P3′	P4′
CI	Ile 56	Val 57	Thr 58	Met 59	Glu 60	Tyr 61	Arg 62	Ile 63
OVO	Ala 15	Cys 16	Thr 17	Leu 18	Flu 19	Tyr 20	Arg 21	Pro 22
PTI	Gly 12	Pro 13	Cys 14	Lys 15	Ala 16	Arg 17	Ile 18	Ile 19

[a] Residue numbers are shown after each residue name.

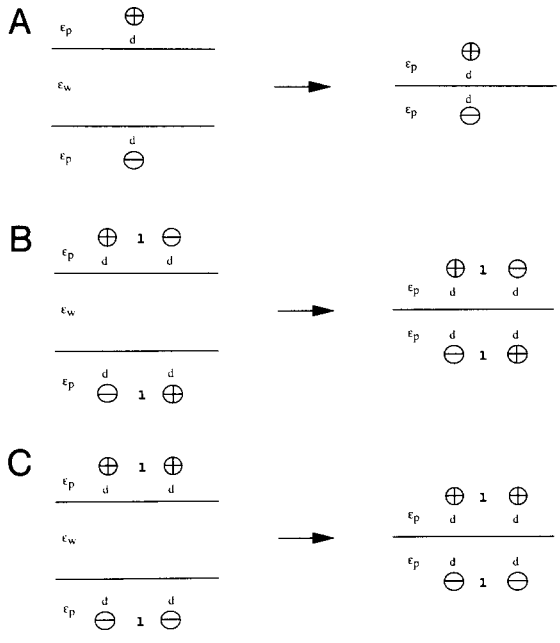

Fig. 12. Electrostatic contributions to the binding free energies of charges embedded in two semi-infinite dielectrics (ϵ_p) separated by a high dielectric (ϵ_w). **A:** One salt bridge. **B:** Two salt bridges with like charges at the same side. **C:** Two salt bridges with opposite charges on the same side.

With $q = 1$, $d = 1.5$ Å, $\epsilon_p = 2$, and $\epsilon_w = 80$, $E_{bind} = -2.7$ kcal/mol. This indicates that the unfavorable desolvation energy is barely compensated for, even under the best circumstances, by a favorable Coulombic interaction. Now consider buried charge pairs (Fig. 12B), each consisting of two equal but opposite charges separated by a distance l. This is a simplified representation of the two complexes SNI/CI and CHO/OVO. In this case, E_s, E_d, and E are:

$$E_s = -332 \times 2 \times \frac{qq'}{2d\epsilon_p} + 332 \times 2 \times \frac{qq'}{\sqrt{l^2 + 4d^2\epsilon_p}}, \quad (11)$$

$$E_d = -332 \times 2 \times \frac{q^2}{2d\epsilon_p} + 332 \times 2 \times \frac{q^2}{\sqrt{l^2 + 4d^2\epsilon_p}}, \quad (12)$$

$$E = -332 \times 2 \times \frac{q^2}{2d\dfrac{\epsilon_w + \epsilon_p}{2}} + 332 \times 2 \times \frac{q^2}{\sqrt{l^2 + 4d^2}\dfrac{\epsilon_w + \epsilon_p}{2}}. \quad (13)$$

The first term on the right in Equation 11 reflects interactions between the charges and their respective images, whereas the second is the cross term stemming from interactions between a charge and the image of its partner. This interference term is clearly unfavorable when the charges on the same side of the plane are of opposite sign, because a positive (negative) charge will interact with the positive (negative) image of a negative (positive) charge.

In Equation 12, the first term represents interactions between a charge and its opposite on the opposing protein (a salt bridge), whereas the second term represents cross interactions. The equations assume that the self energy remains the same after complex formation, and hence cancels.

Finally, consider the situation in which each member of the pair has the same charge (Fig. 12C). This is a simplified representation of 2PTC/PTI complex. It is easy to see that the solution can be obtained from Equations 11, 12, 13 simply by switching the signs of the second terms:

$$E_s = -332 \times 2 \times \frac{qq'}{2d\epsilon_p} - 332 \times 2 \times \frac{qq'}{\sqrt{l^2 + 4d^2\epsilon_p}}, \quad (14)$$

$$E_d = -332 \times 2 \times \frac{q^2}{2d\epsilon_p} - 332 \times 2 \times \frac{q^2}{\sqrt{l^2 + 4d^2\epsilon_p}}, \quad (15)$$

$$E = -332 \times 2 \times \frac{q^2}{2d\dfrac{\epsilon_w + \epsilon_p}{2}} - 332 \times 2 \times \frac{qq'}{\sqrt{l^2 + 4d^2}\dfrac{\epsilon_w + \epsilon_p}{2}}. \quad (16)$$

In the above three examples, the salt bridges are perfectly aligned. E_s and E_d largely cancel each other, and E is almost independent of ϵ_p. In actuality, an alignment for salt bridge formation does not occur readily, and therefore the favorable electrostatic compensation for an unfavorable desolvation energy is somewhat below ideal. In fact, Zachary and Tidor's study (1994) shows that most salt bridges cannot compensate for the desolvation energy, and that E_d is frequently only half of the magnitude of E_s. This is in accord with the calculation of the three protease-inhibitor complexes in the Jackson and Sternberg study (1995).

Under the imperfect salt bridge condition, E_s dominates (e.g., d in the equations for E_d, would be replaced by something larger to take account of nonoptimal alignment). From Equations 7, 11, and 14, it is clear that E_s is inversely proportional to ϵ_p, and the solvation energy ($-E_s$) of a pair of like charges is more favorable than the solvation energy of a pair of opposite charges with similar separations. Hence, like charges will tend to have less favorable binding energy, and the extent to which it is less favorable will be directly dependent on the magnitude of ϵ_p.

As an example, with $l = 10$ Å and $d = 1.5$ Å, the magnitude of E_s in Equation 14 decreases approximately 35 kcal/mol as ϵ_p increases from 2 to 4. Using a dielectric of 2 rather than 4 may in part account for the mispredicted 2PTC/PTI binding free energy by EFE model. This also shows the general uncertainty of EFE method to the change of ϵ_p.

This discussion is not meant to be quantitative, but to illustrate the sensitivity of the predictions to assumptions about dielectric and geometries. The first is fundamental and will undoubtedly require more detailed models that take account of polarization and thereby treat the protein's interior explicitly, rather than by an effective dielectric constant of an assumed magnitude. The second, viz use of correct geometry, is itself cou-

pled to energy, and therefore makes the computation substantially more demanding. There are also a number of less difficult but still important problems related to the solution of the Poisson-Boltzmann equation using finite difference methods. These include the sensitivity of the results to the selection of the boundary separating the protein (ϵ_p) and water (ϵ_w) (J. Novotny, pers. comm.), and the position of the protein relative to the grid (Bruccoleri, 1993). We believe that the solution to Poisson's equation will ultimately prove to be the method of choice for obtaining desolvation energies, and that such solutions, coupled with more accurate procedures for evaluating desolvation entropy, will provide reliable and computationally realistic free energy evaluation methods. However, these methods are likely to be computationally intensive and, at the present time, unsuitable to be used as target functions for docking. Therefore, the empirical procedures used here represent viable and useful alternatives.

Methods

Free energy evaluation using ASP method

The electrostatic interaction energy between receptor and the ligand, E_d^{el} (Equation 5), is calculated with the CHARMm force field (Molecular Simulations, Inc.) using a distance-dependent dielectric coefficient $\epsilon = 4r$ and a nonbonded cutoff of 15 Å.

The desolvation free energy ΔG_s is calculated by Equation 6, and the transfer free energies of the complex (ΔG_{tr}^{rl}), the receptor (ΔG_{tr}^{r}), and ligand (ΔG_{tr}^{l}) are obtained by a linear combination of the solvent-accessible areas weighted by ASPs.

The side-chain entropy loss $T\Delta S_c$ is assumed to be proportional to the loss in solvent-accessible surface (Pickett & Sternberg, 1993), as described in (Vajda et al., 1994).

A basic assumption underlying the free energy function is that the sum of solvent-solute and solute-solute van der Waals energies is the same for all complexes. Briefly, we evaluate the free energies on a set of complexes free of van der Waals clashes. Clashes are removed using a 200-step adopted basis Newton-Raphson (ABNR) minimization, with parameters taken from Version 19 of the CHARMm potential (Molecular Simulations, Inc.). Mass-weighted harmonic constraints with a force constant of 20 kcal/(mol Å^2) are applied to all atoms during the minimization.

Side-chain conformation search

The rigid body assumption was relaxed for buried interfacial inhibitor side chains. Because the side chains remain buried during the search procedure, we assume the entropic and hydrophobic parts of the free energy function are relatively invariant, with the energy term dominating. Using CONGEN algorithm (Bruccoleri & Novotny, 1992), the CHARMm energy function was minimized by an exhaustive search over all side-chain dihedral angles with the grid spacing set to 20°. The van der Waals cutoff is set to be 30 kcal/mol, and the conformations that have van der Waals energies higher than 30 kcal/mol are eliminated. The search for side-chain positions begins by generating a small number of the best conformations for each side chain independently, then these conformations are assembled in all possible combinations and those combinations that do not have bad van der Waals contacts are accepted. Among the retained conformations, the 10 lowest total energy ones are used for the free energy evaluation.

In order to remove any van der Waals clashes that can possibly be introduced by the side-chain search, the free energies are calculated for the CONGEN output complexes after minimizing van der Waals energy for 20 steps and minimizing total CHARMm energy for another 20 steps.

Acknowledgments

We thank Brian Shoichet for providing the coordinates of docked complexes, Jiri Novotny for helpful discussions, and Robert Bruccoleri for providing the program CONGEN. We also thank Chao Zhang for useful discussions about electrostatics. This work was supported by grant AI30535 from the NIH/NIAID (C.D., Z.W.), and by grant DOE-FG02-93ER61656 from the Department of Energy (S.V.).

References

Abagyan R, Totrov M. 1994. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J Mol Biol 235*:983–1002.

Adamson A. 1976. *Physical chemistry of surfaces*. New York: Wiley.

Bacon DJ, Moult J. 1992. Docking by least-squares fitting of molecular surface patterns. *J Mol Biol 225*:849–858.

Bigler TL, Lu W, Park SJ, Tashiro M, Wieczorek M, Wynn R, Laskowski MJ. 1993. Binding of amino acid side chains to preformed cavities: Interaction of serine proteinases with turkey ovomucoid third domains with coded and noncoded P1 residues. *Protein Sci 2*:786–799.

Bruccoleri RE. 1993. Grid positioning independence and the reduction of self-energy in the solution of the Poisson-Boltzmann equation. *J Comput Chem 14*:1417–1422.

Bruccoleri RE, Novotny J. 1992. Antibody modeling using the conformational search program CONGEN. *Immunomethods 1*:96–106.

Cherfils J, Duquerroy S, Janin J. 1991. Protein-protein recognition analyzed by docking simulation. *Proteins Struct Funct Genet 11*:271–280.

Chothia C, Janin J. 1975. Principles of protein-protein recognition. *Nature (Lond) 256*:705–708.

Colonna-Cesari F, Sander C. 1990. Excluded volume approximation to protein-solvent interaction. The solvent contact model. *Biophys J 57*: 1103–1107.

Cramer CJ, Truhlar DG. 1992. AM1-SM2 and PM3-SM3 parameterized SCF solvation models for free energies in aqueous solution. *J Comput-Aided Mol Design 6*:629–666.

Eisenberg D, Mclachlan AD. 1986. Solvation energy in protein folding and binding. *Nature (Lond) 319*:199–203.

Erickson HP. 1989. Co-operativity in protein-protein association: The structure and stability of the actin filament. *J Mol Biol 206*:465–474.

Finkelstein AV, Janin J. 1989. The price of lost freedom: Entropy of biomolecular complex formation. *Protein Eng 3*:1–3.

Helmer-Citterich M, Tramontano A. 1994. PUZZLE: A new method for automated protein docking based on surface shape complementarity. *J Mol Biol 235*:1021–1031.

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

Hubbard SJ, Campbell SF, Thornton JM. 1991. Molecular recognition: Conformational analysis of limited proteolytic sites and serine protease protein inhibitors. *J Mol Biol 220*:507–530.

Jackson JD. 1962. *Classical electrodynamics*. New York: John Wiley and Sons, Inc.

Jackson RM, Sternberg MJE. 1995. A continuum model for protein-protein interactions: Application to the docking problem. *J Mol Biol 250*:258–275.

Jiang F, Kim SH. 1991. Soft docking: Matching of molecular surface cubes. *J Mol Biol 219*:79–102.

Judson RS, Tan YT, Mori E, Melius C, Jaeger EP, Treasurywala AM, Mathiowetz A. 1995. Docking flexible molecules: A case study of three proteins. *J Comput Chem 16*:1405–1419.

Krystek S, Stouch T, Novotny J. 1993. Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J Mol Biol 234*:661–679.

Murphy KP, Xie D, Garcia KC, Amzel LM, Freire E. 1993. Structural energetics of peptide recognition – Angiotensin II antibody binding. *FEBS Lett 15*:113–120.

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

Novotny J, Bruccoleri RE, Saul FA. 1989. On the attribution of binding energy in the antigen-antibody complexes McPc 603, D1.3 and HyHel-5. *Biochemistry 28*:4735–4749.

Ooi T, Oobatake M, Nemethy G, Scheraga HA. 1987. Accessible surface areas as a measure of the thermodynamic parameters of hydration of peptides. *Proc Natl Acad Sci USA 84*:3086–3090.

Page MI, Jenks WP. 1971. Entropic contribution to rate acceleration in enzymic and intramolecular reactions and the chelate effect. *Proc Natl Acad Sci USA 68*:1678–1683.

Pickett SD, Sternberg MJE. 1993. Empirical scale of side-chain conformational entropy in protein folding. *J Mol Biol 231*:825–839.

Shoichet BK, Kuntz ID. 1991. Protein docking and complementarity. *J Mol Biol 221*:327–346.

Smith K, Honig B. 1994. Evaluation of the conformational free energies of loops on proteins. *Proteins Struct Funct Genet 18*:119–132.

Stoddard BL, Koshland DE. 1992. Prediction of the structure of a receptor protein using a binary docking method. *Nature (Lond) 358*:774–776.

Stouten PFW, Frommel C, Nakamura H, Sander C. 1993. An effective solvation term based on atomic occupancies for use in protein simulations. *Molecular Simulation 10*:97–120.

Svendsen I, Jonassen I, Hejgaard J, Boisen S. 1980. Amino acid sequence homology between a serine protease inhibitor from barley hordeum vulgare cultivar hiproly and potato inhibitor I. *Carlsberg Res Commun 45*:389–395.

Tanford C, Kirkwood JG. 1957. Theory of protein titration curves. I. General equations for impenetrable spheres. *J Am Chem Soc 79*:5333–5339.

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

Vajda S, Weng Z, DeLisi C. 1995. Extracting hydrophobicity parameters from solute partition and protein mutation/unfolding experiments. *Protein Eng*. Forthcoming.

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

Vincent JP, Lazdunski M. 1972. Trypsin-pancreatic inhibitor association. Dynamics of the interaction and role of disulfide bridges. *Biochemistry 11*:2967–2977.

Walls PH, Sternberg MJE. 1992. New algorithm to model protein-protein recognition based on surface complementarity. Applications to antibody-antigen docking. *J Mol Biol 228*:277–297.

Wesson L, Eisenberg D. 1992. Atomic solvation parameters applied to molecular dynamics of proteins in solution. *Protein Sci 1*:227–235.

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

Zachary HS, Tidor B. 1994. Do salt bridges stabilize proteins? A continuum electrostatic analysis. *Protein Sci 3*:211–226.