

Research article

Open Access

## In silico screening of mutational effects on enzyme-proteic inhibitor affinity: a docking-based approach

Daniele Dell'Orco<sup>1,2</sup>, Pier Giuseppe De Benedetti<sup>1</sup> and Francesca Fanelli<sup>\*1,2</sup>

Address: <sup>1</sup>Department of Chemistry, University of Modena and Reggio Emilia, Italy and <sup>2</sup>Dulbecco Telethon Institute, via Campi 183 41100 Modena, Italy

Email: Daniele Dell'Orco - dellorco@unimo.it; Pier Giuseppe De Benedetti - deben@unimo.it; Francesca Fanelli\* - fanelli@unimo.it

\* Corresponding author

Published: 8 June 2007

Received: 23 January 2007

BMC Structural Biology 2007, 7:37 doi:10.1186/1472-6807-7-37

Accepted: 8 June 2007

This article is available from: <http://www.biomedcentral.com/1472-6807/7/37>

© 2007 Dell'Orco et al; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

**Background:** Molecular recognition between enzymes and proteic inhibitors is crucial for normal functioning of many biological pathways. Mutations in either the enzyme or the inhibitor protein often lead to a modulation of the binding affinity with no major alterations in the 3D structure of the complex.

**Results:** In this study, a rigid body docking-based approach has been successfully probed in its ability to predict the effects of single and multiple point mutations on the binding energetics in three enzyme-proteic inhibitor systems. The only requirement of the approach is an accurate structural model of the complex between the wild type forms of the interacting proteins, with the assumption that the architecture of the mutated complexes is almost the same as that of the wild type and no major conformational changes occur upon binding. The method was applied to 23 variants of the ribonuclease inhibitor-angiogenin complex, to 15 variants of the barnase-barstar complex, and to 8 variants of the bovine pancreatic trypsin inhibitor- $\beta$  Trypsin system, leading to thermodynamic and kinetic estimates consistent with in vitro data. Furthermore, simulations with and without explicit water molecules at the protein-protein interface suggested that they should be included in the simulations only when their positions are well defined both in the wild type and in the mutants and they result to be relevant for the modulation of mutational effects on the association process.

**Conclusion:** The correlative models built in this study allow for predictions of mutational effects on the thermodynamics and kinetics of association of three substantially different systems, and represent important extensions of our computational approach to cases in which it is not possible to estimate the absolute free energies. Moreover, this study is the first example in the literature of an extensive evaluation of the correlative weights of the single components of the ZDOCK score on the thermodynamics and kinetics of binding of protein mutants compared to the native state.

Finally, the results of this study corroborate and extend a previously developed quantitative model for in silico predictions of absolute protein-protein binding affinities spanning a wide range of values, i.e. from -10 up to -21 kcal/mol.

The computational approach is simple and fast and can be used for structure-based design of protein-protein complexes and for in silico screening of mutational effects on protein-protein recognition.

## Background

Among biological macromolecules, enzymes play a crucial role in every cell as catalysts of virtually any biochemical reaction. Kinetics and binding equilibria of enzyme-substrate and enzyme-proteic inhibitor interactions represent the molecular basis of the complex regulatory mechanisms of biochemical pathways.

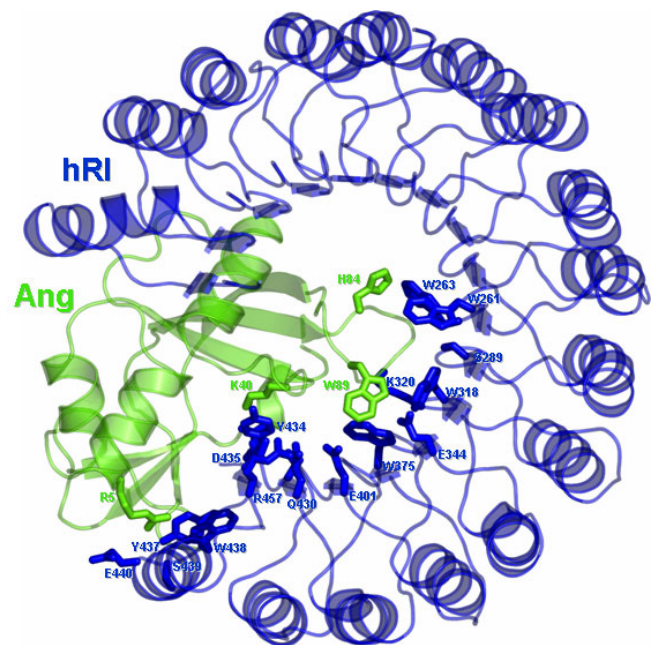
Enzyme-substrate and enzyme-inhibitor constitute the tightest protein-protein complexes [1], i.e. characterized by very low binding free energies ( $\Delta G^\circ$ ). Comparable high affinities characterize the inter-subunit interactions in some protein quaternary structures (i.e. grow factors, multi-domain proteins etc.) [1].

The ability to modulate the binding affinity in enzyme-proteic inhibitor interactions is of high interest, both for probing the molecular determinants involved in recognition and stabilization of the protein-protein complex, and for unravelling the molecular mechanisms that underlie the early onset of pathological conditions (see for instance Refs. [2,3]). Naturally occurring or artificially induced mutations in either the enzyme or the inhibitor protein represent a convenient way to modulate the binding affinity without altering significantly the three dimensional (3D) structure of the proteins.

Recently, we have developed a rigid-body docking-based approach for estimating the effects of point mutations on the thermodynamics and the kinetics of protein reconstitution [4], and protein-nucleic acid binding [5]. Indeed, we found that, under the condition of an exhaustive sampling of the roto-translational space of one protein with respect to the other, the scoring function (ZD-s) of the ZDOCK2.3 protein docking algorithm [6] has the potential of an empirically determined free energy function for protein-protein and protein-DNA interactions, where no major conformational changes occur upon binding [4]. The fundamental requirement of the approach is an accurate structural model of the complex between the wild type forms of the interacting proteins. The variants (i.e. mutations or deletions) of either one or both the partners can be achieved by molecular modelling. Docking simulations on the wild type forms of the two interacting proteins extracted from the X-ray structure of the complex are bound-bound docking cases. In contrast, docking simulations, in which the modelled mutations concern only one or both the interacting partners, should be moderately assimilated, respectively, to bound-unbound and unbound-unbound docking cases. This is particularly true when mutations involve multiple positions that are essential components of the interface. The basic assumption of the approach is that the architecture of the mutated complexes is almost the same as that of the wild type and no major conformational changes occur upon binding.

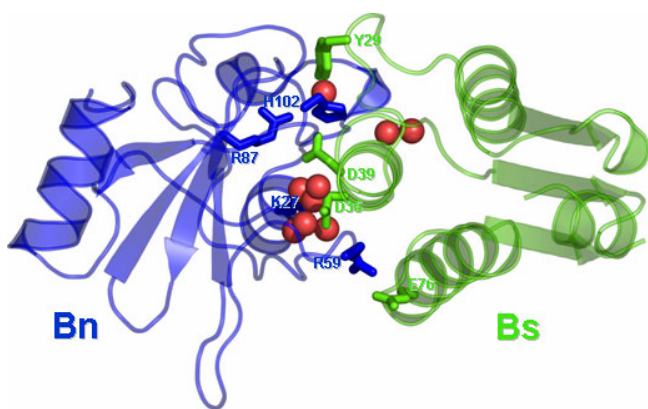
In this study, we extend our protocol to three substantially different cases of enzyme-inhibitor recognition, i.e. the human ribonuclease inhibitor-angiogenin (hRI-Ang), the barnase-barstar (Bn-Bs) and the bovine pancreatic trypsin inhibitor- $\beta$  Trypsin (BPTI- $\beta$ -Tryp) complexes (Figures 1, 2, 3). The effects of 23 and 15 different modifications (i.e. point mutations or deletions) on the thermodynamics and the kinetics of hRI-Ang and Bn-Bs binding, respectively, have been determined by a number of in vitro experiments (Tables 1, 2, 3) [7-13]. Differently from the hRI-Ang and Bn-Bs systems, in which point mutations are located in either one or both the interacting proteins, in the case of BPTI- $\beta$ -Tryp, a single amino acid, i.e. K15 in BPTI (K15<sup>BPTI</sup>), was replaced by eight different residues, each substitution exerting a remarkable influence on the binding equilibrium (Table 4).

In this study, we reconstituted all the variant forms (Tables 1, 2, 3) of hRI-Ang, Bn-Bs and BPTI- $\beta$ -Tryp by rigid-body docking, and used this structure-based approach to build a robust quantitative model for determining the residues most relevant for binding, i.e. the so called hot-spots [14].



**Figure 1**

Cartoon representation of the 3D structure of the hRI-Ang complex. Residues target of mutagenesis are represented in sticks. Here, as in the following drawings, the protein that was kept fixed in docking simulations (i.e. the target) is coloured in blue, whereas the one sampling the rotational and translational space (i.e. the probe) is coloured in green. Drawings were prepared with the software Pymol [52].



**Figure 2**

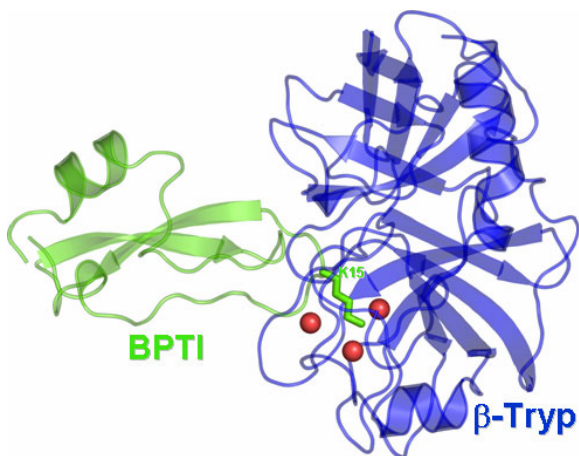
Cartoon representation of the 3D structure of Bn-Bs complex. Residues target of mutagenesis are represented in sticks. Here, as in the following drawings water molecules explicitly included in docking simulations are represented by red spheres.

The results of this study suggest that our computational protocol can be useful for *in silico* mutational analysis of enzyme-inhibitor interactions, where sufficient structural information on the wild-type forms of the interacting proteins is available.

## Results

### A docking score-based correlative approach for fast estimations of $\Delta G^\circ$ in protein-protein interactions

In a recent work we employed rigid body docking simulations for reconstituting nine protein-protein complexes, sharing no structural similarity and characterized by a broad range of  $\Delta G^\circ$ , i.e. from -10 up to -17 kcal/mol [4].



**Figure 3**

Cartoon representation of the 3D structure of  $\beta$ -Tryp-BPTI complex. Lysine 15 is the residue target of multiple mutagenesis, and is represented in sticks.

We found that averaging the docking scores of all the native-like solutions from independent docking runs provides an index, i.e. ZD-s, that is linearly correlated with  $\Delta G^\circ$  [4]. Indeed, we could obtain the empirical equation:

$$\Delta G^\circ = -0.37\text{ZD-s} + 2.9 \quad (R = 0.94; p = 0.00015; N = 9) \quad (1)$$

that was successfully employed to predict the thermodynamics and kinetics of calbindin D9k reconstitution from complementary wild type and mutated fragments [4].

The quality of the correlation and the lack of structural similarity between the training and the test sets led us to the conclusion that this docking-based approach has a general applicability under the condition that no major conformational change occur upon binding. This latter requirement arises from the rigid body approximation and is related to the neglect of conformational and cratic entropy changes upon binding [15].

The cases reported in this study indeed represent an interesting extension of the approach to the important class of enzyme-inhibitor interactions and allowed us to probe the accuracy and the applicability of the method for general protein design purposes.

First of all we probed the ability of the average docking score to discriminate the native-like solutions (i.e. the solutions characterized by a  $C_\alpha$ -Root Mean Square Deviation ( $C_\alpha$ -RMSD) lower than 1.0 Å from the native complex) from the false positives. Thus, we subjected to cluster analysis the 4000 solutions provided by each run, by using an algorithm described previously [4]. A  $C_\alpha$ -RMSD threshold of 1.5 Å was employed. Benchmarks were carried out on the three protein-protein complexes under study in their native forms, because the crystal structures of such complexes are known. Next we computed the docking scores averaged over all the members of each cluster. Interestingly, for each tested system and for each independent docking run, the cluster characterized by the highest average docking score was indeed that of the native-like solutions. The threshold of 1.5 Å established a clean cut between the cluster of native-like solutions, i.e. the best scored one, and all the other clusters. In general, this happens less frequently with the single docking scores as the highest one does not always correspond to a native-like solution. Hence, the average ZD-s performs better than the single value in distinguishing native-like solutions from false positives.

This was all the more reason for using the average ZD-s instead of the single one in the correlative models (see below in the text).

**Table 1: Human Ribonuclease Inhibitor (hRI) – Angiogenin (Ang) interaction: thermodynamic and kinetic data and ZD scores**

hRI	Ang	log $k_d^a$	log $k_a^b$	$\Delta G^\circ^c$ (kcal/mol)	$\Delta\Delta G^d$ (kcal/mol)	$N_{sol}^e$	Best Rank <sup>f</sup>	ZD <sub>best</sub> <sup>g</sup>	ZD-s <sup>h</sup>	ZD <sub>sc</sub> <sup>i</sup>	ZD <sub>el</sub> <sup>j</sup>	ZD <sub>des</sub> <sup>m</sup>
Wt <sup>n</sup>	wt	-6.96	8.30	-20.8	-	43	1	82.4	64.2	47.0	17.1	0.1
des (S460) <sup>n,o</sup>	R5A	-3.80	7.72	-15.7	5.1	38	3	74.5	61.0	45.1	14.7	1.2
des (S460) <sup>n</sup>	K40G	-3.14	7.87	-15.0	5.8	38	1	74.8	58.8	44.5	13.5	0.8
Q430A/V432A <sup>n</sup>	wt	-7.08	8.15	-20.8	0.1	45	1	81.7	63.1	46.0	17.1	0.0
W438A/S439A/E440A <sup>n</sup>	wt	-5.77	8.11	-18.9	1.9	45	1	81.8	63.9	45.9	17.3	0.6
R457A <sup>n</sup>	wt	-7.24	8.18	-21.0	-0.2	46	1	82.2	63.6	46.4	17.0	0.2
I459A <sup>n</sup>	wt	-6.70	8.08	-20.2	0.7	47	1	82.0	63.1	46.0	17.3	-0.2
wt <sup>n</sup>	H84A	-6.85	8.28	-20.6	0.2	41	1	80.4	63.9	46.7	17.1	0.1
wt <sup>n</sup>	W89A	-6.80	8.30	-20.6	0.2	44	1	78.7	61.1	44.4	17.4	-0.7
W261A <sup>n</sup>	wt	-7.00	8.20	-20.7	0.1	41	1	81.8	63.2	45.5	17.3	0.4
W263A <sup>n</sup>	wt	-6.16	8.26	-19.7	1.2	41	1	76.0	61.9	45.1	17.4	-0.6
S289A <sup>n</sup>	wt	-7.20	8.04	-20.8	0.0	44	1	82.7	64.4	47.3	17.0	0.1
W318A <sup>n</sup>	wt	-5.89	8.30	-19.3	1.5	37	1	76.6	62.0	45.7	17.4	-1.1
K320A <sup>n</sup>	wt	-7.03	8.46	-21.1	-0.3	45	1	81.9	63.6	46.0	16.8	0.8
E344A <sup>n</sup>	wt	-7.01	8.15	-20.7	0.2	43	1	81.9	63.0	45.9	16.8	0.3
W375A <sup>n</sup>	wt	-6.51	8.00	-19.8	1.0	48	1	77.4	61.8	44.9	17.7	-0.8
E401A <sup>n</sup>	wt	-6.49	8.15	-20.0	0.9	45	1	81.5	63.5	46.1	17.1	0.3
W261A/W263A/W318A <sup>n</sup>	wt	-3.51	7.96	-15.6	5.2	31	11	68.0	57.9	42.5	17.6	-2.2
W261A/W263A/W318A <sup>n</sup>	R5A	-1.39	7.89	-12.6	8.2	28	39	63.7	55.3	40.8	15.3	-0.8
W261A/W263A/W318A <sup>n</sup>	K40G	-0.96	7.59	-11.7	9.2	24	58	61.1	53.9	41.5	14.0	-1.6
Y434A <sup>p</sup>	wt	-5.14	7.88	-17.7	3.1	45	1	78.1	59.8	43.1	17.2	-0.5
D435A <sup>p</sup>	wt	-4.85	8.00	-17.5	3.3	44	1	77.5	59.6	44.6	14.1	0.9
Y437A <sup>p</sup>	wt	-6.51	8.30	-20.2	0.6	41	1	76.8	60.5	43.3	17.5	-0.3
des (S460) <sup>p</sup>	wt	-6.20	8.28	-19.7	1.1	44	1	81.3	63.5	46.0	17.2	0.3

<sup>a</sup>decimal logarithm of the dissociation rate constant  $k_d$  (s<sup>-1</sup>). Experimental value from Ref. [13].

<sup>b</sup>decimal logarithm of the association rate constant  $k_a$  (M<sup>-1</sup>s<sup>-1</sup>). Experimental value from Ref. [13].

<sup>c</sup>standard free energy of association as obtained from the relationship  $\Delta G^\circ = RT \ln K_D$ , where  $K_D = k_d/k_a$ .

<sup>d</sup>difference in binding free energy for the wild type and variant complexes, calculated from the equation  $\Delta\Delta G = -RT \ln (K_{D,wt}/K_{D,var})$ .

<sup>e</sup>number of native-like solutions among the ensemble of 12000 solutions (i.e. derived from three independent docking runs each ranking 4000 solutions).

<sup>f</sup>rank number of the best native-like solution in the output list (i.e. 1 = highest scored ; 12000 = lowest scored).

<sup>g</sup>ZD scores of the best native-like solution in the output list

<sup>h</sup>ZD scores averaged over all the native-like complexes resulting from three independent runs; <sup>i</sup>shape complementarity term; <sup>j</sup>electrostatic term; <sup>m</sup> desolvation term. See *Methods*.

<sup>n</sup>experimental data from [13] (Tables 2, 3, and 5).

<sup>o</sup>recombinant hRI in which Ser460 is deleted. Nomenclature from Ref. [7] <sup>p</sup> experimental data from Ref. [7] (Table 4).

### **hRI-Ang interaction: effects of point mutations and residue deletion on binding affinity**

The interactions of hRI with both RnaseA and Ang are well characterized examples of tight protein-protein binding [7,13,16].

The binding affinities of hRI for either RnaseA or Ang are very high, i.e. -18.4 kcal/mol and -20.8 kcal/mol, respectively [7]. In vitro site-directed mutagenesis experiments targeting both hRI and Ang, resulted in changes in the binding affinities, ranging from less than 0.5 kcal/mol for single substitutions up to about 9 kcal/mol for multiple mutations (Table 1) [7,13].

The single or multiple point mutations of hRI and Ang are essentially located at the protein-protein interface, although their positions and features vary remarkably from case to case. Two of the four mutations of Ang, i.e.

R5A and K40G, indeed, belong to the interface and reduce the binding affinity because of the breakage of intermolecular salt bridges. In contrast, the remaining two mutants W89A and H84A are less involved in the interface and, consistently, do not substantially affect the binding affinity (Figure 1 and Table 1). As for the hRI mutations, the most detrimental effects on the binding affinity are associated with (a) breakages of H-bonding interactions, like in the case of the Des(S460)<sup>hRI</sup> deletion, (b) breakages of salt bridges, like in the case of the D435A mutant, or (c) loss of aromatic interactions, like in the cases of the W263A and W318A mutants (Figure 1 and Table 1).

We carried out docking simulations between hRI and Ang in their wild type and 23 variants (Table 1). The native structures of both the enzyme and the inhibitor were extracted from the crystal structure of the complex. In contrast, the variant forms, which include a deletion at the C-

**Table 2: Barnase (BN) – Barstar (BS) interaction: thermodynamic and kinetic data and ZD scores from docking simulations without interfacial water molecules**

BS	BN	log $k_d^a$	log $k_a$	$\Delta G^\circ$ (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	$N_{sol}$	Best Rank	ZD-s
wt <sup>b</sup>	wt	-5.43	8.57	-19.00	-	47	1	53.5
wt	H102A	-0.89	8.60	-12.90	6.10	49	1	50.3
Y29A	wt	-3.00	8.46	-15.60	3.40	51	1	49.9
Y29A	H102A	-0.82	8.56	-12.70	6.30	47	1	45.1
Y29F	wt	-5.62	8.48	-19.10	-0.10	47	1	51.3
Y29F	H102A	-1.35	8.59	-13.50	5.50	48	1	47.8
D39A	wt	-0.05	8.28	-11.30	7.70	45	1	48.0
D39A	H102A	1.23	8.64	-10.10	8.90	46	1	46.0
wt	R59A	-2.62	7.53	-13.80	5.20	45	12	44.7
wt	K27A	-2.35	7.71	-13.60	5.40	44	1	51.0
wt	R87A	-1.77	9.33	-13.50	5.50	47	1	51.3
D39A	R59A	n.a.	n.a.	-7.70	11.30	38	32	40.0
D39A	K27A	-0.17	7.76	-10.80	8.20	42	1	46.6
D39A	R87A	-0.52	8.20	-11.90	7.10	46	1	47.7
D35A	wt	-2.42	8.28	-14.50	4.50	48	1	48.9
E76A	wt	-4.68	8.30	-17.70	1.30	52	1	49.7

<sup>a</sup>Details and legend concerning the quantities reported herein were explained in Table 1

<sup>b</sup>Experimental data from [8] (Table 3) and [12] (Table 1)

terminus of hRI (i.e. Des(S460)<sup>hRI</sup>) and single and triple mutations in either one or both the interacting partners, were built by in silico mutation of the wild type form. The selected docking indices are reported in Table 1 together with the in vitro determined thermodynamic and kinetic data.

The results of docking simulations of hRI-Ang interaction are satisfactory for each tested variant. On average, 41 native-like solutions out of 12000 total solutions from three independent sets of docking runs were found considering all the variants (Table 1). In 20 out of the 24 tested cases, the native-like solutions comprise the best

one according to the docking score, i.e. solution N. 1 in the output list (Table 1).

The correlation between in vitro determined relative binding free energy values ( $\Delta\Delta G^\circ$ , see methods) and the relative change of ZD-s for the hRI-Ang systems is reported in Figure 4. The overall trend is clearly linear and the correlation is quantitatively meaningful ( $R = 0.92$ ,  $p < 0.0001$ ,  $N = 23$ ).

In vitro determined  $\Delta\Delta G^\circ$  values were also correlated with the relative changes in each of the three components of the relative ZD score, i.e. the shape and electrostatic com-

**Table 3: Barnase (BN) – Barstar (BS) interaction: thermodynamic and kinetic data and ZD scores from docking simulations with interfacial water molecules**

BS	BN	log $k_d^a$	log $k_a$	$\Delta G^\circ$ (kcal/mol)	$\Delta\Delta G$ (kcal/mol)	$N_{sol}$	Best Rank	ZD <sub>best</sub>	ZD-s	ZD <sub>sc</sub>	ZD <sub>el</sub>	ZD <sub>des</sub>
wt <sup>b</sup>	wt	-5.43	8.57	-19.00	-	50	1	65.5	55.6	41.6	12.0	2.0
wt	H102A	-0.89	8.60	-12.90	6.10	49	1	63.3	53.0	39.4	11.1	2.6
Y29A	wt	-3.00	8.46	-15.60	3.40	53	1	60.4	52.7	40.1	11.0	1.6
Y29A	H102A	-0.82	8.56	-12.70	6.30	50	1	59.5	50.7	38.1	10.6	2.0
Y29F	wt	-5.62	8.48	-19.10	-0.10	52	1	63.1	54.5	41.8	10.1	2.6
Y29F	H102A	-1.35	8.59	-13.50	5.50	48	1	61.7	51.3	38.6	10.6	2.2
D39A	H102A	1.23	8.64	-10.10	8.90	50	1	60.4	48.6	37.8	6.8	4.0
wt	K27A	-2.35	7.71	-13.60	5.40	48	1	61.5	52.6	40.2	9.7	2.7
D39A	K27A	-0.17	7.76	-10.80	8.20	48	1	59.3	48.3	38.6	5.1	4.5
D35A	wt	-2.42	8.28	-14.50	4.50	50	1	58.2	51.4	38.9	9.7	2.8
E76A	wt	-4.68	8.30	-17.70	1.30	49	1	60.6	53.1	42.2	8.5	2.4

<sup>a</sup>Details and legend concerning the quantities reported herein were explained in Table 1.

<sup>b</sup>Experimental data from [8] (Table 3) and [12] (Table 1)

**Table 4:  $\beta$ -Trypsin – BPTI interaction: thermodynamic data and ZD scores for X-ray determined and *in silico* modeled BPTI-Lys 15 mutants**

Mutation <sup>a</sup>	$\Delta G^\circ$ (kcal/mol) <sup>b</sup>	$N_{sol}^{xr,c}$	Best Rank <sup>xr</sup>	ZD <sub>best</sub>	ZD <sub>-s<sup>xr</sup></sub>	ZD <sub>sc<sup>xr</sup></sub>	ZD <sub>el<sup>xr</sup></sub>	ZD <sub>des<sup>xr</sup></sub>	ZD <sub>-s<sup>xr2,d</sup></sub>	$N_{sol}^{xr2}$	Best Rank <sup>xr2</sup>	ZD <sub>-s<sup>mod,e</sup></sub>	$N_{sol}^{mod}$	Best Rank <sup>mod</sup>
GLY	-5.73	63	9	51.7	43.2	34.8	-0.9	9.3	42.3	62	13	42.5	62	1
THR	-7.50	80	1	54.3	45.0	35.8	-1.2	10.4	44.4	79	1	45.5	70	1
ASP	-6.54	76	1	54.9	44.7	37.6	-2.2	9.3	44.8	71	1	44.2	78	1
MET	-10.36	79	1	56.8	47.1	36.7	-1.1	11.5	46.1	78	1	45.4	74	1
GLU	-8.59	73	1	58.7	47.7	38.8	-1.2	10.1	47.3	74	1	44.8	76	1
GLN	-8.73	89	1	55.5	44.8	36.4	-0.8	9.2	45.1	84	1	43.9	74	1
HIS	-9.27	88	1	62.5	48.7	38.6	0.2	9.9	49.4	87	1	47.0	77	1
PHE	-11.04	85	1	67.1	51.8	40.1	-0.1	11.8	51.6	85	1	49.0	71	1

<sup>a</sup>Residue in which BPTI-Lys 15 was mutated. The PDB entries corresponding to these variants are 3BTG, 3BTT, 3BTD, 3BTM, 3BTE, 3BTQ, 3BTH and 3BTF [28]

<sup>b</sup>Standard free energy of association as obtained from the relationship  $\Delta G^\circ = RT \ln K_D$ . Experimental data from Ref [10]

<sup>c</sup>Here, and in the following columns, *xr* refers to the results of docking runs from the experimentally resolved X-ray structures

<sup>d</sup>Here, and in the following columns, *xr2* refers to the results of docking runs from the experimentally resolved X-ray structures in the absence of interface water molecules

<sup>e</sup>Here, and in the following columns, *mod* refers to the results of docking runs from the *in silico* modelled structures of the same complexes

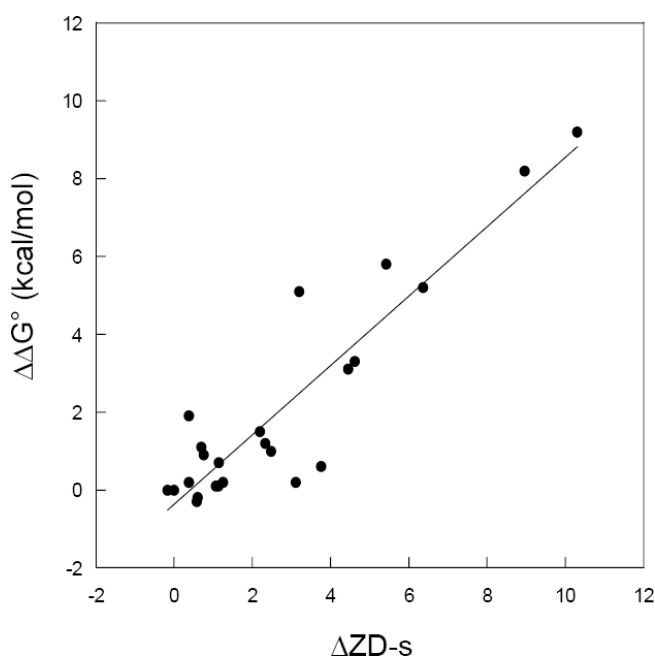
plementarities (ZD<sub>sc</sub> and ZD<sub>el</sub>, respectively) and desolvation (ZD<sub>des</sub>). All the possible combinations of such components were considered as well (Table 5). The most significant correlations concern that between  $\Delta\Delta G^\circ$  and  $\Delta ZD_{sc}$  ( $R = 0.82$ ,  $p < 0.0001$ ),  $\Delta ZD_{sc+el}$  ( $R = 0.93$ ,  $p < 0.0001$ ), or  $\Delta ZD_{el+des}$  ( $R = 0.90$ ,  $p < 0.0001$ ) (Table 5). These results suggest that combined changes in the shape and electrostatic complementarities are more important

contributors to  $\Delta\Delta G^\circ$  than the individual terms, and that changes in desolvation do not appreciably affect the thermodynamics of binding.

#### **Bn-Bs interaction: effects of point mutations on binding affinity**

The interaction of Bn with its inhibitor Bs is another example of high-affinity protein-protein association ( $\Delta G^\circ = -19$  kcal/mol), which has been widely studied in kinetic and thermodynamic detail [8,9,12,17-19].

The importance of electrostatic interactions in driving and stabilizing Bn-Bs association has been widely discussed in the past years [20-26]. In fact, Bs exposes negatively charged amino acids, i.e. D39 and E76, to positively charged amino acids on the Bn surface, i.e. K27, R59 and R87 (Figure 2). Alanine replacement of either one of these amino acids, induces breakages of interfacial salt bridges, therefore diminishing the electrostatic complementarity between the two proteins (Figure 2 and Table 2). *In vitro* experiments demonstrated that the association rate constant of Bn with Bs is approximately constant over the pH range of 4.5–9, whereas the dissociation rate constant increases by a factor of almost 100 when the pH of the solution is lowered from 7 to 5 [17]. A significant reduction in the pH dependence of the dissociation rate constant of the Bn-Bs complex occurs upon alanine substitution for H102<sup>Bn</sup>[17]. Collectively, these data suggest that the ionization state of H102 does not affect the association process, whereas it influences the dissociation kinetics. The increase in the dissociation rate constant with the lowering of the pH of the solution may be caused by protonation of H102<sup>Bn</sup>. In the crystal structure of the wild type Bn-Bs complex, the N $\epsilon$  nitrogen atom of H102<sup>Bn</sup> donates an H-bond to D39<sup>Bs</sup>, whereas the N $\delta$  nitrogen atom of the same histidine accepts an H-bond from the backbone amide group of G31<sup>Bs</sup>[27]. This interaction pat-

**Figure 4**

Experimental relative affinities ( $\Delta\Delta G^\circ$ ) versus relative ZD-s for hRI-Ang interaction. The fitting line equation is  $\Delta\Delta G^\circ = -0.37 + 0.89\Delta ZD-s$ , the correlation coefficient is  $R = 0.92$  and its probability  $p(R) < 0.0001$ . The number of experimental points is  $N = 23$ .

**Table 5: Decomposition of ZD-s in three components for the enzyme-proteic inhibitor complexes. Statistical analysis of the empirical correlations**

	ZD <sub>sc</sub>					ZD <sub>el</sub>				ZD <sub>des</sub>				ZD <sub>sc+el</sub>				ZD <sub>sc+des</sub>				ZD <sub>el+des</sub>			
	N <sup>a</sup>	A <sup>b</sup>	B <sup>b</sup>	R <sup>c</sup>	p(R) <sup>d</sup>	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)
HRI-ANG	24	35.5	-0.5	0.82	< 0.0001	10.2	-0.3	0.73	< 0.0001	-2.0	-0.1	0.34	0.105	45.6	-0.9	0.93	< 0.0001	33.4	-0.6	0.73	< 0.0001	8.1	-0.4	0.90	< 0.0001
BN-BS_1 <sup>e</sup>	11	56.4	-1.8	0.91	0.0001	-6.4	-0.8	0.56	0.073	-20.3	2.2	0.62	0.042	30.2	-0.9	0.87	0.0005	41.1	-1.3	0.61	0.045	-1.1	-1.1	0.46	0.150
BN-BS_2 <sup>f</sup>	11	56.8	-1.8	0.89	0.0002	-9.0	-0.6	0.36	0.282	-19.4	1.7	0.53	0.092	35.5	-1.0	0.81	0.0023	26.3	-1.0	0.50	0.1139	-11	-0.3	0.1	0.7628
BPTI-βtTRYPS	8	31.9	0.9	0.67	0.071	-2.9	0.3	0.57	0.137	6.6	0.6	0.79	0.021	29.0	1.2	0.75	0.032	38.5	1.5	0.82	0.013	3.7	0.9	0.85	0.008

<sup>a</sup>Number of analyzed interactions, i.e. three docking runs for each variant.

<sup>b</sup>Coefficients of the linear regression equation fitting the experimental data. In detail, the equation is in the form  $Y = A + BX$ , where  $Y$  is the experimental  $\Delta G^\circ$  expressed in kcal/mol,  $A$  (kcal/mol) is the intercept,  $B$  (kcal/mol) is the slope and  $X$  is the corresponding ZD-s component.

<sup>c</sup>Linear correlation coefficient.

<sup>d</sup>Linear correlation coefficient probability.

<sup>e</sup>Data referring to the protonated form of H102<sup>Bn</sup>.

<sup>f</sup>Data referring to the neutral form of H102<sup>Bn</sup>.



tern is not compatible with the protonated state of H102 and, consistently, the  $pK_a$  of such histidine in the complex is expected to be lower than 5 [17,24]. Protonation of H102<sup>Bn</sup> with decreases in the pH of the solution is, therefore, expected to destabilize the interaction pattern found in the Bn-Bs complex, thus favouring dissociation. The structural effects of H102A mutation, thus, include the breakage of two intermolecular H-bonds.

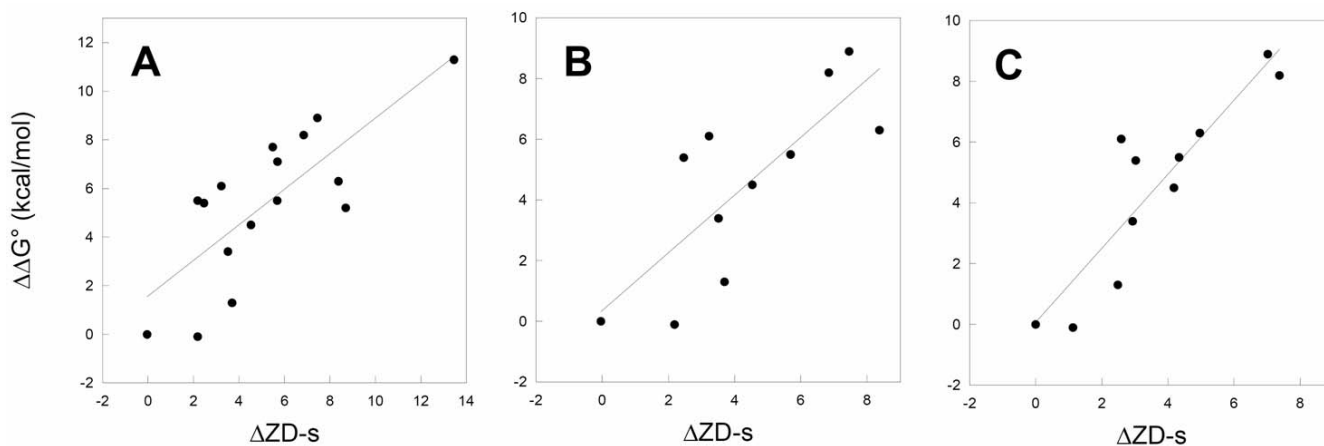
Alanine replacements of R59<sup>Bn</sup>, K27<sup>Bn</sup>, D35<sup>Bs</sup> also perturb H-bonding interactions with interfacial water molecules. In contrast, loss of van der Waals interactions rather than of H-bonding interactions seems to be responsible for the detrimental effect on the binding free energy exerted by the Y29A<sup>Bs</sup> mutation. This is suggested by the evidence that the Y29F<sup>Bs</sup> substitution has a marginal effect on  $\Delta G^\circ$  (Table 2).

We performed docking simulations between Bn and Bs in their wild type and 15 variants characterized by single mutations on either one or both the interacting proteins, selected for the availability of in vitro determined thermodynamic and kinetic data (Tables 2 and 3). The native structures of both the enzyme and the inhibitor were extracted from the crystal structure of the complex. In contrast, the mutants were achieved by in silico mutating the wild type structure. Consistent with simulations on the hRI-Ang system, docking simulations were in principle

carried out without interfacial water molecules. However, since interfacial water molecules are known to be important in Bn-Bs association as they fill regions of poor surface complementarity [27], docking simulations were also carried out by explicitly including water molecules in those cases in which water positions are known at acceptably high resolution (for deep detail, see the Methods section). In particular, these cases include the wild type and 10 Bn and/or Bs mutants (Table 3). Furthermore, both the neutral (i.e. with the hydrogen atom on N $\epsilon$ ) and charged forms of H102<sup>Bn</sup> were considered.

Figure 5A reports a plot of the experimental  $\Delta\Delta G^\circ$  versus the relative change in ZD-s for the docking simulations without interfacial water molecules and with the protonated form of H102<sup>Bn</sup>. While a trend is clearly observed ( $R = 0.77$ ,  $p = 0.00046$ ,  $N = 16$ ), the correlation is quantitatively less striking than that achieved for the hRI-Ang system (Figure 4). Simulations with the neutral form of H102<sup>Bn</sup> gave worse correlative models (i.e.  $R = 0.65$ ,  $p = 0.006$ ,  $N = 16$ ).

Comparisons of the correlative models achieved from docking simulations without and with interfacial water molecules for the 11 molecular systems, for which water positions could be defined at acceptably high resolution, showed a remarkable improvement in the correlative models based upon inclusion of explicit water molecules



**Figure 5**

Experimental relative affinities ( $\Delta\Delta G^\circ$ ) versus relative ZD-s for Bn-Bs interaction. (A) Plot referred to data reported in Table 2, i.e. 16 variants of Bn-Bs without water molecules at the interface and protonated form of H102A<sup>Bn</sup>. The fitting line equation is  $\Delta\Delta G^\circ = 1.57 + 0.73\Delta ZD-s$ ,  $R = 0.77$ ,  $p(R) = 0.00046$ ,  $N = 16$ , where R is the correlation coefficient,  $p(R)$  is the probability of such coefficient and N is the number of points. (B) Correlative model derived from the one at point (A) by leaving out four points. The dataset in this plot is limited to the 11 variants of Bn-Bs, for which water positions could be defined at acceptably high resolution (Table 3). The correlation equation and its parameters are:  $\Delta\Delta G^\circ = 0.35 + 0.95\Delta ZD-s$ ,  $R = 0.79$ ,  $p(R) = 0.0041$ ,  $N = 11$ . (C) Same data set as in B, but with ZD-s derived by docking simulations with explicit interfacial water molecules and H102<sup>Bn</sup> in its protonated state. The correlation equation and its parameters are, respectively:  $\Delta\Delta G^\circ = 0.06 + 1.22\Delta ZD-s$ ,  $R = 0.90$ ,  $p(R) = 0.00017$ ,  $N = 11$ .



(Figures 5B and 5C). Indeed, the correlation coefficient rises from 0.79 to 0.90 following inclusion of water molecules (Figures 5B and 5C).

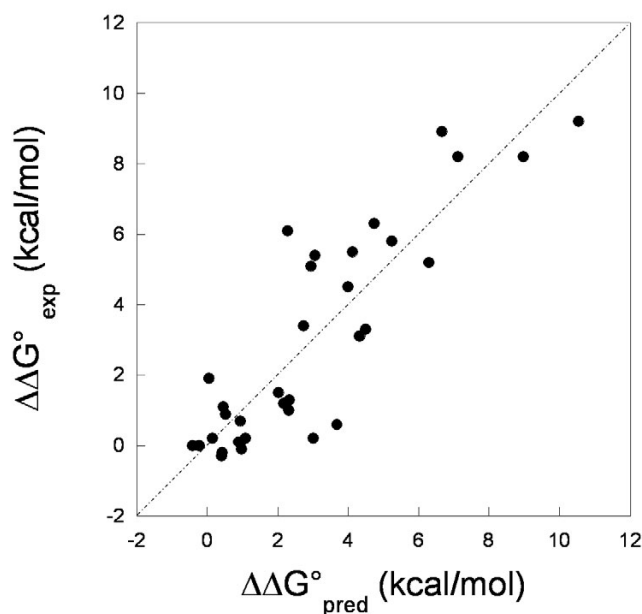
Differently from the results of docking simulations without explicit waters, in the case of simulations with explicit waters, the employment of the neutral form of H102<sup>Bn</sup> does not change substantially the correlative models compared to protonated H102<sup>Bn</sup> ( $R = 0.89$ ,  $N = 11$ ). The ZD-s achieved with protonated and non-protonated H102<sup>Bn</sup> are, indeed, highly correlated ( $R = 0.97$ ).

In general, independent of the presence or absence of water molecules and of the prototropic form of H102<sup>Bn</sup> (i.e. charged or neutral), for each Bn-Bs variant, the best scored solution in the output list (i.e. solution N. 1) was a native-like (Tables 2 and 3).

Decomposition of the ZD-s was carried out only for the cases that produced the best correlative models, i.e. the 11 Bn-Bs variants shown in Table 3, simulated in the presence of explicit water molecules. No significant improvement was observed by using the single components compared to the whole ZD-s (Table 5). More precisely, the improvement on using  $\Delta ZD_{sc}$  was marginal (i.e.  $R = 0.91$  with  $\Delta ZD_{sc}$  versus  $R = 0.90$  with the whole score). The correlations with  $\Delta\Delta G^\circ$  remained acceptable when the  $\Delta ZD_{sc+el}$  is employed ( $R = 0.87$ ,  $p = 0.0005$ ), whereas significantly worse linear trends were observed on using desolvation taken either singularly or paired with the other two components (Table 5). The data reported above for the correlative analyses with the ZD-s components refer to docking simulations with protonated H102. The employment of neutral H102 gave very similar results (Table 5).

#### Relative binding affinity predictions by merging data from the hRI-Ang and the Bn-Bs systems

As shown in the previous sections, we found convincing correlations between  $\Delta\Delta G^\circ$  and the change in the total docking score ( $\Delta ZD$ -s) for both hRI-Ang and Bn-Bs complexes (see Figures 4 and 5). Since such quantity is by definition relative, it was possible to perform a global correlative analysis by merging the two sets of data in a unique set of 34 cases, including 23 data for the hRI-Ang system and 11 data for the Bn-Bs one. The latter concerns only docking simulations with explicit water molecules (i.e. the data set in Table 3). The thereof obtained equation has been used to predict the association free energy changes caused by mutations or deletions in either the enzyme or the inhibitor, or both, by a leave-one-out approach (Figure 6). The trend is significantly linear, the fitting line slope is about 1, and the coefficient  $R$  as well as its probability have good values ( $R = 0.87$ ,  $p < 0.0001$ ,  $N = 34$ ).



**Figure 6**  
Experimental ( $\Delta\Delta G^\circ_{exp}$ ) versus predicted ( $\Delta\Delta G^\circ_{pred}$ ) relative affinities for hRI-Ang and Bn-Bs data, analyzed as a unique set. The predicted values refer to a leave-one-out test. The fitting line equation is  $\Delta\Delta G^\circ_{exp} = 0.04 + 0.98\Delta\Delta G^\circ_{pred}$ ,  $R = 0.87$ ,  $p(R) < 0.0001$  and  $N = 34$ .

#### $\beta$ -Trypsin-BPTI interaction: effects of multiple replacements of K15<sup>BPTI</sup> on the binding equilibrium

The last case considered in this study, i.e. the  $\beta$ -Tryp-BPTI interaction, differs from the other two in that here a single amino acid, i.e. K15<sup>BPTI</sup>, was subjected to eight different substitutions, whose effect on the association equilibrium constant  $K_A$ , i.e.  $\Delta G^\circ$ , has been determined in vitro (Table 4) [10].

In the wild type structure, the residue target of mutagenesis, K15<sup>BPTI</sup>, protrudes towards a  $\beta$ -Tryp cavity, being highly packed in the binding site (Figure 3). Its functional importance is highlighted also by its central role in coordinating the overall interaction between the two proteins. Indeed, its protonated nitrogen atom is involved in H-bonds with two water molecules and a salt bridge with D189 <sup>$\beta$ -Tryp</sup> (Figure 3). The latter interaction is considered fundamental for the inhibitory ability of BPTI [28,29].

The heterogeneous physico-chemical nature of the perturbations introduced by the K15<sup>BPTI</sup> substitutions is reflected by the broad variation of  $K_A$ , i.e. about 4 pK<sub>A</sub> units (Table 4).

Docking simulations were carried out between  $\beta$ -Tryp and eight mutants of K15<sup>BPTI</sup>. Three different molecular systems were considered, characterized by decreasing detail and resolution. The most complete and resolved one consists of the mutants and the relevant interfacial water molecules extracted from the X-ray structures of the complexes with  $\beta$ -Tryp (see Methods for details). The second molecular system, characterized by intermediate resolution, differs from the first one for the lack of interfacial water molecules. Finally, the third and less resolved molecular system is characterized by the lack of interfacial water molecules and by *in silico* built amino acid substitutions for K15.

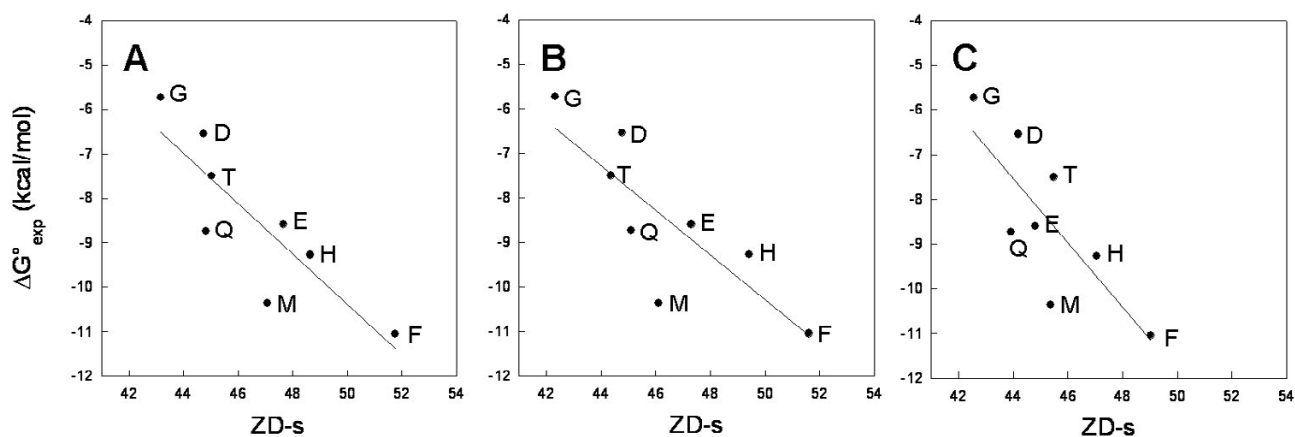
As for docking simulations by using the X-ray structures of the K15<sup>BPTI</sup> mutants and interfacial water molecules, in each case, the best ranked solution (i.e. solution N. 1) is a native-like, except for K15G (Table 4). For each mutant, the average number of native-like solutions out of the total 12000 is about 79.

A linear trend was obtained by plotting  $\Delta G^\circ$  versus the ZD-s ( $R = 0.86$ , and  $p = 0.006$ ,  $N = 8$ ; Figure 7A). Predicted versus experimental  $\Delta G^\circ$ 's through a leave-one-out test led to accurate estimations ( $R = 0.80$ ;  $p(R) = 0.017$ ,  $N = 8$ ), which significantly improve if the K15Q substitution is omitted ( $R = 0.85$ ,  $p(R) = 0.016$ ,  $N = 7$ ) [15].

Interestingly, the goodness of the correlative models decreases with the resolution of the structural models

(Figure 7). The first slight worsening is observed on neglecting interfacial water molecules in docking simulations ( $R = 0.86$  and  $R = 0.83$  with and without water molecules, respectively, Figures 7A and 7B). A further slight worsening of the correlative models is observed with docking simulations carried out on *in silico* modelled mutant side chains, neglecting interfacial water molecules ( $R = 0.79$ ,  $p = 0.0019$ ,  $N = 8$ , Figure 7C). The three linear trends are, however, very similar and share the same outliers, i.e. K15M and K15Q mutants (Figure 7). Although these results support the common knowledge that the performance of the docking algorithms improves with the resolution of the structural models, the results of this comparative study suggest that, at least for the K15<sup>BPTI</sup> mutants, the performance of the docking algorithm is good even in the case of *in silico* modelled mutants. It is worth noting, indeed, that, if both K15M and K15Q mutants are deleted, correlations improve significantly for all the three molecular systems, i.e. the ones with highest, intermediate, and lowest resolution (i.e. the correlation coefficients become 0.99, 0.98, and 0.95, respectively). For the K15Q mutant, the side chain conformation of the mutated amino acid is ill-defined even in the crystal structure. Leaving out only this mutant, the correlation for the three systems significantly improves as well (i.e. the correlation coefficients become 0.90, 0.85, and 0.85, respectively).

Dissecting the components of the relative ZD-s derived from the runs on the highest resolved molecular systems



**Figure 7**

Plot of experimental  $\Delta G^\circ$  versus ZD-s for the  $\beta$ -Tryp-BPTI complex, where K15<sup>BPTI</sup> was substituted in eight different amino acids named by their one-letter code. (A) Results of docking simulations starting from the X-ray structures of the eight mutants, including explicit water molecules. The fitting line equation is  $\Delta G^\circ = 17.9 - 0.57ZD-s$ ,  $R = 0.86$ ,  $p(R) = 0.006$  and  $N = 8$ . (B) Correlation model derived by docking simulations on the same molecular models as in (A) but without explicit water molecules. The fitting line equation is  $\Delta G^\circ = 14.8 - 0.50ZD-s$ ,  $R = 0.82$ ,  $p(R) = 0.012$  and  $N = 8$ . (C) Correlation model derived from docking simulations starting from the *in silico*-modelled structures of the eight mutants, with no interface water molecules. The fitting line equation is  $\Delta G^\circ = 24.2 - 0.72ZD-s$ ,  $R = 0.79$ ,  $p(R) = 0.019$  and  $N = 8$ .

shows that the most significant correlations with  $\Delta G^\circ$  concern  $ZD_{el+des}$  ( $R = 0.85$ ,  $p = 0.008$ ),  $ZD_{sc+des}$  ( $R = 0.82$ ,  $0.013$ ), or  $ZD_{des}$  ( $R = 0.79$ ,  $p = 0.021$ ) (Table 5). Unlike the previous two cases, desolvation seems to play a relevant role in modulating the effects of K15<sup>BPTI</sup> mutations on the binding free energy.

#### **Kinetics of enzyme-inhibitor interaction and docking simulations: correlative analysis**

Kinetic data (i.e. the association ( $k_a$ ) and the dissociation ( $k_d$ ) rate constants) are available for each variant of the hRI-Ang and Bn-Bs complexes analyzed in the present study (Tables 1 and 3) [12,13]. For the hRI-Ang system, the association rate constants vary about one order of magnitude, whereas, for the Bn-Bs system, they vary two orders of magnitude (Tables 1, 2, 3). Different from the  $k_a$ 's, the  $k_d$ 's vary to a larger extent, i.e. about six orders of magnitude for hRI-Ang and seven orders of magnitude for Bn-Bs (Tables 1, 2, 3).

For both the hRI-Ang and Bn-Bs systems, changes in the dissociation kinetics, i.e.  $k_d$ , are highly correlated with changes in the equilibrium constant, i.e.  $K_A$  (data not shown).

We searched for empirical correlations between ZD-s and the rate constants, considering also all the combinations of ZD-s components. The results of the analyses are reported in Table 6. For the hRI-Ang system, the most significant correlations with  $k_a$  were achieved by the total ZD-s ( $R = 0.71$ ,  $p = 0.0001$ ,  $N = 24$ ) and  $ZD_{el+des}$  ( $R = 0.76$ ,  $p < 0.0001$ ,  $N = 24$ ; Table 6).

In contrast, for the Bn-Bs system, no significant correlation could be found between  $k_a$  and any combination of the ZD-s components (Table 6).

Interestingly, significant correlations were found between the total ZD-s and  $k_d$  for both hRI-Ang ( $R = 0.92$ ,  $p < 0.0001$ ,  $N = 24$ ) and Bn-Bs ( $R = 0.88$ ,  $p = 0.0003$ ,  $N = 11$ , Table 6). Dissecting the three components of the ZD score shows that, for hRI-Ang, the most significant correlations with  $k_d$  are given by  $ZD_{sc+el}$  ( $R = 0.93$ ,  $p < 0.0001$ ) and  $ZD_{el+des}$  ( $R = 0.90$ ,  $p < 0.0001$ ), whereas, for Bn-Bs, the most significant correlations are given by  $ZD_{sc}$  ( $R = 0.93$ ,  $p = 0.0001$ ,  $N = 11$ ) and  $ZD_{sc+el}$  ( $R = 0.84$ ,  $p = 0.0012$ ,  $N = 11$ , Table 6).

The statistic parameters reported above for Bn-Bs concern simulations with protonated H102<sup>Bn</sup>. The employment of neutral H102 gave slightly worsened correlations (Table 6).

#### **An extended correlative model for absolute binding free energy predictions in protein-protein interactions**

In this study, we reconsidered the quantitative model previously obtained on 9 highly heterogeneous protein systems (see equation (1)), by incorporating the new in vitro and in silico data concerning wild type hRI-Ang. This upgrade extends the binding affinity range by more than 2 kcal/mol (Figure 8 and Ref. [4]). Furthermore, the old ZD-s, concerning docking simulations on wild type Bn-Bs without water molecules, was replaced by the average score from simulations with explicit water molecules. The  $\beta$ -Tryp-BPTI wild type system could not be included in the training set due to ambiguities in the in vitro data [30,31]. The extended equation:

$$\Delta G^\circ = -0.37ZD-s + 3.13 \quad (R = 0.97; p < 0.00015; N = 10) \quad (2)$$

is strikingly similar to the previous one (equation (1)), i.e. slope and intercept are almost unchanged. This result, considering also that the updated training set covers a wider range of affinities, i.e. from -10 up to -21 kcal/mol, is supportive of the robustness of the quantitative model. The model is, however, more robust in the affinity range from -12 to -17 kcal/mol, Figure 8A). The predictive ability of the extended model was successfully probed by a leave-one-out test concerning the same set of protein-protein complexes (Figure 8B). It is worth noting that the predicted  $\Delta G^\circ$  for hRI-Ang corresponds to the in vitro value, i.e. -20.8 kcal/mol (Table 1). The same happens for Bn-Bs, for which the predicted  $\Delta G^\circ$  (i.e. -17.3 kcal/mol) is the same as the in vitro measured value reported by Hartley for the pseudo-wild type C40A<sup>Bn</sup>-C82A<sup>Bs</sup> (i.e. the form, for which the X-ray structure is available and was employed in this study; see the Methods section) [9]. The value  $\Delta G^\circ = -19$  kcal/mol reported in Table 2, indeed, refers to in vitro experiments performed on the wild type Bn-Bs complex.

The model correctly estimates also the binding affinity of the two complementary EF-hand sub-domains that constitute calbindin D9k (i.e. the experimental and predicted  $\Delta G^\circ$  values are -13.8 and -14.2 kcal/mol, respectively), by using ZD-s values previously derived by docking simulations of fragment complementation [4].

#### **Discussion**

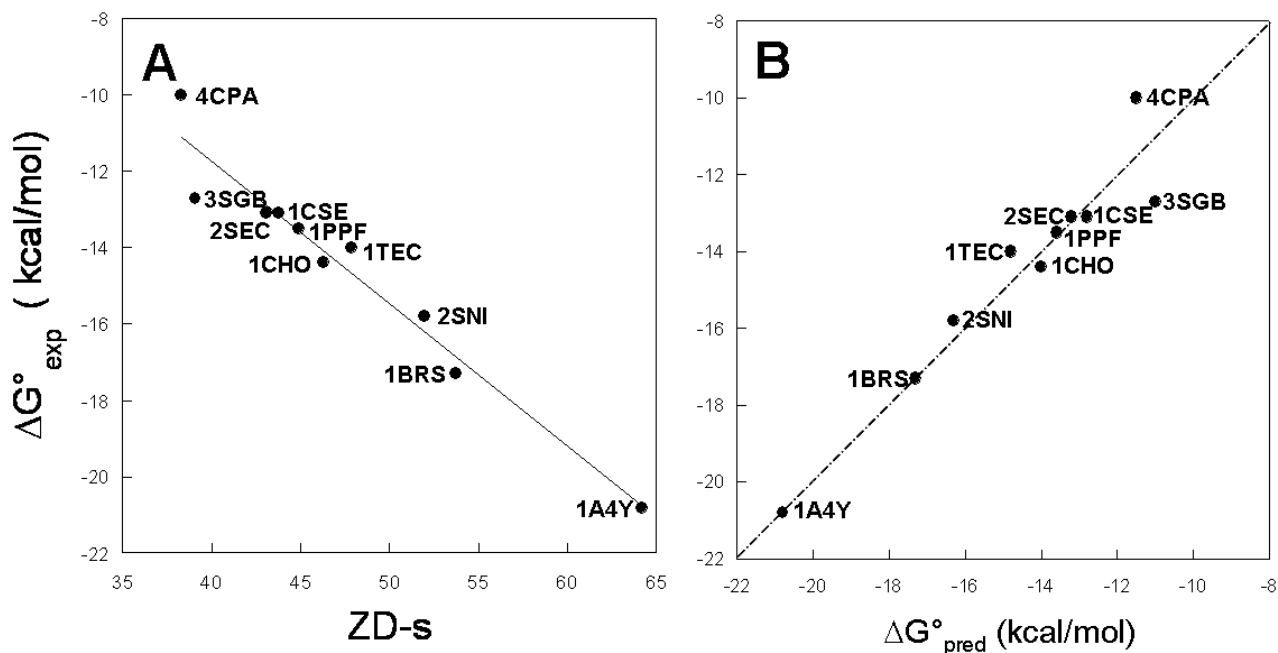
Recently, we developed a rigid-body docking based approach to estimate mutational effects on the thermodynamics and the kinetics of protein fragment complementation [4] and protein-DNA binding [5]. The approach does not require decomposition of  $\Delta G^\circ$  in enthalpic and entropic components. The major requirement is, indeed, an accurate structural model of the complex between the wild type forms of the interacting proteins. The variants

**Table 6: Kinetic parameters for hRI-Ang and Bn-Bs interaction. Statistical analysis of the empirical correlations for the association and dissociation rate constants**

Association ( $\log k_a$ ) <sup>a</sup>																									
ZD <sub>sc</sub>					ZD <sub>el</sub>				ZD <sub>des</sub>				ZD <sub>sc+el</sub>				ZD <sub>sc+des</sub>				ZD <sub>el+des</sub>				
	N	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)	A	B	R	p(R)
<b>HRI-ANG</b>	24	6.3	4.8	0.59	0.002	-16.6	4.1	0.67	0.0003	-6.4	0.7	0.19	0.363	-10.4	8.9	0.74	< 0.0001	-0.1	5.6	0.51	0.0011	-23.0	4.9	0.76	< 0.0001
<b>BN-BS_1</b>	11	41.8	-0.2	0.05	0.878	-16.3	3.1	0.50	0.114	12.1	-1.2	0.43	0.188	25.6	2.8	0.32	0.338	53.9	-1.4	0.31	0.343	-4.2	2.0	0.50	0.116
<b>BN-BS_2</b>	11	43.0	-0.4	0.08	0.807	-18.7	3.3	0.61	0.045	18.5	-1.9	0.62	0.04	24.3	2.9	0.40	0.225	61.5	-2.2	0.46	0.156	-0.2	1.5	0.44	0.169
Dissociation ( $\log k_d$ ) <sup>b</sup>																									
<b>HRI-ANG</b>	24	40.7	-0.8	0.83	< 0.0001	13.8	-0.5	0.72	< 0.0001	-1.0	-0.2	0.35	0.095	54.4	-1.3	0.93	< 0.0001	39.7	-0.9	0.74	< 0.0001	12.8	-0.7	0.90	< 0.0001
<b>BN-BS_1</b>	11	38.3	-0.7	0.93	< 0.0001	8.5	-0.5	0.50	0.116	3.2	0.2	0.57	0.068	46.8	-1.1	0.84	0.0012	41.4	-0.4	0.67	0.025	11.7	-0.2	0.41	0.217
<b>BN-BS_2</b>	11	38.4	-0.6	0.91	< 0.0001	8.5	-0.2	0.28	0.410	3.4	0.2	0.45	0.164	46.8	-0.9	0.77	0.0054	41.8	-0.4	0.58	0.063	11.9	-0.02	0.05	0.895

<sup>a</sup>Association process, described by the  $k_a$  rate constant. The linear regression equation for the overall ZD-s and  $\log k_a$  is, similarly to Table 4, in the form  $Y = A + BX$ , where  $Y$  is ZD-s,  $X$  is  $\log k_a$  and  $A$  and  $B$  are the coefficients. For hRI-Ang complexes, when the overall ZD-s is correlated with  $\log k_a$  the correlation coefficient and its probability are found to be, respectively,  $R = 0.71$  and  $p(R) = 0.0001$  and the parameters are  $A = -15.2$  and  $B = 9.5$ . For Bn-Bs\_1 complexes,  $A = 37.9$ ,  $B = 1.7$ ,  $R = 0.25$  and  $p(R) = 0.465$ , whereas for Bn-Bs\_2,  $A = 43.5$ ,  $B = 1.0$ ,  $R = 0.18$  and  $p(R) = 0.593$

<sup>b</sup>Dissociation process, described by the  $k_d$  rate constant. For hRI-Ang complexes, when considering the overall ZD-s, the correlative parameters are  $A = 53.5$ ,  $B = -1.5$ ,  $R = 0.92$  and  $p(R) < 0.0001$ . For Bn-Bs\_1 the correlative parameters are  $A = 49.9$ ,  $B = -0.9$ ,  $R = 0.88$  and  $p(R) = 0.0003$ , whereas for Bn-Bs\_2,  $A = 50.3$ ,  $B = -0.7$ ,  $R = 0.81$  and  $p(R) = 0.002$ .



**Figure 8**

A general quantitative model for docking score-based free energy predictions in protein-protein interactions. (A) Linear correlation between average ZD-s and in vitro-determined standard free energy of association for a set of ten protein-protein complexes. Each dot is labelled according to the PDB code of the complex. Experimental and computational data are reported from Ref [4], except for 1BRS and 1A4Y, which both refer to this study. The linear correlation equation is  $\Delta G^\circ = 3.13 - 0.37ZD\text{-}s$  ( $R = 0.97$ ,  $p(R) < 0.0001$ ,  $N = 10$ ). (B) Predicted versus in vitro-determined free energy of association of the same ten complexes. The predicted values refer to a leave-one-out test. The fitting equation is  $\Delta G^\circ_{\text{exp}} = -0.51 + 0.96\Delta G^\circ_{\text{pred}}$  ( $R = 0.96$ ,  $p(R) < 0.0001$ ,  $N = 10$ ).

(i.e. mutations or deletions) of either one or both of the partners can be achieved by molecular modelling. The basic assumption is that the architecture of the mutated complexes is almost the same as that of the wild type and no major conformational changes occur upon binding.

In this study, the same approach has been probed in its ability to estimate mutational effects on enzyme-proteic inhibitor binding, focusing on three substantially different systems, i.e. hRI-Ang, Bn-Bs, and BPTI- $\beta$ -Tryp complexes.

In all the tested cases, the simulated mutations essentially occur at the enzyme-inhibitor interface and their contribution to  $\Delta G^\circ$  remarkably varies from case to case, depending on whether or not they concern hot-spot regions for the specific interaction. The physico-chemical and geometric features of the interface are quite different in the three considered systems. In fact, a) in the case of the hRI-Ang complex, the interface is mostly made of contacts between aromatic amino acids; b) in the Bn-Bs complex, it is essentially made of ionic interactions; and c) in

the BPTI- $\beta$ -Tryp system, a major contribution from one partner, i.e. BPTI, arises from the amino acid K15, which is involved in charge-reinforced H-bonding interactions (Figures 1, 2, 3). Thus, the extension of the interface is larger for the hRI-Ang and Bn-Bs compared to BPTI- $\beta$ -Tryp. As a consequence, in the case of BPTI- $\beta$ -Tryp, computational analyses focused on the functionally important amino acid K15<sup>BPTI</sup>, which was, indeed, subjected to eight different replacements. In contrast, for the hRI-Ang and Bn-Bs systems, mutational analyses involved different amino acid residues on the enzyme, the inhibitor, or both.

Despite the relevant differences in size/shape and electrostatic features of the three considered interfaces, the ZDOCK algorithm was always able to reconstitute the native complex, which was assigned the best docking score in almost every case. Intriguingly, the approach could correctly estimate the effects of point mutations on the binding thermodynamics, hence resulting in good linear trends between experimental  $\Delta\Delta G^\circ$  and  $\Delta ZD\text{-}s$  (Figures 4, 5, 6, 7).

The relative ZD-s employed in such correlations are simple arithmetic averages over the scores of all the native-like solutions obtained from three independent sets of docking runs. Consistent with previous observations [4,5], the employment of ZDOCK scores averaged over ensembles of native-like solutions performs better in distinguishing native-like solutions from false positives than the ZDOCK scores computed on the best native-like complex. Moreover, the averages perform better than the single score in correlation analyses. In detail, for the hRI-Ang case, the score of the best-ranked solution of each variant ( $ZD_{best}$ , Table 1) leads to similar correlations with respect to the average ( $R_{ZD_{best}} = 0.91$  versus  $R_{ZD-s} = 0.92$ ), yet with worsened statistical parameters (i.e. standard error on  $\Delta ZD$ -s equal to 0.57, versus 1.20 for  $\Delta ZD_{best}$ ). As a more general trend, and consistent with previous observations [4,5],  $ZD_{best}$  performs significantly worse in correlations with  $\Delta G^\circ$ 's for both Bn-Bs ( $R_{ZD_{best}} = 0.55$  versus  $R_{ZD-s} = 0.90$ ; Table 3) and BPTI- $\beta$  Tryp ( $R_{ZD_{best}} = 0.80$  versus  $R_{ZD-s} = 0.86$ ; Table 4) interactions. Moreover, the employment of averages overcomes, at least in part, the low resolution of the complexes involving the modelled mutants taking also into account possible slight changes in the binding modes of the mutant structures compared to the wild type. On this line, we have found that the accuracy of predictions is related to the sampling of the native-like solutions; in other words, performance improves with the increase in the number of randomized docking runs and in the consequent number of native-like solutions. Redocking is, hence, necessary for computing such an average score and is worth doing even in the rare cases, in which the crystal structures are available for both the wild type and the mutated complexes and a simple scoring of such complexes would be possible.

The linear trends remain good when merging the  $\Delta \Delta G^\circ$ 's and  $\Delta ZD$ -s's obtained for the hRI-Ang and Bn-Bs systems. The good predictive ability of the quantitative model extended to the two different systems (Figure 6) highlights the potential of our computational approach to handle cases in which it is not possible to estimate the absolute free energies. These correlative models are also suitable for in silico screening of the amino acids relevant to the stability of the enzyme-inhibitor complexes, i.e. hot-spot residues.

Interestingly, our computational approach is effective in handling the peculiar case of the  $\beta$ -Tryp-BPTI system, in which K15, the major contributor from BPTI to the enzyme-inhibitor interface, was replaced by eight physico-chemically different amino acids. In fact, good linear trends were obtained in this case by correlating  $\Delta G^\circ$  and  $\Delta ZD$ -s, independent of the conformation of the replacing amino acid and of the presence of explicit interfacial water molecules in the simulation. In a very recent study [15],

we compared the results obtained with our method for  $\beta$ -Tryp-BPTI with those obtained by applying a well-known empirical approach to  $K_A$  predictions based on changes in the polar/apolar solvent accessible surface area ( $\Delta ASA$ ) [32]. As already demonstrated in a previous study on calbindin D9k reconstitution from complementary fragments [4], the predictive ability of our method, concerning trend and order of magnitude of  $\Delta \Delta G^\circ$ , is significantly better compared to the  $\Delta ASA$ -based method [15].

We have also investigated the role of interfacial water molecules in mediating mutational effects on the binding thermodynamics. We found that, for systems like BPTI- $\beta$ -Tryp, the presence or absence of explicit water molecules in docking simulations does not affect substantially the linear trends. In contrast, in the case of Bn-Bs, the essential role of interfacial water molecules in mediating protein-protein contacts would require the inclusion of such molecules in the simulations, under the condition that their positions both in the wild type and the mutated complexes are defined.

Collectively, the results of this study suggest that interfacial water molecules should be included in simulations only when their positions are well defined both in the wild type and in the mutants and they result to be relevant for the modulation of mutational effects on the association process.

The correlative analyses by dissecting the ZD-s in the three components, i.e.  $ZD_{scr}$ ,  $ZD_{el}$  and  $ZD_{des}$ , also led to interesting results. Collectively, the amelioration of the correlation coefficients on using the single or paired components instead of the whole score did not occur in all the test cases and, when it occurred, it was marginal. Current and previous results suggest that the whole ZD-s is more suitable for absolute  $\Delta G^\circ$  estimations concerning heterogeneous molecular systems than the single or differently paired components. However, dissection of the score may provide hints about the physico-chemical determinants of mutational effects on the association process. In detail, for the hRI-Ang and Bn-Bs systems, shape and electrostatic complementarities, considered singularly or in combination, play a major role in modulating the effects of point mutations on the binding free energy compared to desolvation (Table 5). In contrast, in the case of the  $\beta$ -Tryp-BPTI system, desolvation taken alone or paired with each of the other two components plays a major role in modulating mutational effects.

The interpretative and predictive abilities of the correlations concerning the kinetic properties of the enzyme-inhibitor interactions are less striking than those concerning  $\Delta G^\circ$ . However, in the general context of protein-pro-

tein interaction kinetics, the results of this study allow for some interesting remarks. While general rules have been established concerning the diffusion-limited association rates of protein-protein interactions, geometric constraints characterizing the binding sites and physico-chemical processes intervening upon binding cannot be neglected in most of the cases [33]. Indeed, a number of factors influence the kinetics of protein-protein interactions, such as viscosity, ionic strength, pH, and mutations [33,34]. Here, we focus only on the latter contributor. It has been demonstrated that only mutations involving charged residues significantly affect  $k_a$ , although the magnitude of the effect is strictly related to the specific location of the mutation [33,34]. The most general model of association of two proteins to form a complex is indeed a four-state model including the formation of an unstable encounter-complex and a following intermediate committed to form the final complex [33]. Every step is in principle characterized by specific rate constants, each contributing to the overall rate of association [33]. Simplified models, such as two-state kinetic models, should then be considered approximations that require experimental validation. At best, a rule of thumb for interpreting kinetic data is the notion that short-range interactions mostly affect  $k_d$  whereas long-range interactions mostly affect  $k_a$ . For the systems analyzed in this study, i.e. hRI-Ang and Bn-Bs, the changes in association rate constant upon mutation are of one and two orders of magnitude, respectively (Tables 1, 2, 3). The fact that, for hRI-Ang,  $k_a$  correlates linearly with  $ZD_{el+des}$  is in line with the major role played by desolvation and charge-charge interactions in steering the association process. However, despite the relevant role of electrostatics in Bn-Bs association, no significant correlation was found between any combination of ZD-s components and the  $k_a$ 's for such system, indicative of the low predictive ability of the method concerning mutational effects on the association kinetics. In contrast, the overall ZD-s resulted to be a good predictor of mutational effects on the dissociation kinetics, in line with the high correlation between  $k_d$ 's and  $K_A$ 's. Collectively, we conclude that our protocol is not suitable for accurately predicting association rate constants, although some interesting insights can be achieved in some cases. A combined approach, including for instance the use of the program PARE, developed by Schreiber and co-workers [35], may eventually lead to a more complete kinetic characterization of the designed complexes.

The extension of equation (1) to equation (2) with the improvement of the statistical parameters, on one hand, and the substantially unchanged fitting curve concerning the  $\Delta G^\circ$ -ZD-s correlation, on the other, clearly shows that ZD-s, which has now been tested on a broad range of experimental affinities (Figure 8), is, indeed, a good empirical descriptor of protein-protein absolute associa-

tion free energy. In fact, the developed quantitative model covers a wide range of affinities, i.e. from -10 up to -21 kcal/mol, being, however, more robust in the affinity range from -12 to -17 kcal/mol (Figure 8).

A number of elegant computational approaches have been recently employed to explore the binding energetics and predict mutational effects on the thermodynamics of association of a number of protein-protein systems, including the Bn-Bs and BPTI- $\beta$ -Trypsin considered in this study [25,36-38]. In this respect, Guerios et al. developed the FOLDEF (i.e. FOLD-X energy function) algorithm for fast estimations of mutational effects on protein stability and protein-protein association [37]. The FOLD-X energy function includes several terms: van der Waals interactions, solvation effects, hydrogen bonds, water bridges, electrostatic and entropic effects for the backbone and side-chains. Backbone entropy is calculated from the secondary structure amino acid preferences derived from the statistical analysis of a protein structure database. Side-chain entropy is calculated by estimated values scaled by a factor that accounts for the loss of side-chain mobility upon burying [37]. Predictions of mutation-induced binding free energy changes were carried out on 82 mutants, whose structural models were obtained by mutating the X-ray structure of the wild type complex. Since the structures of the mutants are identical to that of the wild type except for the mutated amino acid side chain, a necessary condition for the good performance of the approach is that mutation does not induce any structural change in the complex, compared to the wild type. This is a less stringent condition in our approach since the complexes of the mutants are predicted by the docking algorithm and differ, though by a low extent, from that of the wild type. Moreover, as already stated above, the employment of a docking score averaged over the scores of an ensemble of native-like solutions contributes to take into account possible deviation of the mutant complex from the wild type one. The correlation coefficient for the binding free energy changes predicted by FOLDEF versus the calculated ones was 0.64, considering the whole set of 82 mutants [37]. We could compare the predictive ability of our approach with that of the FOLDEF algorithm for the hRI-Ang and Bn-Bs wild type complexes. While our docking-based approach predicts with acceptable accuracy the experimental values for both the hRI-Ang and Bn-Bs  $\Delta G^\circ$ 's (i.e. -20.8 kcal/mol and -17.3 kcal/mol, respectively (Figure 8B)), the FOLD-X algorithm overestimates the affinity in both cases (i.e. it predicts -31.1 kcal/mol  $\Delta G^\circ$ , for hRI-Ang, and -21.4 kcal/mol  $\Delta G^\circ$ , for Bn-Bs). Another approach based on energetics calculation on the X-ray structures of the complexes subjected to global energy optimization of the hydrogen bonding network is that proposed by Kortemme and Baker [36]. The model successfully predicted the results of alanine scanning



experiments, i.e. 743 mutations in globular proteins and 233 mutations at 19 protein-protein interfaces [36]. Tested protein-protein complexes included also the Bn-Bs one object of our study. The results by Kortemme and Baker concerning such system corroborate the inference of our study that the explicit inclusion of interfacial water molecules improves the estimation of mutational effects on the thermodynamics of Bn-Bs association. Indeed, the model by Kortemme and Baker neglects water molecules and did not perform so well in mutational analysis of Bn-Bs [36].

Other approaches relied on different molecular simulations protocols [25,38]. In detail, the work by Wang et al. concerned the study of a large set of mutants of the Bn-Bs system by complementary computational methods, in order to estimate the relative importance of different contributions to the binding affinity of the two proteins, to predict mutation-induced  $\Delta\Delta G^\circ$ 's, and to conceive a way for designing mutants able to improve Bn-Bs binding compared to wild type [25]. The approach relied on the knowledge of the crystal structure of the wild type complex that was employed to build the structures of the mutants. The minimized structures of the wild type and mutant Bn-Bs were subjected to electrostatic binding free energy calculations by means of the UHBD6.1 program and to Coulombic and Lennard-Jones interaction energy calculations by means of Amber7.0. Decomposition of the Coulombic and Lennard-Jones interaction energies on a per residue pair basis generated 19580 descriptors for each Bn-Bs complex. Multivariate analysis by means of COMBINE PLS led to good predictive models of the quantitative effects of mutations on Bn-Bs binding. The COMBINE analysis model, together with Poisson-Boltzmann electrostatics calculations and Brownian Dynamics simulations gave predictions of the mutants that should bind faster and with higher affinity than the wild type proteins [25]. The QSAR (Quantitative Structure Activity Relationships) model by Wang et al. rely on a multivariate analysis, whereas our correlative models rely on the average docking score taken as a whole or decomposed. Our method takes into account possible differences in the binding mode of the mutants by a larger extent compared to the approach by Wang et al. Comparisons of the predictive ability of our approach with that of Wang et al. for the set of 11 Bn-Bs complexes (i.e. the wild type and 10 mutants) that gave the best correlative models (Table 2 and Figure 5C) shows a slightly better performance for the COMBINE QSAR model. In fact, the correlation coefficient and the standard deviation concerning the experimental versus calculated  $\Delta\Delta G$  by the COMBINE model are 0.95 and 1.13, respectively, whereas those computed through the correlative model reported in Figure 5C are 0.90 and 1.34, respectively.

The work by Almlöf et al. concerned an adaptation of the Linear Interaction Energy approach (LIE) to predictions of mutational effects of K15<sup>BPTI</sup> on the binding to  $\beta$ -trypsin [38]. Where possible, the crystal structures of the wild type and of the mutants were employed. In the absence of the crystal structures, the structural models of the mutants were obtained by mutating K15 in the crystal structure of BPTI. The method, so far limited to single point mutations, consisted of Molecular Dynamics (MD) simulations of the considered mutants and led to accurate predictions [38]. Also in this case, the basic assumption is that mutants bind in the same way as the wild type.

The differences between our approach and that by Almlöf et al. are that the former does not require any energy optimization or MD simulations and can handle multiple point mutations with acceptable accuracy. Comparisons of the predicted versus calculated binding  $\Delta G$  by the LIE method and by the correlative model reported in Figure 7A show the slightly better performance of our model. In fact, the correlation coefficient and the standard deviation are, respectively, 0.75 and 1.62, for the LIE model, and 0.80 and 1.71, for our model.

## Conclusion

Considering the outstanding importance of enzyme activity within cells, computational methods that allow simple and fast predictions of mutational effects on the binding free energy are extremely useful, especially because they may lead to the effective design of specific inhibitors. Whereas a number of first principle-based methods are available today for such purpose, the actual performance is often not balanced by the cost in computer time. So far, our method has been successfully tested on: a) 67 variants of protein-protein associations in water environment (see this work and Ref[4]); b) 32 variants of transmembrane helices dimerization [39]; and c) 10 variants of protein-DNA binding [5], in each case showing consistency with *in vitro* data.

Our approach is not aimed at investigating the subtle determinants of association for a given protein system, but it is rather aimed at finding the amino acid mutations that cause detrimental effects on the binding energy. The approach is independent of the physico-chemical nature of the protein-protein interface and, since it is based on docking simulations, it can be also applied to cases in which the structure of the complex is unknown even for the wild type.

The method is simple and fast and can be used for structure-based design of protein-protein complexes and for *in silico* screening of mutational effects on protein-protein recognition.

## Methods

### Rigid-body docking simulations of enzyme-inhibitor interaction and correlation with relative stabilities

Molecular simulations of enzyme-protein inhibitor interactions were carried out by means of the rigid-body algorithm ZDOCK2.3 [6]. The Fast Fourier Transform-based algorithm performs a search for optimization of molecular complementarity and provides a score (ZD-s) for each docking solution, which can be summarized as a combination of three components:

$$\text{ZD-s} = \alpha S_{sc} + \beta S_{el} + \gamma S_{des} = \text{ZD}_{sc} + \text{ZD}_{el} + \text{ZD}_{des}$$

where *sc*, *el* and *des* indicate, respectively, the shape complementarity, electrostatic and desolvation terms, and  $\alpha = 0.01$ ,  $\beta = 0.06$ , and  $\gamma = 1$  are scaling factors of the energy terms (S) proposed by the ZDOCK developers [40].

When necessary, structural water molecules at the interface were included in the atomic coordinates of the molecule used as a target. The atomic radius and the atomic contact energy (ACE)[41] for the water oxygen atom were thus included in the parameters file, whereas the partial charge was set equal to zero. A  $128 \times 128 \times 128$  point grid with a 1.2 Å spacing was used for digitalizing the interacting molecules. As a general rule, the protein with higher molecular weight was kept fixed (i.e. target), whereas the smaller was allowed to rotate and translate around the target (i.e. probe). A rotational sampling interval of 6° was employed, i.e. dense sampling, and the best 4000 solutions were retained for each run and ranked according to the ZDOCK score. Three independent sets of docking runs were performed for each complex, i.e. one starting from the X-ray coordinates and the other two randomizing the initial positions of the probe, in a fashion described previously [4]. We selected as native-like structures all the docked complexes characterized by a  $C_{\alpha}$ -RMSD lower than 1.0 Å from the native complex. This criterion was employed previously [4,5] and demonstrated to be a reasonable threshold. For each docking simulation, both the total score and the three components were averaged over all the native-like complexes resulting from the three independent runs, and employed in the correlation analysis with the thermodynamic and kinetic data. For each tested system, we employed the relative ZD-s index in correlative studies with the relative binding affinity ( $\Delta\Delta G^{\circ}$ ). The physical quantity  $\Delta\Delta G^{\circ}$  is defined, for each variant tested in this study, by the relationship:

$$\Delta\Delta G^{\circ} = \Delta G^{\circ}_{\text{var}} - \Delta G^{\circ}_{\text{wt}} = -RT \ln(K_{D,\text{wt}}/K_{D,\text{var}})$$

where "var" refers to either mutated or deleted forms and  $K_D$  is the equilibrium dissociation constant. All the quantities used in this relationship were taken from in vitro studies (Tables 1, 2, 3 and reference therein).

The CPU time for each docking simulation was found to vary significantly with the size of the docked system. Approximately, each docking run performed on a 2.2 GHz Opteron processor with 2GB RAM took about 23 hours for hRI-Ang (i.e. 3410 and 993 atoms, respectively) and about 8.5 hours for Bn-Bs (i.e. 864 and 693 atoms, respectively) and BPTI-β Tryp (i.e. 427 and 1557 atoms, respectively).

### hRI-Ang interaction: structural information

The structures of both the free and hRI-bound forms of Ang have been solved by X-ray crystallography. The  $C_{\alpha}$ -RMSD between isolated (PDB entry 1B1I, resolution 1.8 Å) [42] and hRI-bound Ang (PDB entry 1A4Y, resolution 2.0 Å) is 1.07 Å. Reasonably, the highest deviations occur in the 84–90 region, which is involved in the binding interface, and in the 65–69 region, which does not belong to the interface. The overall RMSD is 1.79 Å, indicative of some rearrangements involving the side chains.

The segments named A and B in the 2.0 Å resolution structure corresponding to the PDB entry 1A4Y[11] were used to model the bound structures of hRI and Ang, respectively. Table 1 shows the single or multiple point mutations performed in either hRI or Ang in each docking run. Mutations were performed by means of the Protein Design module within the QUANTA2005 package [43]. For each docking simulation, hRI was used as a target while Ang was used as a probe.

### Bn-Bs interaction: structural information

The structures of the Bn-Bs complex and of the isolated components have been determined at the atomic detail. The  $C_{\alpha}$ -RMSD between unbound (PDB entry 1A2P, resolution: 1.5 Å) [44] and Bs-bound Bn (PDB entry 1BRS, resolution: 2.0 Å)[27] is 0.46 Å, whereas the overall RMSD equals 0.77 Å. In contrast, the  $C_{\alpha}$ - and the overall RMSDs between unbound (PDB entry 1BTA, average NMR structure)[45] and Bn-bound Bs (PDB entry 1BRS, resolution: 2.0 Å) are 0.87 Å and 1.41 Å, respectively.

The Bn and Bs structures extracted from the 2.0 Å resolution complex (PDB entry 1BRS [27]) were employed in docking simulations. Three complexes are present in the crystallographic unit, which slightly differ in completeness and few side chain assignments. Quality checks through the Protein Health module of QUANTA2005 led to the final choice of the complex corresponding to the segment A for Bn (residues 3–110) and D for Bs (residues 1–89). In case of multiple side chain assignment, i.e. that of S14<sup>Bs</sup>, the criterion was to choose the conformation improving the stereochemical quality of the structural model. Tables 2 and 3 report the mutations performed in either Bn or Bs in each docking run. Mutations were performed by means of the Protein Design module within

the QUANTA2005 package, and the Ponder's rotamer library [46] was applied to assign the conformation to the Y29F substitution. The amino acid residue H102 from Bn was considered both in its neutral (i.e. with the hydrogen atom on N $\epsilon$ ) and charged forms. Docking runs on the wild type and mutated forms of Bn and Bs were performed both in the absence and in the presence of structural water molecules. Simulations without explicit water were carried out on 16 molecular systems (i.e. the wild type and 15 mutants (Table 2). In contrast, simulations with explicit water concerned a more limited set of variants (i.e. the wild type and 10 mutants, Table 3), for which information on the positions of interfacial water molecules could be obtained from the X-ray structures. In detail, the included water molecules are numbered as 14, 22, 29, 33, 36, 48, 93, 128 and 155 in 1BRS, which mediate Bn-Bs contacts and show an almost fixed position in all the X-ray structures of the complexes resolved so far, i.e. that of the wild type (1BRS [27]) and those of the mutants K27A<sup>Bn</sup>/D35A<sup>Bs</sup> (1B2U [47]), H102A<sup>Bn</sup>/Y29F<sup>Bs</sup> (1B3S [47]), K27A<sup>Bn</sup>/T42A<sup>Bs</sup> (1B2S [47]), E76A<sup>Bs</sup> (1X1X [48]) and Q2A/D35A<sup>Bs</sup> (1X1Y [48]). Only for the mutant H102A<sup>Bn</sup>/Y29F<sup>Bs</sup>, interfacial water molecules 22, 33, 93 and 128 (the numbering is that from the 1BRS structure) have been omitted because they are not present in the crystal structure of the complex (1B3S [47]). The inclusion of these water molecules was done because of their known importance in improving the shape complementarity between Bn and Bs. In addition to the interfacial water molecules common to the wild type and the mutants, ad hoc water molecules that occupy the cavity formed upon alanine replacements were included in the modelled structures of selected mutants. In detail, (a) the water molecule 48 (i.e. wat48), extracted from the structure of the K27A<sup>Bn</sup>/D35A<sup>Bs</sup> mutant (1B2U), was included in the K27A<sup>Bn</sup> mutant, whereas wat2 and wat35, extracted from the same crystal structure, were included in the D35A<sup>Bs</sup> mutant; (b) wat58 and wat98, extracted from the structure of the H102A<sup>Bn</sup>/Y29F<sup>Bs</sup> mutant (1B3S), were included in the H102A<sup>Bn</sup> mutant, whereas wat65 and wat85, extracted from the same crystal structure, were included in the Y29F<sup>Bs</sup> and Y29A<sup>Bs</sup> mutants; finally (c) wat152 and wat211, extracted from the crystal structure of the E76A<sup>Bs</sup> (1X1X [48]) mutant, were included in the E76A<sup>Bs</sup> mutant. The mutants D39A<sup>Bs</sup>, R87A<sup>Bn</sup>/D39A<sup>Bs</sup>, R59A<sup>Bs</sup>/D39A<sup>Bn</sup> and R87A<sup>Bn</sup> were excluded from calculations with explicit water because information on ad hoc water molecules that would fill the cavity left by mutation is lacking. We decided to consider D39A<sup>Bs</sup> only in docking simulations with the K27A and H102A Bn mutants, for which information on the positions of ad hoc water molecules is available (see above). For each mutant, water molecules were considered as belonging to Bn, which was kept fixed (i.e. target), whereas Bs was chosen as a probe.

It is worth noting that, in this work, the "wild type Bn-Bs" indeed refers to the complex holding the C40A/C82A double Bs mutant, elsewhere named as "pseudo-wild type" [47]. This form of Bs shows a slightly lower affinity for Bn compared to the wild type (i.e.,  $\Delta G^\circ = -17.3$  kcal/mol)[9] and the high resolution structure in complex with Bn is indeed available only for such mutant [27]. NMR relaxation spectroscopy, however, clearly showed that the structure of barstar C40A/C82A is essentially the same as that of the wild type [49], hence validating the use of the pseudo-wild type as a template in our study.

#### **BPTI- $\beta$ -Tryp interaction: structural information**

Smith et al. evaluated the total interface RMSD between the bound and unbound forms of BPTI interacting with  $\beta$ -Tryp to be, respectively, 0.57 Å for  $\beta$ -Tryp and 0.98 Å for BPTI [50].

From X-ray crystallography, two 3D-structures of the wild type BPTI- $\beta$ -trypsin complex are available at 1.85 Å (PDB entry 2PTC)[51] and 1.90 Å resolution (PDB entry 3BTK) [28]. Moreover, the structures of the complexes between ten mutants of the BPTI residue K15 and  $\beta$ -Tryp are available at an average resolution of 1.90 Å (PDB entry 3BTX, where X indicates the code of the amino acid substituting Lys) [28]. Two of these structures, i.e. 3BTK and 3BTW, were excluded from the correlative analysis because the relative in vitro data were not homogeneous with respect to the data concerning the other K15 mutants. In detail, the thermodynamic data concerning the wild type form, i.e. 3BTK, refer to an early work [31] and were assessed in different conditions than those of all the mutants [10]. On the other hand, the K15W mutant, 3BTW, presents high B factors of several side chains in the binding pocket as well as two less clearly defined water molecules [28], which make it a more ambiguous case if compared to other mutants. Docking simulations of the eight variants of BPTI, i.e. the probe, were run both by including and excluding structural water molecules at the interface as bound to  $\beta$ -Tryp, i.e. the target. Where ambiguous side chains assignments were present, alternative conformations were tested (data not shown). Moreover, the same eight BPTI mutants of known 3D structure were built by introducing the point mutation in the structure of the wild type protein, and then used as probes in docking simulations. In silico mutations were carried out on the 2PTC structure, characterized by better stereochemical quality compared to the 3BTK structure. The Ponder's rotamer library [46] was employed to assign the conformation to the mutated amino acids. No water molecules were added at the interface in order to avoid any possible bias arising from experimental information.

## Authors' contributions

DDO conceived the present study, carried out docking simulations and analyses. PGDB contributed to supervise the work. FF initiated the project and supervised the work. DDO and FF shared the writing of the manuscript. All authors read and approved the final manuscript.

## Acknowledgements

This work was supported by a Telethon-Italy grant No. S00068TELA (to FF). We are extremely grateful to Dr. Zhiping Weng for kindly providing us with the executable to dissect ZD-s in its three components. DDO gratefully acknowledges Emiliano Specchia for helpful technical suggestions.

## References

- Brooijmans N, Sharp KA, Kuntz ID: **Stability of macromolecular complexes.** *Proteins* 2002, **48(4)**:645-653.
- Kuhnel K, Veltel S, Schlichting I, Wittinghofer A: **Crystal structure of the human retinitis pigmentosa 2 protein and its interaction with Arl3.** *Structure* 2006, **14(2)**:367-378.
- Molinari F, Meskanaitė V, Munnich A, Sonderegger P, Colleaux L: **Extracellular proteases and their inhibitors in genetic diseases of the central nervous system.** *Hum Mol Genet* 2003, **12 Spec No 2**:R195-200.
- Dell'Orco D, Seeber M, De Benedetti PG, Fanelli F: **Probing fragment complementation by rigid-body docking: in silico reconstitution of calbindin D9k.** *J Chem Inf Model* 2005, **45(5)**:1429-1438.
- Fanelli F, Ferrari S: **Prediction of MEF2A-DNA interface by rigid body docking: a tool for fast estimation of protein mutational effects on DNA binding.** *J Struct Biol* 2006, **153(3)**:278-283.
- Chen R, Li L, Weng Z: **ZDOCK: an initial-stage protein-docking algorithm.** *Proteins* 2003, **52(1)**:80-87.
- Chen CZ, Shapiro R: **Site-specific mutagenesis reveals differences in the structural bases for tight binding of RNase inhibitor to angiogenin and RNase A.** *Proc Natl Acad Sci U S A* 1997, **94(5)**:1761-1766.
- Frisch C, Schreiber G, Johnson CM, Fersht AR: **Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation.** *J Mol Biol* 1997, **267(3)**:696-706.
- Hartley RV: **Directed mutagenesis and barnase-barstar recognition.** *Biochemistry* 1993, **32(23)**:5978-5984.
- Krowarsch D, Dadlez M, Buczek O, Krokoszynska I, Smalas AO, Otlewski J: **Interscaffolding additivity: binding of PI variants of bovine pancreatic trypsin inhibitor to four serine proteases.** *J Mol Biol* 1999, **289(1)**:175-186.
- Papageorgiou AC, Shapiro R, Acharya KR: **Molecular recognition of human angiogenin by placental ribonuclease inhibitor--an X-ray crystallographic study at 2.0 Å resolution.** *Embo J* 1997, **16(17)**:5162-5177.
- Schreiber G, Fersht AR: **Energetics of protein-protein interactions: analysis of the barnase-barstar interface by single mutations and double mutant cycles.** *J Mol Biol* 1995, **248(2)**:478-486.
- Shapiro R, Ruiz-Gutierrez M, Chen CZ: **Analysis of the interactions of human ribonuclease inhibitor with angiogenin and ribonuclease A by mutagenesis: importance of inhibitor residues inside versus outside the C-terminal "hot spot".** *J Mol Biol* 2000, **302(2)**:497-519.
- Clackson T, Wells JA: **A hot spot of binding energy in a hormone-receptor interface.** *Science* 1995, **267(5196)**:383-386.
- Dell'Orco D, De Benedetti PG, Fanelli F: **Single amino acid contributions to binding affinity in enzyme-inhibitor interactions: a docking-based screening of BPTI-Beta Trypsin interaction.** In *From Computational Biophysics to Systems Biology Workshop, Volume NIC Series, vol.34.* NIC, Julich, Germany ; 2006:67-72.
- Chen CZ, Shapiro R: **Superadditive and subadditive effects of "hot spot" mutations within the interfaces of placental ribonuclease inhibitor with angiogenin and ribonuclease A.** *Biochemistry* 1999, **38(29)**:9273-9285.
- Schreiber G, Fersht AR: **Interaction of barnase with its polypeptide inhibitor barstar studied by protein engineering.** *Biochemistry* 1993, **32(19)**:5145-5150.
- Schreiber G, Buckle AM, Fersht AR: **Stability and function: two constraints in the evolution of barstar and other proteins.** *Structure* 1994, **2(10)**:945-951.
- Frisch C, Fersht AR, Schreiber G: **Experimental assignment of the structure of the transition state for the association of barnase and barstar.** *J Mol Biol* 2001, **308(1)**:69-77.
- Schreiber G, Fersht AR: **Rapid, electrostatically assisted association of proteins.** *Nat Struct Biol* 1996, **3(5)**:427-431.
- Gabdoulline RR, Wade RC: **Brownian dynamics simulation of protein-protein diffusional encounter.** *Methods* 1998, **14(3)**:329-341.
- Vijayakumar M, Wong KY, Schreiber G, Fersht AR, Szabo A, Zhou HX: **Electrostatic enhancement of diffusion-controlled protein-protein association: comparison of theory and experiment on barnase and barstar.** *J Mol Biol* 1998, **278(5)**:1015-1024.
- Gabdoulline RR, Wade RC: **Protein-protein association: investigation of factors influencing association rates by brownian dynamics simulations.** *J Mol Biol* 2001, **306(5)**:1139-1155.
- Dong F, Vijayakumar M, Zhou HX: **Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of barnase and barstar.** *Biophys J* 2003, **85(1)**:49-60.
- Wang T, Tomic S, Gabdoulline RR, Wade RC: **How optimal are the binding energetics of barnase and barstar?** *Biophys J* 2004, **87(3)**:1618-1630.
- Spaar A, Dammer C, Gabdoulline RR, Wade RC, Helms V: **Diffusional encounter of barnase and barstar.** *Biophys J* 2006, **90(6)**:1913-1924.
- Buckle AM, Schreiber G, Fersht AR: **Protein-protein recognition: crystal structural analysis of a barnase-barstar complex at 2.0-Å resolution.** *Biochemistry* 1994, **33(30)**:8878-8889.
- Helland R, Otlewski J, Sundheim O, Dadlez M, Smalas AO: **The crystal structures of the complexes between bovine beta-trypsin and ten PI variants of BPTI.** *J Mol Biol* 1999, **287(5)**:923-942.
- Perona JJ, Craik CS: **Structural basis of substrate specificity in the serine proteases.** *Protein Sci* 1995, **4(3)**:337-360.
- Castro MJ, Anderson S: **Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to beta-trypsin and alpha-chymotrypsin.** *Biochemistry* 1996, **35(35)**:11435-11446.
- Vincent JP, Lazdunski M: **The interaction between alpha-chymotrypsin and pancreatic trypsin inhibitor (Kunitz inhibitor). Kinetic and thermodynamic properties.** *Eur J Biochem* 1973, **38(2)**:365-372.
- Luque I, Freire E: **Structure-based prediction of binding affinities and molecular design of peptide ligands.** *Methods Enzymol* 1998, **295**:100-127.
- Schreiber G: **Kinetic studies of protein-protein interactions.** *Curr Opin Struct Biol* 2002, **12(1)**:41-47.
- Dell'Orco D, Xue WF, Thulin E, Linse S: **Electrostatic contributions to the kinetics and thermodynamics of protein assembly.** *Biophys J* 2005, **88(3)**:1991-2002.
- Selzer T, Albeck S, Schreiber G: **Rational design of faster associating and tighter binding protein complexes.** *Nat Struct Biol* 2000, **7(7)**:537-541.
- Kortemme T, Baker D: **A simple physical model for binding energy hot spots in protein-protein complexes.** *Proc Natl Acad Sci U S A* 2002, **99(22)**:14116-14121.
- Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations.** *J Mol Biol* 2002, **320(2)**:369-387.
- Almlöf M, Aqvist J, Smalas AO, Brandsdal BO: **Probing the effect of point mutations at protein-protein interfaces with free energy calculations.** *Biophys J* 2006, **90(2)**:433-442.
- Dell'Orco D, De Benedetti PG, Fanelli F: **In silico screening of mutational effects on transmembrane helix dimerization: insights from rigid-body docking and molecular dynamics simulations.** *J Phys Chem B* 2007 in press.
- Chen R, Weng Z: **Docking unbound proteins using shape complementarity, desolvation, and electrostatics.** *Proteins* 2002, **47(3)**:281-294.

41. Zhang C, Vasmatzis G, Cornette JL, DeLisi C: **Determination of atomic desolvation energies from the structures of crystallized proteins.** *J Mol Biol* 1997, **267(3)**:707-726.
42. Leonidas DD, Shapiro R, Allen SC, Subbarao GV, Veluraja K, Acharya KR: **Refined crystal structures of native human angiogenin and two active site variants: implications for the unique functional properties of an enzyme involved in neovascularisation during tumour growth.** *J Mol Biol* 1999, **285(3)**:1209-1233.
43. [www.accelrys.com](http://www.accelrys.com).
44. Mauguen Y, Hartley RW, Dodson EJ, Dodson GG, Bricogne G, Chothia C, Jack A: **Molecular structure of a new family of ribonucleases.** *Nature* 1982, **297(5862)**:162-164.
45. Lubienski MJ, Bycroft M, Freund SM, Fersht AR: **Three-dimensional solution structure and <sup>13</sup>C assignments of barstar using nuclear magnetic resonance spectroscopy.** *Biochemistry* 1994, **33(30)**:8866-8877.
46. Ponder JW, Richards FM: **Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes.** *J Mol Biol* 1987, **193(4)**:775-791.
47. Vaughan CK, Buckle AM, Fersht AR: **Structural response to mutation at a protein-protein interface.** *J Mol Biol* 1999, **286(5)**:1487-1506.
48. Ikura T, Urakubo Y, Ito N: **Water-mediated interaction at a protein-protein interface.** *Chem Phys* 2004, **307**:111-119.
49. Wong KB, Fersht AR, Freund SM: **NMR <sup>15</sup>N relaxation and structural studies reveal slow conformational exchange in barstar C40/82A.** *J Mol Biol* 1997, **268(2)**:494-511.
50. Smith GR, Sternberg MJ, Bates PA: **The relationship between the flexibility of proteins and their conformational states on forming protein-protein complexes with an application to protein-protein docking.** *J Mol Biol* 2005, **347(5)**:1077-1101.
51. Marquart M, Walter J, Deisenhofer J, Bode W, Huber R: **The Geometry of the Reactive Site and of the Peptide Groups in Trypsin, Trypsinogen and its Complexes with Inhibitors.** *Acta Crystallogr, Sect B* 1983, **39**:480.
52. DeLano WL: **The PyMOL Molecular Graphics System.** 0.95th edition. 2002 [<http://www.pymol.org>]. San Carlos, CA, USA, DeLano Scientific

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:  
[http://www.biomedcentral.com/info/publishing\\_adv.asp](http://www.biomedcentral.com/info/publishing_adv.asp)

