How Optimal Are the Binding Energetics of Barnase and Barstar?

Ting Wang,* Sanja Tomic,† Razif R. Gabdoulline,* and Rebecca C. Wade*
*Molecular and Cellular Modeling Group, EML Research, 69118 Heidelberg, Germany; and †Ruder Boskovic Institute, HR-10001 Zagreb, Croatia

ABSTRACT The extracellular ribonuclease barnase and its intracellular inhibitor barstar bind fast and with high affinity. Although extensive experimental and theoretical studies have been carried out on this system, it is unclear what the relative importance of different contributions to the high affinity is and whether binding can be improved through point mutations. In this work, we first applied Poisson-Boltzmann electrostatic calculations to 65 barnase-barstar complexes with mutations in both barnase and barstar. The continuum electrostatic calculations with a van der Waals surface dielectric boundary definition result in the electrostatic interaction free energy providing the dominant contribution favoring barnase-barstar binding. The results show that the computed electrostatic binding free energy can be improved through mutations at W44/barstar and E73/barnase. Furthermore, the determinants of binding affinity were quantified by applying COMparative BINding Energy (COMBINE) analysis to derive quantitative structure-activity relationships (QSARs) for the 65 complexes. The COMBINE QSAR model highlights ~20 interfacial residue pairs as responsible for most of the differences in binding affinity between the mutant complexes, mainly due to electrostatic interactions. Based on the COMBINE model, together with Brownian dynamics simulations to compute diffusional association rate constants, several mutants were designed to have higher binding affinities than the wild-type proteins.

INTRODUCTION

The binding energetics of a protein-protein complex are governed by several different factors including electrostatic and hydrophobic interactions between the proteins and solvent-protein interactions. Individually, these factors can favor or disfavor binding; the binding affinity is determined by the net effect. Stronger binding can be achieved by balancing and optimizing the individual energy terms through protein engineering. One of the recent advances is the design of a high-affinity variant of human growth hormone containing 15 mutations and binding to the human growth hormone receptor ~400-fold tighter than the wild-type protein (Pal et al., 2003). But improving the affinity of barnase and barstar is more challenging because the wild-type proteins already bind very tightly.

The extracellular ribonuclease of *Bacillus amyloliquefaciens,* barnase, and its intracellular inhibitor barstar bind fast ($k_{on} \sim 10^8$ M^{-1} s^{-1}) and with high affinity ($k_d \sim 10^{-14}$M). The binding interface consists mainly of polar and charged residues, and contains a number of buried water molecules. It shows high electrostatic complementarity, and the electrostatic attraction between the charged and polar residues of barnase and barstar acts to stabilize the bound complex. On the other hand, the desolvation cost for these residues when the proteins bind destabilizes the bound complex. Prior theoretical studies (Chong et al., 1998; Dong et al., 2003; Lee and Tidor, 2001a,b; Sheinerman and Honig, 2002) on the net contribution of electrostatics to the thermodynamics of binding have shown contradictory results. Tidor and co-workers (Chong et al., 1998; Lee and Tidor, 2001a,b), using a continuum solvent electrostatic model, reported an unfavorable computed electrostatic binding free energy of +14 kcal/mol (Lee and Tidor, 2001b) even though they found that barstar has an electrostatically optimized charge distribution for tight binding to barnase. Continuum solvent electrostatic calculations by Sheinerman and Honig (2002) showed, on the other hand, that the electrostatic attraction and the charge desolvation cost almost cancel each other and result in a net contribution to binding affinity close to zero. Very recently, Zhou and co-workers (Dong et al., 2003) computed favorable electrostatic contributions to binding affinity and pointed out that the electrostatic contribution to barnase-barstar binding strongly depends on the dielectric treatment in the calculations.

Another notable characteristic of barnase-barstar binding is the highly cooperative interactions between some interfacial residues observed in double mutant cycle experiments (Schreiber and Fersht, 1995) that result in nonadditivity of the contributions of the interfacial residues to the binding affinity. This cooperativity enhances the difficulty of prediction of the effects of mutation on binding affinity. Therefore, there is a need for models that quantitatively interpret the correlation between mutation and binding free energy, and which are useful in guiding the design of proteins to alter their binding affinity.

A set of 32 mutant barnase-barstar complexes has been analyzed by Covell and Wallqvist (1997) by using a model of binding free energy that is based on pairwise surface preferences. The effects of mutations were predicted within an error margin of 1.5 kcal/mol and it was found that interfacial water molecules contributed 25% of the binding free energy.

Submitted January 30, 2004, and accepted for publication June 14, 2004.

Address reprint requests to Rebecca C. Wade, Tel.: 49-6221-533-247; Fax: 49-6221-533-298; Email: rebecca.wade@eml-r.villa-bosch.de.

0006-3495/04/09/1618/13 $2.00

doi: 10.1529/biophysj.104.040964

Recently, Kortemme and Baker (2002) calculated the effects of 14 single mutations on the binding free energy of barnase and barstar by using a model based on an all-atom rotamer description of the side chains with an energy function dominated by Lennard-Jones interactions, solvation interactions, and hydrogen bonding. The effects on binding of some mutated residues involved in water-mediated hydrogen bonds in the interface were underpredicted.

In this work, we studied the wild-type barnase-barstar complex and 64 mutant complexes by using a Poisson-Boltzmann continuum model for electrostatics calculations (Madura et al., 1995) and performing COMparative BINding Energy (COMBINE) analysis (Ortiz et al., 1995; Wade, 2001) to derive a system-specific quantitative structure-activity relationship (QSAR) model for estimating overall binding free-energy differences (electrostatic and nonelectrostatic). The aim was, through studying a large set of mutants of the barnase-barstar system by complementary theoretical methods, to estimate the relative importance of different contributions to the binding affinity of barnase and barstar, and to assess how optimal these proteins are for binding by using the models to investigate whether mutants could be designed to bind with higher affinity than the wild-type proteins.

COMBINE analysis is based upon the premise that the binding free energy (ΔG) can be correlated with a subset of suitably weighted energy components determined from the structures of the receptor(s) and ligands in bound and unbound forms. In this study, the energy terms computed are the electrostatic desolvation energies of barnase (bn) and barstar (bs) upon binding, $\Delta G_{ele}^{desol\_bn}$ and $\Delta G_{ele}^{desol\_bs}$, respectively, and the pairwise electrostatic, E_i^{ele}, and Lennard-Jones, E_i^{vdw}, interaction energies between each barnase and each barstar residue in energy-minimized structures of barnase-barstar complexes (see Methods section for details). The binding free energy, ΔG, is estimated as a weighted linear sum of these energy terms as given in Eq. 1:

$$\Delta G = w_{bn}^{desol}\Delta G_{ele}^{desol\_bn} + w_{bs}^{desol}\Delta G_{ele}^{desol\_bs} + \sum_i w_i^{vdw}E_i^{vdw} + \sum_i w_i^{ele}E_i^{ele} + C. \quad (1)$$

The contribution of each interaction energy term is represented by its weight, namely the parameter w_{bn}^{desol}, w_{bs}^{desol}, w_i^{vdw}, or w_i^{ele} in Eq. 1. The weights are obtained by partial least-squares (PLS) analysis using a training set of complexes with experimentally determined binding affinities. The COMBINE analysis method has proved successful for deriving high quality QSAR models for a variety of protein-ligand complexes including enzyme-inhibitor (Ortiz et al., 1997, 1995; Pastor et al., 2000; Perez et al., 1998b; Wang and Wade, 2001), enzyme-substrate (Lozano et al., 2000; Tomic and Kojic-Prodic, 2002), protein-peptide (Wang and Wade, 2002), and nuclear receptor-DNA complexes (Tomic et al., 2000). This study is the first application of COMBINE analysis to protein-protein complexes.

The procedure and results are summarized here as follows. First, the structures of 64 barnase-barstar complexes with different interfacial mutations were modeled and energy minimized using the structure of the wild-type protein complex as the template. See Table 1 for the 65 complexes and their experimental binding free-energy values, which were taken from two references (Frisch et al., 1997; Schreiber and Fersht, 1995). In this article, each complex is designated by the mutation in barstar (bs) followed by the mutation in barnase (bn). Fig. 1 shows the binding interface and the location of the mutated residues (six residues in barnase and eight residues in barstar).

The electrostatic contributions to binding in all 65 complexes were computed by solving the finite difference Poisson-Boltzmann equation (Madura et al., 1995). In these continuum electrostatic calculations, with a van der Waals surface dielectric boundary definition, we find that electrostatic interactions are the dominant contribution favoring barnase-barstar binding. On the other hand, we find that wild-type barnase and barstar are not fully electrostatically optimized at the binding interface.

Then the molecular mechanics interaction energies between barnase and barstar were decomposed on a per residue pair basis and subjected to two chemometric analyses: a principal component analysis (PCA) to investigate the distribution of the 65 complexes in the energy space, and PLS analysis to derive Eq.1. The PCA analysis highlights three barnase residues and three barstar residues for which mutations have substantial effects on the energetics of barnase-barstar binding. The PLS analysis indicates that the overall effects of interfacial mutations can be quantitatively represented by the interaction energies between 16 barnase and 11 barstar residues and the electrostatic desolvation energies of barnase and barstar upon binding.

Based on the chemometric analysis results, some mutants were designed to optimize binding affinity and their binding free energies were predicted. The association rates of the designed mutants were calculated by using Brownian dynamics simulations (Gabdoulline and Wade, 2001). The results provide mutants of barnase and barstar that are predicted to have higher binding affinities than the wild-type proteins due to both increased association rates and decreased dissociation rates.

MATERIALS AND METHODS

Molecular mechanics modeling

Preparation of mutant complexes

The crystallographic structure of the pseudo wild-type barnase-barstar complex(Vaughan et al., 1999) (Protein Data Bank code 1b27) was used as the template for preparing most mutant complexes. The complex formed by chains A and D, including 213 surrounding bound water molecules, was extracted. Chain A was used for modeling barnase and chain D was used for modeling barstar. A40 and A82 in barstar were mutated to cysteines to

TABLE 1 Energetics of the wild-type barnase-barstar complex and 64 mutant complexes

No.	Complex (bs:bn)*	ΔG_{exp} (kcal/mol)	$\Delta G_{ele}^{desol\_bn}$ (kcal/mol)	$\Delta G_{ele}^{desol\_bs}$ (kcal/mol)	E_{ele}^{bn-bs} (kcal/mol)	$\Delta G_{ele}^{binding}$ (kcal/mol)	ΔG_{pred} (kcal/mol)
1	WTWT	−19.0	28.18	25.26	−75.67	−22.22	−18.35
2	WTK27Awm	−13.6	25.59	24.61	−67.31	−17.11	−14.15
3	WTR59A	−13.8	23.68	22.66	−61.22	−14.89	−14.49
4	D39AR59A	−7.7	15.57	15.13	−35.04	−4.35	−9.28
5	WTR87A	−13.5	23.61	23.45	−62.73	−15.66	−14.22
6	WTH102Awm	−12.9	23.33	21.55	−64.37	−19.48	−13.66
7	Y29FWTwm	−19.1	28.33	22.61	−72.53	−21.59	−17.60
8	Y29AWT	−15.6	25.81	21.63	−68.16	−20.71	−16.11
9	D35AWTwm	−14.5	27.41	18.75	−64.82	−18.65	−13.75
10	W38FWT	−17.4	26.88	25.13	−73.77	−21.77	−16.97
11	D39AWT	−11.3	20.66	17.23	−49.38	−11.48	−13.87
12	T42AWT	−17.2	26.46	24.3	−72.70	−21.94	−16.62
13	W44FWT	−19.0	27.32	24.33	−74.43	−22.78	−18.23
14	E76AWT	−17.7	24.42	23.48	−64.70	−16.80	−16.06
15	E80AWT	−18.5	25.79	23.5	−70.50	−21.22	−16.44
16	Y29AK27Awm	−10.4	23.17	20.85	−59.01	−14.99	−11.97
17	D35AK27Awm	−9.5	23.34	17.83	−54.84	−13.67	−8.33
18	W38FK27Awm	−12.6	23.81	24.02	−63.76	−15.92	−13.64
19	D39AK27Awm	−10.8	16.67	16.72	−41.94	−8.55	−8.69
20	T42AK27Awm	−13.3	24.46	23.7	−64.91	−16.74	−14.00
21	E76AK27Awm	−12.3	22.00	22.49	−55.88	−11.40	−12.70
22	E80AK27Awm	−13.5	24.22	22.61	−63.90	−17.07	−13.35
23	Y29AR59A	−10.9	21.04	19.12	−52.71	−12.54	−12.19
24	D35AR59Awm	−12.7	21.01	17.26	−50.83	−12.56	−9.61
25	W38FR59A	−12.8	22.87	22.78	−59.07	−13.43	−13.03
26	T42AR59A	−12.2	21.67	21.95	−57.22	−13.61	−13.19
27	E76AR59A	−14.1	23.09	21.59	−59.26	−14.59	−13.84
28	E80AR59A	−13.9	22.27	21.55	−58.12	−14.30	−13.67
29	Y29AR83Q	−10.7	17.79	19.47	−47.48	−10.22	−11.94
30	D35AR83Qwm	−9.4	18.56	16.75	−44.22	−8.91	−9.05
31	W38FR83Q	−12.3	19.29	23.38	−53.60	−10.94	−12.86
32	D39AR83Q	−12.6	15.75	16.79	−43.28	−10.74	−11.62
33	T42AR83Qwm	−12.9	18.40	22.26	−51.98	−11.33	−12.67
34	E76AR83Q	−12.3	16.79	21.61	−44.89	−6.50	−12.45
35	E80AR83Q	−13.3	19.62	22.15	−53.15	−11.37	−13.40
36	Y29AR87A	−11.0	21.29	19.76	−55.44	−14.39	−12.14
37	T42AR87A	−12.0	21.71	22.55	−58.37	−14.11	−12.31
38	E76AR87A	−12.2	19.93	21.76	−51.13	−9.44	−11.84
39	E80AR87A	−12.9	22.48	22.07	−58.82	−14.26	−12.73
40	Y29AH102Awm	−12.7	20.74	18.16	−56.24	−17.34	−10.92
41	Y29FH102Awm	−13.5	22.43	18.38	−59.78	−18.97	−11.51
42	W38FH102Awm	−11.4	22.79	21.69	−63.07	−18.59	−12.49
43	D39AH102Awm	−10.1	16.94	14.97	−42.81	−10.91	−8.54
44	T42AH102Awm	−10.9	21.41	20.53	−60.20	−18.25	−12.26
45	E76AH102Awm	−11.5	20.94	19.80	−55.40	−14.65	−12.40
46	E80AH102Awm	−12.3	23.40	20.31	−62.35	−18.64	−13.17
47	D35AE73Wwm	−13.3	33.59	19.88	−74.47	−20.99	−14.04
48	D39AE73A	−11.9	25.28	17.30	−53.31	−10.73	−11.38
49	D39AE73Q	−11.8	26.44	17.61	−55.21	−11.16	−11.17
50	E76AE73Q	−15.5	30.91	23.73	−75.93	−21.28	−14.97
51	WTE73A	−16.7	32.27	24.82	−83.56	−26.46	−16.91
52	WTE73C	−16.5	33.02	24.96	−84.69	−26.71	−17.09
53	WTE73F	−16.8	34.67	25.67	−88.44	−28.10	−16.98
54	WTE73Q	−17.6	32.52	25.32	−84.39	−26.55	−15.89
55	WTE73S	−16.0	33.59	25.24	−85.71	−26.89	−17.33
56	WTE73Y	−16.6	33.74	25.35	−86.94	−27.85	−16.68
57	D35AE73Awm	−12.6	32.74	18.67	−71.09	−19.68	−13.92
58	D35AE73Fwm	−12.6	33.05	19.05	−72.08	−19.98	−13.72
59	D39AE73F	−11.6	25.46	17.33	−53.44	−10.66	−11.38
60	D39AR87A	−11.9	18.35	17.19	−47.45	−11.92	−12.78
61	E76AE73W	−15.5	31.89	24.87	−79.09	−22.33	−16.03
62	W38FR87A	−12.0	22.37	23.50	−60.10	−14.23	−12.38

(*continued*)

TABLE 1 (*Continued*)

No.	Complex (bs:bn)*	ΔG_{exp} (kcal/mol)	$\Delta G_{ele}^{desol\_bn}$ (kcal/mol)	$\Delta G_{ele}^{desol\_bs}$ (kcal/mol)	E_{ele}^{bn-bs} (kcal/mol)	$\Delta G_{ele}^{binding}$ (kcal/mol)	ΔG_{pred} (kcal/mol)
63	WTE73W	−17.4	34.16	26.42	−88.83	−28.25	−17.21
64	WTR83Q	−13.6	18.52	22.96	−52.92	−11.44	−13.75
65	D39AE73W	−12.4	26.77	17.52	−54.68	−10.39	−11.73

*Complexes are named by the mutation in barstar (bs) followed by the mutation in barnase (bn). Complexes with additional water molecules are indicated by the last two letters ''wm'' in the names. The wild-type is indicated by ''WT''.

restore the wild-type barstar sequence, and the N-terminal methionine residue in barstar, which is located far from the interface and not noted in most of the previous literature, was deleted for ease of cross-referencing. Finally, the template contained 110 barnase residues, 89 barstar residues, and 213 bound water molecules. The polar hydrogen atoms of barnase, barstar, and the bound water molecules were added using the program WHATIF (Hooft et al., 1996; Vriend, 1990). H102 in barnase was protonated on Hε and H17 in barstar was protonated on Hδ. The nonpolar hydrogens were added using the tLeap module of AMBER7.0 (http://amber.scripps.edu).

The mutant complexes were chosen based on two references (Frisch et al., 1997; Schreiber and Fersht, 1995). The mutations were introduced on six barnase interfacial residues (K27, R59, E73, R83, R87, and H102) and eight barstar interfacial residues (Y29, D35, W38, D39, T42, W44, E76, E80) (see Fig. 1). For the 36 complexes in which the mutations are only $X \rightarrow$ alanine, the structures were modeled by deleting the side chains of the mutated residues in the template. For the other 28 complexes, which involve $X \rightarrow$ nonalanine mutations, the InsightII rotamer library (http://www.accelrys.com/insight/) was used to determine the side-chain conformation of the mutated residue while the backbone and the side chains of other residues remained frozen.

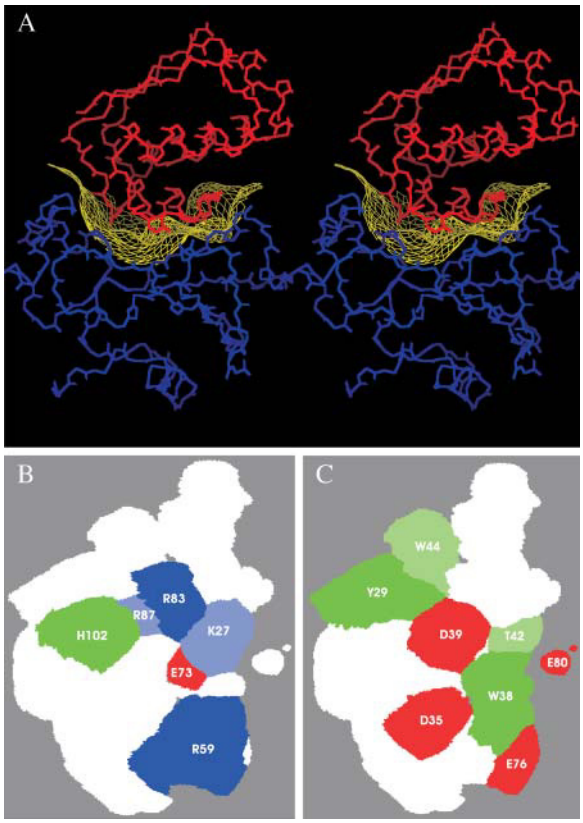

FIGURE 1 (*A*) Stereo view of barnase (*blue*)-barstar (*red*) binding interface (*yellow mesh*). 2-D projections onto the interface of (*B*) the six residues in barnase and (*C*) the eight residues in barstar that are mutated in the data set studied. All pictures were made with MolSurfer (Gabdoulline et al., 1999, 2003).

Some mutations, particularly those with $X \rightarrow$ alanine, created cavities accessible to water molecules. Therefore, three mutant complexes (Vaughan et al., 1999), which are the only mutants with available crystallographic structures in the Protein Data Bank (PDB) in the data set studied here, were used as templates for adding additional interfacial water molecules. These three complexes are K27A/bn-D35A/bs with the PDB entry code of 1b2u, H102A/bn-Y29F/bs with the entry code of 1b3s, and K27A/bn-T42A/bs with the entry code of 1b2s. Comparison of interfacial water molecules in these mutants and the wild-type complex indicated the following additional water molecules to be added to the mutant complexes: HOH48 in 1b2u for the K27A/bn mutant, HOH58 and HOH98 in 1b3s for the H102A/bn mutant, HOH2 and HOH35 in 1b2u for the D35A/bs mutant, and HOH85 and HOH65 in 1b3s for the Y29F/bs mutant.

All mutant complexes were prepared according to the above strategy with two exceptions: mutants with Y29F/bs-H102A/bn or E73W/bs. In the crystallographic structure of the mutant Y29F/bs-H102A/bn (PDB code 1b3s), F29 shows two conformations (Vaughan et al., 1999): swinging out of the interface in the complexes formed by chains A:D and chains C:F with significant backbone movement and side-chain conformational changes on barstar residues 28–30 and barnase residue Ser-85, and pointing to the interface in the complex formed by chains B:E. The inward-pointing conformation of F29 in the complex of chains B:E may be due to crystal packing. Therefore, the barstar residues 28–30 and the barnase residue Ser-85 in the mutant Y29F/bs-H102A/bn were modeled by using chains A:D in 1b3s as a reference structure, as the outward-facing orientation of F29 would not be obtained by energy minimization alone from the wild-type protein conformation. For complexes with the mutation E73W/bs, the best rotamer still had a serious steric clash with the neighboring residues, and energy minimization by DISCOVER in InsightII was carried out on this residue. The final conformation required the removal of three water molecules.

The 65 complexes are listed in Table 1 with their experimental binding free energies. The binding free energy of the wild-type complex is −19.0 kcal/mol (Schreiber and Fersht, 1995) and all mutants except Y29F/bs-WT/bn and W44F/bs-WT/bn show lower binding affinities.

Energy minimization

The all-atom AMBER 95 force field (Cornell et al., 1995) was used to obtain all the parameters for the proteins and water molecules. The tLeap module of AMBER7.0 was used to obtain the topology and coordinate files of each complex. Then the energy minimization of each complex was carried out using the Sander module of AMBER7.0 and consisted of three stages. In the first stage of 200 steps, the protein nonhydrogen atoms were restrained to their starting positions by a harmonic potential with a force constant of 32 kcal/(mol.Å^2) whereas the hydrogen atoms and the water molecules were unrestrained. In the second stage of 200 steps, the constraint was released from the side-chain atoms of the proteins and remained on the backbone atoms only. In the third stage of 400 steps, no constraint was used at all. A nonbonded cutoff of 10.0 Å and a distance-dependent dielectric constant ($\varepsilon = r_{ij}$) were used throughout. In each stage, the first 100 steps were

performed with the steepest descent algorithm and the rest of the steps were performed with the conjugate gradient method.

During the minimization, the backbone atoms of the proteins did not show observable movement, and only water molecules, particularly the additional interfacial water molecules, and some side-chain atoms showed significant movements. More extensive optimization of the structures of the mutants, e.g., by simulated annealing, which might help to predict the structures of mutants like Y29FH102A, would be possible but is beyond the scope of this work, which relied on the observation that most single-point mutants result in little effect on protein structure.

Electrostatic binding free-energy calculations

To investigate the electrostatic contributions to barnase-barstar binding, we first carried out continuum electrostatic calculations by using the University of Houston Brownian Dynamics (UHBD) Program UHBD6.1 (Madura et al., 1995). The electrostatic binding free energy $\Delta G_{ele}^{binding}$ is defined as in Eq. 2:

$$\Delta G_{ele}^{binding} = \Delta G_{ele}^{desol\_bn} + \Delta G_{ele}^{desol\_bs} + E_{ele}^{bn-bs}. \quad (2)$$

The electrostatic contribution to the desolvation energy of barnase, $\Delta G_{ele}^{desol\_bn}$ (or barstar, $\Delta G_{ele}^{desol\_bs}$), was defined as the loss of the electrostatic interaction between the solvent and barnase (or barstar) upon binding, as calculated by the two-step procedure described by Perez et al.(1998a): 1), calculate the electrostatic energy of barnase (or barstar) and the surrounding solvent in the absence of barstar (or barnase); 2), calculate the electrostatic energy of barnase (or barstar) and the surrounding solvent with the second protein bound but without partial charges. The electrostatic desolvation energy ($\Delta G_{ele}^{desol\_bn}$ or $\Delta G_{ele}^{desol\_bs}$) is the difference between the electrostatic energies computed from these two steps.

The electrostatic interaction E_{ele}^{bn-bs} was calculated by Eq. 3. ϕ_i is the electrostatic potential generated by barstar at the position of atomic charges q_i of barnase:

$$E_{ele}^{bn-bs} = \sum_{i}^{N} \phi_i q_i. \quad (3)$$

The finite difference method implemented in UHBD6.1 was used to solve the Poisson-Boltzmann equation. The interior dielectric constant of the protein was set to 2 and the solvent dielectric constant was set to 78 with an ionic strength of 50 mM and ionic radius of 1.5 Å. The coarse grid spacing was set to 0.80 Å and the fine grid spacing was set to 0.35 Å. The dielectric boundary was defined as the van der Waals surface. Both the coarse grid and the fine grid were dimensioned to $110 \times 110 \times 110$ with the center on the position of the Cγ atom of D39/bs. The coarse grid enclosed the whole protein complex whereas the fine grid enclosed the interface including all residues mutated.

Before doing the UHBD calculations, the minimized structures of all the complexes were superposed with the minimized structure of the wild-type barnase-barstar complex to ensure the same reference coordinates. Then, a separate program was written to convert the superposed structures of the complexes to *qcd* format files for input to UHBD6.1, with all water molecules removed.

In both steps 1 and 2 described above, structures of barnase and barstar as found in the bound conformation in the complexes were used.

Interaction energy decomposition

For the COMBINE analysis, the interaction energies between barnase and barstar were decomposed on a per residue pair basis. The ANAL module of AMBER7.0 (slightly modified) was used to calculate the Coulombic and the Lennard-Jones interaction energies between each protein residue in the energy minimized complex. A separate code was written to extract the intermolecular energy terms between each barnase residue and each barstar residue, generating 19580 (=110 barnase residues $\times$ 89 barstar residues $\times$ 2) energy descriptors for each complex.

Chemometric analysis

The GOLPE4.5.1 program (Baroni et al., 1993) was used to carry out the chemometric analysis. A matrix was constructed with each row representing an object, in this case a protein-protein complex. Each object is represented by the same number of columns in the matrix corresponding to the chemical/physical descriptors (called *X* variables) and responsive variables (called *Y* variables). Here, there are 19582 *X* variables: 9790 Coulombic energy terms, 9790 Lennard-Jones energy terms, and two desolvation energies. There is one *Y* variable, assigned as the binding free energy.

To reduce the size of the matrix, the *X* variables showing little variation among the complexes, below 0.1 kcal/mol for Lennard-Jones terms and below 0.9 kcal/mol for the Coulombic terms, were zeroed. As a result, 233 *X* variables were retained for further analysis. To investigate the distribution of the 65 complexes in the energy space defined by these *X* variables, a PCA was performed. The distances between complexes were measured by the PCA scores. Then, the *X* variables were correlated with the *Y* variable by PLS analysis to yield initial PLS models of varying dimensionality. Leave-one-out cross-validation was performed to determine the optimal dimensionality for predictive performance. To remove the noisy variables and improve the predictive abilities of the PLS models, an *X* variable selection procedure consisting of a D-optimal preselection and a fractional factorial design was performed for up to eight latent variables. The D-optimal preselection removed nearly half of the *X* variables without affecting model quality, and the fractional factorial design further removed a few *X* variables while retaining uncertain variables. Final PLS models were built for the remaining 111 *X* variables, with interaction energies between 28 barnase and 21 barstar residues. The final models displayed significantly higher predictive cross-validation and slightly higher fitting performance than the initial PLS models.

To further evaluate the robustness of the data and the models, we randomly selected five test sets, each containing 55 complexes for training and 10 complexes for external prediction.

BD calculations

Bimolecular diffusional association rates were computed at 50 mM ionic strength (the ionic strength used in experiments) using the same protocol as described in Gabdoulline and Wade (2001) with minor differences. As previously, the rates to form two hydrogen bond donor-acceptor contacts (observed in the bound complex of the wild-type proteins) were computed by Brownian dynamics simulation using SDA software (Gabdoulline and Wade, 1997, 1998). Here, however, we used the energy minimized structures of the complexes of barnase and barstar prepared in this work and the AMBER 95 force field rather than the OPLS force field. As a result, because the energetic optimization of the structures improves their electrostatic interactions and because the 2 times smaller radii of polar hydrogen atoms in the AMBER force field compared to the OPLS force field allow shorter and stronger intermolecular interactions, the computed association rate for the wild-type barnase-barstar complex was 2.5 times higher than that computed for barnase and barstar earlier (Gabdoulline and Wade, 2001). Therefore, we adjusted the interatomic contact distance to define encounter complex formation from 6 Å to 5 Å. This resulted in satisfactory computed association rates ~2 times higher than experimental values. The rate constants for formation of two donor-acceptor contacts at 5 Å were therefore used to quantify association rate changes due to mutation of the proteins.

RESULTS AND DISCUSSION

Computed electrostatic binding free energies are favorable

The calculated $\Delta G_{ele}^{deso\_bn}$, $\Delta G_{ele}^{desol\_bs}$, E_{ele}^{bn-bs}, and $\Delta G_{ele}^{binding}$ values from the Poisson-Boltzmann electrostatics calculations are listed in Table 1 along with the values of the experimental binding free energy, ΔG_{exp}. With the parameterization used in the calculations, the electrostatic interactions can compensate and surpass the desolvation costs in the complexes analyzed. For the wild-type complex, the electrostatic binding free energy is −22.2 kcal/mol, exceeding the experimental binding free energy (−19.0 kcal/mol) by 3.2 kcal/mol. Most mutants (46 out of 64) also have an electrostatic binding free energy that is more negative (favorable) than the experimental binding free energy.

It should be noted that the values reported were computed with a dielectric boundary defined by the protein van der Waals surface. We also did the calculations with the dielectric boundary defined by the solvent-accessible molecular surface (SAMS) (composed of the contact and reentrant surfaces of a solvent probe of radius 1.4 Å), which gave positive (unfavorable) electrostatic binding free energies for almost all the complexes. For the wild-type complex, the electrostatic binding free energy was +9 kcal/mol. Literature reports (Dong et al., 2003; Dong and Zhou, 2002; Vijayakumar and Zhou, 2001; Xu et al., 1997) have also shown that continuum electrostatic calculations are sensitive to the dielectric boundary definition, and greater agreement with experiment has been found by using the van der Waals surface boundary definition (Dong et al., 2003). The SAMS boundary definition leads to higher desolvation costs.

Computed electrostatic binding free energies correlate with experimental binding free energies

The linear correlation coefficients between the experimental binding free energy ΔG_{exp}, and the electrostatic binding free energy $\Delta G_{ele}^{binding}$, the electrostatic interaction energy E_{ele}^{bn-bs}, the electrostatic desolvation energy of barstar $\Delta G_{ele}^{desol\_bs}$, and the electrostatic desolvation energy of barnase $\Delta G_{ele}^{desol\_bn}$ are 0.777, 0.796, 0.739, and 0.655, respectively (see Fig. 2). The unfavorable desolvation energies oppose the favorable interaction energies. Thus, the experimental binding free energy is positively correlated with the electrostatic interaction energy between barnase and barstar and negatively correlated with the desolvation energies of both barnase and barstar. It is worth mentioning that the experimental binding free energy ΔG_{exp} shows a much poorer correlation (with a coefficient of 0.4) with the electrostatic binding free energy $\Delta G_{ele}^{binding}$ computed by using the SAMS dielectric boundary definition (data not shown).

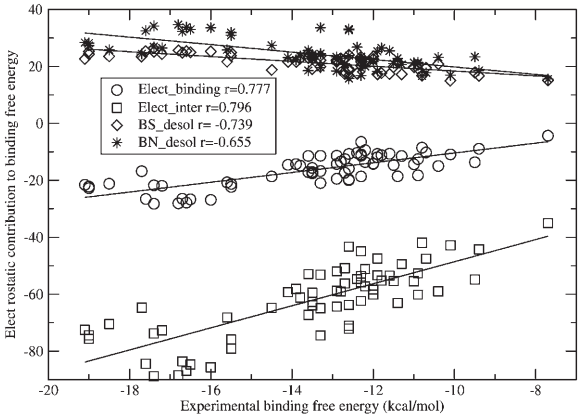

FIGURE 2 Linear correlations between the experimental binding free energy ΔG_{exp} and the computed continuum electrostatics binding free energy $\Delta G_{ele}^{binding}$, the electrostatic interaction energy E_{ele}^{bn-bs}, the electrostatic desolvation energy of barstar $\Delta G_{ele}^{desol\_bs}$, and the electrostatic desolvation energy of barnase $\Delta G_{ele}^{desol\_bn}$.

The mutations, with the exception of those at W44/bs and E73/bn, are unfavorable for the computed electrostatic binding free energies

Mutations of interfacial residues in most cases resulted in a decrease of the magnitude of the electrostatic binding free energy, the exceptions being the complexes with W44F/bs and those with single mutations of E73/bn.

It appears that W44/bs can be a site to design mutants with improved electrostatics and therefore we modeled three single-point mutants: W44Y/bs, W44E/bs, and W44D/bs. All three mutants have computed electrostatic binding free energies that are more favorable than that of the wild-type protein (−22.2 kcal/mol) but similar to that of W44F/bs (−22.8 kcal/mol). The experimental binding free energy of the W44F/bs mutant (−19 kcal/mol) is the same as that of the wild-type complex.

E73/bn is critical for the catalytic activity of barnase (Schreiber et al., 1997) and, in the complex, it is located in the vicinity of the negatively charged binding surface residues of barstar (in most complexes, these are D35 and D39). Although the attractive electrostatic interaction between barnase and wild-type barstar increases with the mutation of E73 to neutral residue types (by 4–6 kcal/mol), the experimental binding affinities actually decrease by 1.5–3 kcal/mol. This is mainly because an indirect favorable interaction observed in double-mutant experiments (Schreiber et al., 1997) between E73/bn and D39/bs is lost upon mutation. This stabilizing interaction is between E73/bn and its neighboring positively charged residues K27/bn, R83/bn and R87/bn, which are close to D39/bs in the complex.

The complexes with W44/bs and E73/bn mutants demonstrate that the binding of barstar and barnase is not fully electrostatically optimized. When Lee and Tidor considered single-point mutations of barstar to the 20 common amino acids (Lee and Tidor, 2001a), they found that wild-type

barstar is electrostatically optimized in the sense that, although charge optimization (allowing unnatural charges) could result in improved electrostatic binding free energies of up to ~1 kcal/mol at residues 38, 72, and 76 of barstar, no modeled mutations of these residues resulted in greater computed binding affinity than the wild-type proteins. On the other hand, Lee and Tidor (2001b) found that the electrostatic complementarity of the barnase-barstar complex could be significantly improved by optimizing the charge distribution by allowing for unnatural charge distributions and that the charge distribution of barstar was more optimal for binding than that of barnase.

For the set of mutations that we have studied, we find, in accord with Lee and Tidor, that barstar appears to be more electrostatically optimized than barnase. This is consistent with barstar's function as an inhibitor and barnase's need to have a charge distribution suitable for catalysis as well as binding barstar. We find that it is possible to modestly improve the electrostatic free energy of binding of barstar by mutation to one of the 20 common amino acids at position 44, as well as the positions 38 and 72 identified by Lee and Tidor (2001a). The differences in the results of the Poisson-Boltzmann calculations most likely arise primarily from the fact that Lee and Tidor optimized only the partial atomic charges of the amino acids in barstar, whereas our models of mutants accounted for changes in charge magnitude, position, and dielectric boundary location, and employed a different dielectric boundary definition. It should be noted that improved electrostatic binding free energy for the W44F/bs and E73/bn mutants does not result in improved overall binding affinity: the experimentally measured binding affinity is comparable to or weaker for these mutants than for the wild-type proteins.

TABLE 2 Coupling energies (kcal/mol) of six barstar (bs)-barnase (bn) residue pairs

		Calculated	
Residue pair	Experimental	Electrostatic Poisson-Boltzmann	COMBINE analysis
D35A/bs-R59A/bn	3.4	1.3	−0.3
D39A/bs-H102A/bn	4.9	2.2	−1.7
D39A/bs-K27A/bn	4.8	2.2	−1.0
D39A/bs-R87A/bn	6.1	7.0	3.0
D39A/bs-R83Q/bn	6.7	10.0	2.4
Y29A/bs-H102A/bn	3.3	−0.6	−0.5

Computed electrostatic binding free energies account to varying extents for pairwise residue cooperativity

The cooperative interaction energy or the coupling energy between two residues X and Y, $\Delta\Delta G_{\text{int (X-Y)}}$, is defined as

$$\Delta\Delta G_{\text{int}(X\rightarrow Y)} = \Delta\Delta G_{X\rightarrow A,Y\rightarrow B} - \Delta\Delta G_{X\rightarrow A} - \Delta\Delta G_{Y\rightarrow B}, \quad (4)$$

where $\Delta\Delta G_{X\rightarrow A}$ (or $\Delta\Delta G_{Y\rightarrow B}$) is the change in binding free energy on mutation of X to A (or Y to B), and $\Delta\Delta G_{X\rightarrow A, Y\rightarrow B}$ is the change upon the simultaneous mutation of X to A and Y to B. Based on Eq. 4, the electrostatic coupling energies of the six residue pairs showing significant experimental coupling energies were computed as shown in Table 2. It can be seen that the extent to which the continuum electrostatics interaction energy accounts for the cooperativities depends on the relative orientations and electrostatic properties of the residue pairs: completely for a charged-charged residue pair making side-chain hydrogen bonds like D39/bs and R87/bn with a 6–7 kcal/mol coupling energy, partially for a charged-polar residue pair making side-chain hydrogen bonds like D39/bs and H102/bn (coupling energy of ~5 kcal/mol and computed electrostatic energy of ~2 kcal/mol), and not at all for a polar-polar residue pair making no hydrogen bond like Y29/bs and H102/bn (with a coupling energy of ~3 kcal/mol). D35/bs and R59/bn form only a backbone hydrogen bond (D35:OD1-R59:N) and therefore the alanine truncation of R59 still retained most of the electrostatic interaction. This makes the cooperativity defined in Eq. 4 smaller (coupling energy of ~3 kcal/mol) than when side-chain hydrogen bonds only are involved and this is only partially accounted for by the electrostatic energy (1.3 kcal/mol). The cooperativity between D39A/bs and K27A/bn that is unaccounted by electrostatic interactions (2.6 of the 4.8 kcal/mol coupling energy) may be due to the hydrophobic effect generated by the long exposed side chain of K27/bn, although there is a hydrogen bond between K27:NZ and D39:OD1.

Principle component analysis highlights six important residues for barnase-barstar binding energetics

The score plot of the first two principle components (PC1 and PC2) is given in Fig. 3 and shows how the barnase-barstar complexes are distributed in interaction energy space. All complexes except D39AR59A, E76AR83Q, and E76AR87A are clustered into five groups: A), mutants of D39A/bs with the largest positive PC1; B), mutants of R83Q/bn with the second largest positive PC1; C), mutants of R87A/bn; D), mutants of E76A/bs and mutants of R59A/bn; and E), most single mutants. In PC3 and PC4 (Fig. 4), the complexes are distributed more widely, but it is clear that all mutants of D35A/bs have positive values of PC4. The PCA results imply that mutations at D35/bs, D39/bs, E76/bs, R59/bn, R83/bn, and R87/bn have substantial effects on the energetics of barnase-barstar binding. The correlation between these effects and the binding affinities will be illustrated by the following PLS analysis.

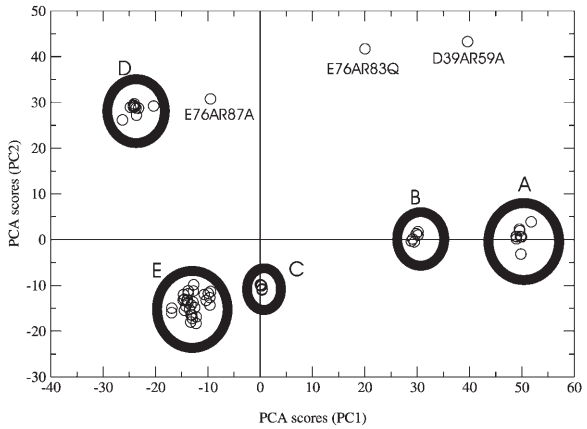

FIGURE 3 Score plot of the first (PC1) and second (PC2) principal components of the interaction energy terms for the 65 complexes. All the complexes except D39AR59A, E76AR83A, and E76AR87A are clustered into five groups: (*A*) D39A/barstar mutants with the largest positive PC1; (*B*) R83Q/barnase mutants with the second largest positive PC1; (*C*) R87A/barnase mutants; (*D*) E76A/bs and R59A/bn mutants; and (*E*) most single mutants.

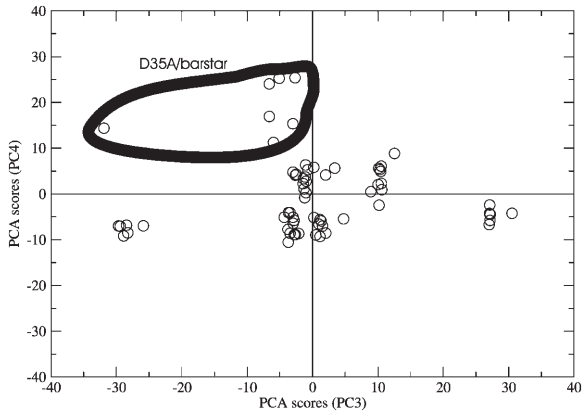

FIGURE 4 Score plot of the third (PC3) and fourth (PC4) principal components of the interaction energy terms for the 65 complexes. All mutants of D35A/bs were distinguished by a large positive PC4.

COMBINE analysis models are predictive for binding affinity

The statistical parameters of the COMBINE PLS models are given in Table 3. The optimal dimensionality was determined as seven latent variables because the model quality (as measured by fitting (R^2 and standard deviation of the error of calculation), and cross-validation (Q^2 and standard deviation of the error of prediction (SDEP)) parameters—see Table 3 for the definitions of these parameters) does not increase significantly by adding more latent variables (see Fig. 5). At this number of latent variables, the predictive Q^2 value is 0.82 (where a Q^2 value of 1 corresponds to perfect prediction and a Q^2 value >0.4 indicates a predictive model). The SDEP is 1.07 kcal/mol. The predicted binding free energies are listed in Table 1 and a plot against the experimental values is shown in Fig. 6. In leave-one-out cross-validation, 54 of the 65 complexes were predicted with an error <1.5 kcal/mol. In five external random test sets (see Methods), the external SDEP values are 1.16, 0.80, 1.17, 0.96, and 1.37 kcal/mol, respectively, at the optimal latent variable number of 7. These SDEP values are mostly close to the SDEP value obtained in leave-one-out cross-validation, indicating robustness of the model.

TABLE 3 Predictive performance of the COMBINE analysis models derived for 65 wild-type and mutant barnase-barstar complexes

*LV	R^2	SDEC (kcal/mol)	Q^2	SDEP (kcal/mol)	Constant C (kcal/mol)
6	0.87	0.90	0.77	1.21	7.04
7	0.91	0.75	0.82	1.07	8.83
8	0.93	0.68	0.85	0.97	7.39

*LV is the number of latent variables. Q^2 is the cross-validated predictive performance and is given by

$$Q^2 = 1 - \frac{\sum_{i=1}^{n}\left(y_{\mathrm{exp(i)}} - y_{\mathrm{pred(i)}}\right)^2}{\sum_{i-1}^{n}\left(y_{\mathrm{exp(i)}} - \langle y_{\mathrm{exp}}\rangle\right)^2},$$

where $y_{\mathrm{pred(i)}}$ corresponds to the value of the quantity predicted with the model for complex i, $y_{\mathrm{exp(i)}}$ is the experimental value of the quantity for complex i, and $\langle y_{\mathrm{exp}}\rangle$ is the average experimental value of the quantity for the complete set of n complexes. A Q^2 value of 1 corresponds to a perfect prediction and a Q^2 value >0.4 is considered indicative of a predictive model. R^2 is the equivalent of Q^2 calculated for fitting. SDEP is the standard deviation in cross-validated prediction and is given by

$$\mathrm{SDEP} = \sqrt{\frac{\sum_{i=1}^{n}\left(y_{\mathrm{exp(i)}} - y_{\mathrm{pred(i)}}\right)^2}{n}}.$$

SDEC is the equivalent of SDEP for fitting. The constant C is as given in Eq. 1 for each COMBINE analysis model.

The predictive ability of the COMBINE model compares favorably with other published models for barnase-barstar

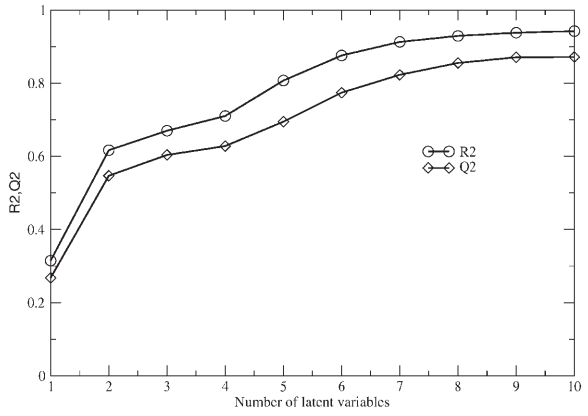

FIGURE 5 Development of the fitting and predictive cross-validation performance of the COMBINE analysis models during derivation. The R^2 and Q^2 values are plotted against the number of latent variables in the COMBINE analysis model.

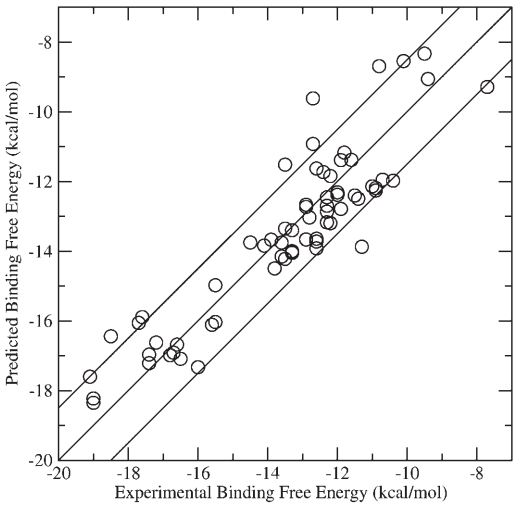

FIGURE 6 Plot of experimental versus predicted binding free energies (kcal/mol) for the COMBINE analysis model for the 65 complexes from leave-one-out cross-validation at seven latent variables. Fifty-four complexes were predicted to within 1.5 kcal/mol (as indicated by the flanking *diagonal lines*).

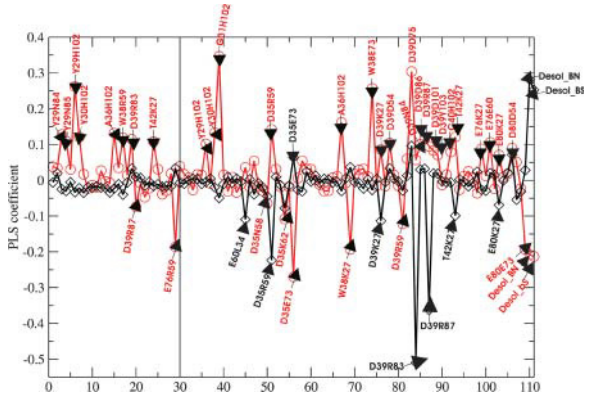

FIGURE 7 PLS coefficients (in *red*) of the 111 selected energy terms (*X* variables) in the COMBINE model. Those with absolute values >0.06 are labeled by the barnase and barstar residue pairs. The energy terms on the left and right sides of the vertical line are from Lennard-Jones and Coulombic interactions, respectively. The values of the 111 energy terms in the wild-type complex after scaling by 100 kcal/mol are shown in black.

binding affinity, providing overall better accuracy over a larger set. However, it should be born in mind that the COMBINE model is system-specific and its derivation requires a number of complexes with experimentally determined binding affinities that can constitute a training set. Once the COMBINE QSAR model has been derived, it can be applied to any number of designed mutations. The complexes for these mutations should be constructed and energy minimized, and the energy components fed into the COMBINE QSAR model.

COMBINE analysis models highlight the interaction energies of 27 residues and the protein desolvation energies as particularly important for binding affinity

To investigate how the binding free energy was weighted by the energy terms, we plotted the PLS coefficients of the 111 selected *X* variables (energy terms), namely the weight parameters w_{bn}^{desol}, w_{bs}^{desol}, w_{i}^{vdw}, and w_{i}^{ele} in Eq. 1, in Fig. 7. The most significant coefficients, with absolute values >0.06, are labeled in Fig. 7. They relate to the two desolvation terms and the interaction terms between 16 barnase residues and 11 barstar residues, including the 11 hot-spot residues defined by Bogan and Thorn (1998). All these residues are located in the binding interface. In particular, most of the large coefficients involve the barnase residues K27, E73, and H102, and the barstar residues D35 and D39. This means that, given the same values (interaction energies), these energy terms will have larger effects on the binding free energies than the other energy terms in the COMBINE model.

The positive value of the constant *C* (8.83 kcal/mol at 7 latent variables) suggests that the overall effect of the energy terms in the COMBINE analysis model is favorable to binding in all the complexes. The constant *C* can be interpreted as in part due to the translational, rotational, and conformational entropy lost upon protein binding, which is largely independent of residue mutations. However, as shown in Fig. 7, an individual residue can contribute to the binding free energy in opposing ways, depending on its interactions with other residues. For example, a positively charged residue at the position of D39/bs would favor binding by the attractive electrostatic interaction with K27/bn but disfavor binding by the repulsive electrostatic interaction with D75/barnase.

Both barnase and barstar electrostatic desolvation energies have negative PLS coefficients (−0.208 for barnase and −0.213 for barstar). This means that, surprisingly, the electrostatic desolvation cost has a favorable contribution to the binding free energy represented in this model. However, we should understand this as the net effect of all the energy terms used in the model, not the effect of the electrostatic desolvation cost alone, i.e., the electrostatic desolvation energies implicitly include the effects of other, favorable, energy terms.

Overall, COMBINE analysis provides a model for binding free energy with terms and weights that can be well related to the underlying physics determining binding affinity. It is worth noting, though, that it is possible in COMBINE analysis for a particular variable to be included in the model that is of less physical relevance than a variable, or combination of variables, with which it varies approximately collinearly over the data set. A variable may also have an apparently unphysical weight for similar reasons, as is the case for the electrostatic desolvation terms discussed above. These potential problems can be minimized by using suitable structure preparation procedures and conservative variable selection procedures (see Methods). Apparently unphysical terms or

weights in COMBINE models may also arise if the training set is not sufficiently broad and balanced to provide the required information for the prediction problem. For the case of models for protein-protein binding affinity, the training set should include information about mutations of the major ''hot-spot'' residues.

COMBINE models only partially represent pairwise residue cooperativity

Although for the majority of the complexes (54 out of 65) the binding free energies were predicted with errors <1.5 kcal/mol, some single-point mutants, especially D39/bs, were overpredicted whereas the related double mutants were underpredicted, which resulted in underprediction of the cooperativities of these residues. Only the cooperativities of D39A/bs-R87A/bn and D39A/bs-R83Q/bn were partially reproduced (3.0 kcal/mol and 2.4 kcal/mol) (see Table 2).

As shown in the Electrostatics section, continuum electrostatics alone can fully or partially account for the cooperativities of five of the six residue pairs. So we decomposed the continuum electrostatic interaction E_{ele}^{bn-bs} on a per residue pair basis and used it to replace the Coulombic interaction variables and rebuild the COMBINE analysis models. However, we obtained very similar COMBINE models with the incomplete representation of cooperativity remaining.

It is generally believed that cooperativity mainly arises from the close interactions between the residues (Buczek et al., 2001; Li et al., 2003; Pielak and Wang, 2001; Roisman et al., 2001; Wells and Cunningham, 1993; Zhang et al., 2002). As electrostatic interactions and Lennard-Jones interactions are included in our models, one possible missing interaction is the hydrophobic interaction. Following this consideration, we calculated the solvent accessible surface area (SASA) change of each residue upon binding by using the program NACCESS2.1 (Hubbard and Thornton, 1993) and separated it into two parts: SASA of polar atoms and SASA of nonpolar atoms. This resulted in an additional $2 \times 199 = 398$ terms (X variables) that were used for building the COMBINE models. However, no improvement in model predictive ability was obtained.

The COMBINE model, together with Poisson-Boltzmann electrostatic and Brownian dynamics calculations, assist design of mutants

For the wild-type complex, the individual Lennard-Jones and Coulombic contributions of each residue pair were obtained by multiplying the corresponding PLS coefficient and the X variable (interaction energy to which the PLS coefficient belongs) in Fig. 7. The favorable (more negative than -0.5 kcal/mol) and unfavorable (more positive than 0.3 kcal/mol) contributions are shown in Fig. 8, *A* and *B*, respectively. For

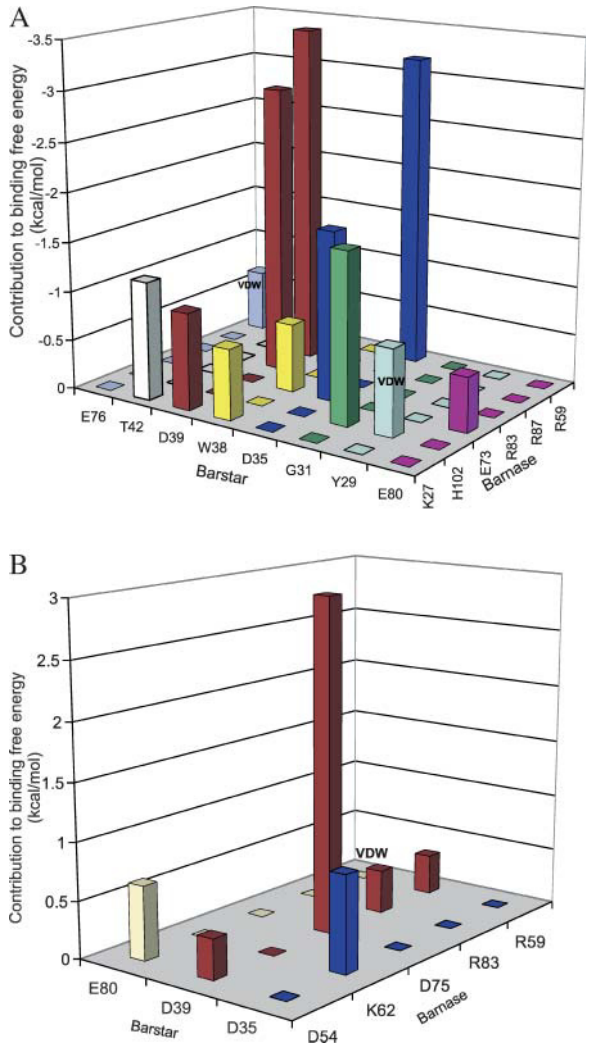

FIGURE 8 The individual contribution of each residue pair to the binding free energy of the wild-type barnase-barstar complex. (*A*) Favorable contributions (>0.5 kcal/mol in magnitude) and (*B*) unfavorable contributions (>0.3 kcal/mol). The contributions are Coulombic interactions unless labeled *VDW*, indicating a Lennard-Jones interaction.

the purpose of designing mutants with higher binding affinity, one should think about eliminating unfavorable contributions while keeping favorable contributions intact. As can be seen in Fig. 8 *b*, unfavorable contributions came from interactions between both oppositely charged residues (D39/bs-R59/bn and D35/bs-K62/bn) and residues of like charge (D39/bs-D75/bn and E80/bs-D54/bn). The former can be due to indirect interactions, which are hard to consider for designing mutants. Therefore, we designed mutants by taking into account the residue pairs of like charge and modeled the following single-point mutants: D54N/bn and D75N/bn. The training data set of 65 complexes does not contain any complexes with mutations at these two positions. Nevertheless, predictions of the effects of mutations can be made due to the energetic description employed in the COMBINE QSAR model. For both mutants, the predicted

binding affinities are stronger than for the wild-type (−18.4 kcal/mol). With seven latent variables in the COMBINE analysis model, the predicted binding free energy is −19.7 kcal/mol for the D54N/bn mutant and −21.6 kcal/mol for the D75N/bn mutant. The calculated electrostatic binding free energies are −25.1 kcal/mol for the D54N/bn mutant and −24.9 kcal/mol for the D75N/bn mutant (compared to −22.2 kcal/mol for the wild-type). These predictions could be modulated by changes in the stability of barnase due to mutation. For E73A/bn (a very low activity mutant), the measured change in stability of barnase upon mutation is unfavorable (2.30 kcal/mol) (Meiering et al., 1992). Although this change in stability is much larger than for most mutants for which this has been measured and which alter stability by <0.6 kcal/mol, the COMBINE model gives a good prediction with an error of only 0.2 kcal/mol. For D54N/bn, the change in stability upon mutation is also unfavorable and of similar magnitude (2.67 kcal/mol) but, although the k_{cat} and K_m values for activity are altered, the enzyme activity (as quantified by k_{cat}/K_m) is slightly greater than for the wild-type barnase (Meiering et al., 1992).

In addition, we also made mutations of D86/bn as its interaction with D39/bs has a positive PLS coefficient (see Fig. 7) and so an unfavorable contribution (small, not shown in Fig. 8) to the wild-type binding free energy. Two mutants D86N/bn and D86K/bn were constructed and their binding free energies were predicted. As expected, only D86K/bn showed notable improvement, with a COMBINE-predicted binding free energy of −19.2 kcal/mol and a calculated electrostatic binding free energy of −23.6 kcal/mol.

As described in the Electrostatics section, we modeled three single-point mutants of W44/bs: W44Y/bs, W44E/bs, and W44D/bs. Their binding free energies were predicted with the COMBINE model as −19.4 kcal/mol for W44Y/bs, −19.3 kcal/mol for W44E/bs, and −18.8 kcal/mol for W44D/bs, indicating modestly improved binding compared to the wild-type protein.

The designed mutants were further quantified by computing their association rate constants, k_{on}, using Brownian dynamics simulations (Gabdoulline and Wade, 2001). A change in binding free energy can be due to a change in association or dissociation rate constant or both. For example, Schreiber and co-workers (Selzer et al., 2000) designed faster associating mutants for binding of β-lactamase and an inhibitor protein that were also shown to bind tighter than the wild-type proteins but did not affect the dissociation rate. If the results of Brownian dynamics calculations of k_{on}, and the COMBINE analysis calculations of binding free energy can, despite the differences in the theoretical models, be combined, then k_{off} values can be derived. For the barnase and barstar mutants studied here, the calculations indicate that both changes in association and dissociation rate constants contribute to the binding free-energy differences. The complexes with D54N/bn and D75N/bn mutants have computed association rate constants 3.3 and 2.4 times, higher, respectively, than the wild-type proteins at 50 mM ionic strength. This means that the contribution of the change in association rate to overall binding free energy, derived from the relation $\Delta G = k_B T \times \ln(k_d) = k_B T \times \ln(k_{off}/k_{on})$ is ~−0.7 and −0.5 kcal/mol, respectively. The remaining −0.7 and −2.8 kcal/mol changes are, respectively, expected to be due to a changed k_{off}. The computed enhancement of association rate for the D54N/bn mutant is consistent with the experimental assignment of Schreiber and Fersht of a 4.4-fold increase in the association rate constant at zero ionic strength (Vijayakumar et al., 1998) and a 1.5-fold increase in the association rate constant at 100 mM ionic strength (Schreiber and Fersht, 1993) for the D54A/bn mutant compared to the wild-type protein. Moreover, the binding affinity of the D54A/bn mutant has been measured to be 0.2 kcal/mol more favorable than for wild-type proteins at 100mM ionic strength (Schreiber and Fersht, 1993) with a slightly higher dissociation constant (1.8 vs. 1.5 s^{-1}).

The mutant D86K/bn also showed an increased computed association rate and contributions of −0.4 and −0.5 kcal/mol to $\Delta\Delta G$ from changes in k_{on} and k_{off}, respectively.

The mutant W44Y/bs does not show any changes in computed association rate, but mutants W44E/bs and W44D/bs associate ~2 times faster. This means that in the case of W44E/bs, $\Delta\Delta G$ is made of −0.4 and −0.6 kcal/mol contributions from changes in k_{on} and k_{off}, respectively. For the W44D/bs mutant, almost all the change in computed binding free energy is due solely to the change in k_{on}.

CONCLUDING REMARKS

A large set of 65 wild-type and mutant complexes of barnase and barstar was studied. For these complexes, Poisson-Boltzmann electrostatic calculations with the van der Waals surface dielectric boundary definition show electrostatic interactions as the dominant term favoring barnase-barstar binding. The electrostatic binding free energies became slightly more favorable in the single mutants W44/bs → F, Y, E, and D, and significantly more favorable in the single mutants E73/bn → A, C, F, Q, S, Y, and W. This indicates that barstar is more electrostatically optimized than barnase but neither wild-type barnase nor wild-type barstar is fully electrostatically optimized for binding with respect to mutation to the 20 common amino acids.

The overall effects of interfacial mutations can be quantitatively predicted by the COMBINE QSAR model and are represented mainly by the interaction energies between 16 barnase and 11 barstar residues and the electrostatic desolvation energies of barnase and barstar upon binding. The COMBINE analysis model can provide a useful guide for interface design, as shown by the examples of the mutant complexes D54N/bn and D75N/bn. The COMBINE analysis model, together with Poisson-Boltzmann electrostatics calculations and Brownian dynamics simulations, give predictions

of mutants that should bind faster and with higher affinity than the wild-type proteins.

We thank Karin Schleinkofer for critical reading of the manuscript. We thank Dr. Gabriele Cruciani (Perugia) for provision of the GOLPE program.

We gratefully acknowledge the financial support of the Klaus Tschira Foundation and the German-Croatian scientific collaboration project WTZ-HRV01/010.

REFERENCES

Baroni, M., G. Costantino, G. Cruciani, D. Riganelli, R. Valigi, and S. Clementi. 1993. Generating optimal linear PLS estimations (GOLPE): an advanced chemometric tool for handling 3D-QSAR problems. *Quant. Struct.-Act. Rel.* 12:9–20.

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

Buczek, O., K. Koscielska-kasprzak, D. Krowarsch, M. Dadlez, and J. Otlewski. 2001. Analysis of serine proteinase-inhibitor interaction by alanine shaving. *Protein Sci.* 11:806–819.

Chong, L. T., S. E. Dempster, Z. S. Hendsch, L. P. Lee, and B. Tidor. 1998. Computation of electrostatic complements to proteins: a case of charge stabilized binding. *Protein Sci.* 7:206–210.

Cornell, W. D., P. Cieplak, C. I. Payly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

Covell, D. G., and A. Wallqvist. 1997. Analysis of protein-protein interactions and the effects of amino acid mutations on their energetics. The importance of water molecules in the binding epitope. *J. Mol. Biol.* 269:281–297.

Dong, F., M. Vijayakumar, and H.-Y. Zhou. 2003. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar. *Biophys. J.* 85:49–60.

Dong, F., and H.-Y. Zhou. 2002. Electrostatic contributions to exposed charges versus semi-buried salt bridges. *Biophys. J.* 83:1341–1347.

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

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

Gabdoulline, R. R., and R. C. Wade. 1998. Brownian dynamics simulation of protein-protein diffusional encounter. *Methods.* 3:329–341.

Gabdoulline, R. R., and R. C. Wade. 2001. Protein-protein association: investigation of factors influencing association rates by Brownian dynamics simulations. *J. Mol. Biol.* 306:1139–1155.

Gabdoulline, R. R., R. C. Wade, and D. Walther. 1999. MolSurfer: two-dimensional maps for navigating three-dimensional structures of proteins. *Trends Biochem. Sci.* 24:285–287.

Gabdoulline, R. R., D. Walther, and R. C. Wade. 2003. MolSurfer: a macromolecular interface navigator. *Nucleic Acids Res.* 31:3349–3351.

Hooft, R. W. W., C. Sander, and G. Vriend. 1996. Positioning hydrogen atoms by optimizing hydrogen bond networks in protein structures. *Proteins.* 26:363–376.

Hubbard, S. J., and J. M. Thornton. 1993. NACCESS2.1. Department of Biochemical and Molecular Biology, University College London, London, UK.

Kortemme, T., and D. Baker. 2002. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc. Natl. Acad. Sci. USA.* 99:14116–14121.

Lee, L. P., and B. Tidor. 2001a. Barstar is electrostatically optimized for tight binding to barnase. *Nat. Struct. Biol.* 8:73–76.

Lee, L. P., and B. Tidor. 2001b. Optimization of binding electrostatics: charge complementarity in the barnase-barstar protein complex. *Protein Sci.* 10:362–377.

Li, Y., M. Urrutia, and S. J. Smith-Gill. 2003. Dissection of binding interactions in the complex between the anti-lysozyme antibody HyHEL-63 and its antigen. *Biochemistry.* 42:11–22.

Lozano, J. J., M. Pastor, G. Cruciani, K. Gaedt, N. B. Centeno, F. Gago, and F. Sanz. 2000. 3D-QSAR methods on the basis of ligand receptor complexes. Application of COMBINE and GRID/GOLPE methodologies to a series of CYP1A2 ligands. *J. Comput. Aided Mol. Des.* 14:341–353.

Madura, J. D., J. M. Briggs, R. C. Wade, M. E. Davis, B. A. Luty, A. Ilin, J. Antosiewicz, M. K. Gilson, B. Bagheri, L. R. Scott, and J. A. McCammon. 1995. Electrostatics and diffusion of molecules in solution: simulations with the University of Houston Brownian Dynamics Program. *Comput. Phys. Commun.* 91:57–95.

Meiering, E. M., L. Serrano, and A. R. Fersht. 1992. Effect of active site residues in barnase on activity and stability. *J. Mol. Biol.* 225:585–589.

Ortiz, A. R., M. Pastor, A. Palomer, G. Cruciani, F. Gago, and R. C. Wade. 1997. Reliability of comparative molecular field analysis models: effects of data scaling and variable selection using a set of human synovial fluid phospholipase A2 inhibitors. *J. Med. Chem.* 40:1136–1148.

Ortiz, A. R., M. T. Pisabarro, F. Gago, and R. C. Wade. 1995. Prediction of drug binding affinities by comparative binding energy analysis. *J. Med. Chem.* 38:2681–2691.

Pal, G., A. A. Kossiakoff, and S. S. Sidhu. 2003. The functional binding epitope of a high affinity variant of human growth hormone mapped by shotgun alanine-scanning mutagenesis: insights into the mechanisms responsible for improved affinity. *J. Mol. Biol.* 332:195–204.

Pastor, M., F. Gago, and G. Cruciani. 2000. Comparative binding energy (COMBINE) analysis on a series of glycogen phosphorylase inhibitors: comparison with GRID/GOLPE methods. *In* Molecular Modeling and Prediction of Bioactivity. K. Gundertofter and F. S. Jorgensen, editors. Kluwer, New York. 329–330.

Perez, C., M. Pastor, A. R. Ortiz, and F. Gago. 1998a. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J. Med. Chem.* 41:836–852.

Perez, C., M. Pastor, A. R. Ortiz, and F. Gago. 1998b. Comparative binding energy analysis of HIV-1 protease inhibitors: incorporation of solvent effects and validation as a powerful tool in receptor-based drug design. *J. Med. Chem.* 41:836–852.

Pielak, G. J., and X. Wang. 2001. Interactions between yeast iso-1-cytochrome c and its peroxidase. *Biochemistry.* 40:422–428.

Roisman, L., J. Piehler, J. Trosset, H. A. Scheraga, and G. Schreiber. 2001. Structure of the interferon-receptor complex determined by distance constraints from double-mutant cycles and flexible docking. *Proc. Natl. Acad. Sci. USA.* 98:13231–13236.

Schreiber, G., and A. R. Fersht. 1993. Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering. *Biochemistry.* 32:5145–5150.

Schreiber, G., and A. R. Fersht. 1995. Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles. *J. Mol. Biol.* 248:478–486.

Schreiber, G., C. Frisch, and A. R. Fersht. 1997. The role of Glu73 of barnase in catalysis and the binding of barstar. *J. Mol. Biol.* 270:111–122.

Selzer, T., S. Albeck, and G. Schreiber. 2000. Rational design of faster associating and tighter binding protein complexes. *Nat. Struct. Biol.* 7:537–541.

Sheinerman, F. B., and B. Honig. 2002. On the role of electrostatic interactions in the design of protein-protein interfaces. *J. Mol. Biol.* 318:161–177.

Tomic, S., and B. Kojic-Prodic. 2002. A quantitative model for predicting enzyme enantioselectivity: application to *Burkholderia cepacia* lipase and 3-(aryloxy)-1,2-propanediol derivatives. *J. Mol. Graph. Model.* 21: 241–252.

Tomic, S., L. Nilsson, and R. C. Wade. 2000. Nuclear receptor-DNA binding specificity: a COMBINE and Free-Wilson QSAR analysis. *J. Med. Chem.* 43:1780–1792.

Vaughan, C. K., A. M. Buckle, and A. R. Fersht. 1999. Structural response to mutation at a protein-protein interface. *J. Mol. Biol.* 286:1487–1506.

Vijayakumar, M., K.-Y. Wong, G. Schreiber, A. R. Fersht, A. Szabo, and H.-X. Zhou. 1998. Electrostatic enhancement of diffusion-controlled protein-protein association: Comparison of theory and experiment on barnase and barstar. *J. Mol. Biol.* 278:1015–1024.

Vijayakumar, M., and H.-Y. Zhou. 2001. Salt bridges stabilize the folded structure of barnase. *J. Phys. Chem. B.* 105:7334–7340.

Vriend, G. 1990. WHAT IF: a molecular modeling and drug design program. *J. Mol. Graph.* 8:52–56.

Wade, R. C. 2001. Derivation of QSARs using 3D structural models of protein-ligand complexes by COMBINE analysis. *In* Rational Approaches to Drug Design: 13th European Symposium on Quantitative Structure-Activity Relationships. H.-D. Holtje and W. Sippl, editors. Prous Science, Barcelona, Spain. 23–28.

Wang, T., and R. C. Wade. 2001. Comparative binding energy (COMBINE) analysis of influenza neuraminidase-inhibitor complexes. *J. Med. Chem.* 44:961–971.

Wang, T., and R. C. Wade. 2002. Comparative binding energy (COMBINE) analysis of OppA-peptide complexes to relate structure to binding thermodynamics. *J. Med. Chem.* 45:4828–4837.

Wells, J. A., and B. C. Cunningham. 1993. Comparison of a structural and a functional epitope. *J. Mol. Biol.* 234:554–563.

Xu, D., S. L. Lin, and R. Nussinov. 1997. Extending the applicability of the non-linear Poisson-Boltzmann equation: multiple dielectric constants and multivalent ions. *J. Mol. Biol.* 265:68–84.

Zhang, J., I. Simeonowa, Y. Wang, and W. Sebald. 2002. The high-affinity interaction of human IL-4 and the receptor alpha chain is constituted by two independent binding clusters. *J. Mol. Biol.* 315:399–407.