J Comput Aided Mol Des (2009) 23:645–654
DOI 10.1007/s10822-009-9282-3

Empirical estimation of the energetic contribution of individual interface residues in structures of protein–protein complexes

Mainak Guharoy · Pinak Chakrabarti

Received: 3 April 2009 / Accepted: 12 May 2009 / Published online: 29 May 2009

Abstract We report a simple algorithm to scan interfaces in protein–protein complexes for identifying binding 'hot spots'. The change in side-chain solvent accessible area (ΔASA) of interface residues has been related to change in binding energy due to mutating interface residues to Ala ($\Delta\Delta G_{X \rightarrow ALA}$) based on two criteria—hydrogen bonding across the interface and location in the interface core—both of which are major determinants in specific, high-affinity binding. These relationships are used to predict the energetic contribution of individual interface residues. The predictions are tested against 462 experimental X → ALA mutations from 28 interfaces with an average unsigned error of 1.04 kcal/mol. More than 80% of interface hot spots (with experimental $\Delta\Delta G \geq 2$ kcal/mol) could be identified as being energetically important. From the experimental values, Asp, Lys, Tyr and Trp are found to contribute most of the binding energy, burying >45 Å^2 on average. The method described here would be useful to understand and interfere with protein interactions by assessing the energetic importance of individual interface residues.

Keywords Protein–protein interaction · Hot spots in the interface · Alanine scanning mutagenesis · Molecular recognition · Binding energy prediction

Electronic supplementary material The online version of this article (doi:10.1007/s10822-009-9282-3) contains supplementary material, which is available to authorized users.

M. Guharoy · P. Chakrabarti (✉)
Department of Biochemistry, Bose Institute, P-1/12 CIT Scheme VIIM, Calcutta 700054, India
e-mail: pinak@boseinst.ernet.in; pinak_chak@yahoo.co.in

Introduction

The affinity and specificity of binding between molecules is central to both chemistry and biology. Protein–protein interactions play a pivotal role in the regulation of most cellular processes and the physicochemical basis of molecular recognition has been a subject of intense investigation [1–4]. In particular, biophysical data from alanine scanning mutagenesis have revealed that the presence of a small subset of interface residues contributing a disproportionately large amount of the interaction energy is a characteristic feature of most interfaces [5–8]. These often constitute the key residues that provide the specificity that is important in the context of protein–protein interaction networks [9, 10]. Experimental alanine scanning, required to identify the interaction hot spots, is often laborious; therefore, a viable alternative is the computational identification of candidate residues for further experimental analysis [11–21].

Translating the structural knowledge of the interface into energetic parameters such as binding free energies requires an in-depth understanding of the major forces that drive protein binding. The hydrophobic effect is held as the principal driving force for protein folding [22] as well as for protein–protein association [23–27]. On the other hand, electrostatic interactions and hydrogen-bond forming capability may be more useful in defining the binding sites in proteins [28, 29]. Polar residues are also often conserved at interfaces and constitute hot spots [30]. Extending our earlier work [31] we show in this study how the burial of the surface area during complex formation may be incorporated to reconcile both hydrophobic and polar interactions.

Protein–protein interfaces can be dissected into 'core' and 'rim' regions based on the degree of burial of the constituent residues [3]. Residues having one or more completely buried atoms are said to be in the interface core, whereas those in which all atoms retain partial solvent accessibility in the complexed state constitute the rim. Residues in the core are usually more conserved than those in the rim and the contribution of core residues to the free energy of binding ($\Delta\Delta G$) correlates well with and can be quantified as a function of the loss of accessible surface area, ΔASA due to side-chain burial [31]. The relationship obtained was 26 cal/mol per Å^2 of side-chain area burial—the value resembles the estimate for the hydrophobic contribution to the free energy of protein folding [32, 33]. Such a correlation is significant only for the core residues, but not the rim. Similarly, for the hydrogen-bonded interface residues (irrespective of their location in core or rim) ΔASA was also found to be correlated to $\Delta\Delta G$. This has been quantified in the present article to provide a value of 29 cal/mol per Å^2 of the side-chain burial. These two relationships provided a model for the estimation of the energetic contribution of individual interface residues. In this model if a residue is hydrogen-bonded across the interface it contributes 29 cal/mol for each Å^2 of the area buried. If there is no hydrogen bond, but the residue is located in the core its involvement is to the tune of 26 cal/mol per Å^2. If a residue does not belong to either of these two categories it has no contribution.

Materials and methods

Datasets and initial calculations

Thermodynamic $\Delta\Delta G$ data for single X → Ala mutations of protein–protein interface residues were obtained from the Alanine Scanning Energetics database, ASEdb [34], the PINT database [35] and other individual reports (Table 1). Alanine scanning data from 13 complexes in ASEdb were used to establish the quantitative relationships between $\Delta\Delta G$ and ΔASA. These relationships were then used to predict $\Delta\Delta G$ for interface residues in 28 interfaces (including 15 additional complexes). Atomic coordinates for structures solved using X-ray crystallography were taken from the Protein Data Bank (PDB) [36]. Identification of interface residues and other interfacial features (such as ΔASA involving the accessible surface area) were carried out using the program ProFace [37]. The interface area is given by the ASA buried between the two components. Hydrogen bonds involving protein atoms across the interface were selected from the output of HBPLUS [38].

Calculation of $\Delta\Delta G$ for X → Ala mutation in interface

The first check was to identify if the interface residue in question forms one (or more) hydrogen bond(s) using its side-chain with another interface atom (which may be from the main- or side-chain) belonging to the partner protein. Based on Fig. 2b the contribution of such a residue to $\Delta\Delta G$ was estimated as 29 cal/mol per Å^2 of the side-chain (C$^\beta$ onwards) area buried, ΔASA (Eq. 1). If the side-chain of the residue does not form a hydrogen bond we noted if it belongs to the interface core, in which case its $\Delta\Delta G$ was calculated (Fig. 2a) as 26 cal/mol per Å^2 of side-chain burial (Eq. 2).

$$\Delta\Delta G_{\mathrm{calc}}(\mathrm{cal/mol}) = 29 \times \Delta\mathrm{ASA}_{\mathrm{side-chain\,C^\beta\,onwards}} \quad (1)$$

$$\Delta\Delta G_{\mathrm{calc}}(\mathrm{cal/mol}) = 26 \times \Delta\mathrm{ASA}_{\mathrm{side-chain\,C^\beta\,onwards}} \quad (2)$$

Thus the method relies solely on the calculation of the change in side-chain accessibility due to complex formation and ascertaining whether the residue participates in hydrogen bonding, and in its absence, its location in the core. It is assumed that rim residues that do not form hydrogen bonds are mostly unimportant and assign them a $\Delta\Delta G$ value of 0.0.

Results

Partitioning of residues based on the combination of $\Delta\Delta G$ and ΔASA

We begin by investigating if the nature of the residue has any influence on the qualitative nature of the relationship between side-chain burial (ΔASA) and $\Delta\Delta G$, the change in the free energy of binding of the protein complex upon mutation of the side-chain to Ala. The alanine-scanning mutagenesis data were collected from ASEdb [34] and the average $\Delta\Delta G$ values for each of the 19 residues (except Ala) were found out along with their average ΔASA (Fig. 1a). Though the standard deviations are rather large and for some residues the number of examples is low (Table S1 of supplementary material), some general trend can be discerned. Asp and Tyr show the highest contribution occurring consistently in hot spots ($\Delta\Delta G \geq 2$ kcal/mol). The former is often implicated in salt-bridge formation and both residues can form hydrogen bonds across the interface. Tyr has a high degree of residue burial (~60 Å^2) and most of its bulky side-chain gets buried upon association. Asp too shows an average burial of ~45 Å^2. Lys, Trp and Arg, known to occur frequently in hot spots [6], have high ΔASA values (>50 Å^2). Among the aromatic residues, Phe contributes less than Tyr and Trp, probably because of its inability to form interface hydrogen bonds. His, with a small heteroaromatic ring capable of forming hydrogen bonds, shows an average $\Delta\Delta G$ of ~1 kcal/mol. Ile and Leu constitute an anomalous pair—though the latter buries more area the former provides a greater $\Delta\Delta G$. Another

Table 1 Warm and hot spot ($\Delta\Delta G_{exp} \geq 1$ kcal/mol), hot spot ($\Delta\Delta G_{exp} \geq 2$ kcal/mol) and neutral ($\Delta\Delta G_{exp} < 1$ kcal/mol) interface residues from 28 protein–protein complexes and their prediction

PDB code	Protein (number of mutations in interface)		Warm/hot spot residues		Hot spot residues		Neutral residues		Data ($\Delta\Delta G_{exp}$) Source[a]
	Partner 1	Partner 2	Number (total, core)[b]	Fraction correctly predicted (total, core)[b]	Number (total, core)	Fraction correctly predicted (total, core)	Total number	Fraction correctly predicted	
1a22	hGH (25)	hGHbp (30)	15, 12 [11, 10]	0.4, 0.5 [0.55, 0.6]	7, 7	0.57, 0.57	40	0.68	KB
1a4y	RNase inhibitor (13)	Angiogenin (11)	6, 6 [6, 6]	1.0, 1.0 [1.0, 1.0]	3, 3	1.0, 1.0	18	0.72	ASEdb, KB
1ahw	Tissue Factor (8)	IgG Fab (−)	5, 3 [1, 1]	0.2, 0.33 [1.0, 1.0]	1, 1	1.0, 1.0	3	0.67	ASEdb, KB
1aie	P53 tetramer (17)	–	4, 3	0.75, 1.0	1, 1	1.0, 1.0	13	0.62	GNS
1bp3	hGH (19)	Prolactin receptor (−)	7, 6	0.43, 0.33	2, 2	0.5, 0.5	12	0.75	PINT
1brs	Barnase (7)	Barstar (5)	11, 7 [9, 7]	0.55, 0.71 [0.67, 0.71]	9, 6	0.67, 0.83	1	0.00	ASEdb, KB, PINT
1bxi	Im9 (17)	ColicinE9 (−)	9, 8 [9, 8]	0.56, 0.50 [0.56, 0.5]	7, 6	0.57, 0.50	8	1.0	ASEdb, KB
1cbw	BPTI (6)	Chymotrypsin (−)	1, 1 [1, 1]	1.0, 1.0 [1.0, 1.0]	1, 1	1.0, 1.0	5	0.6	ASEdb, KB
1dan	Tissue Factor (21)	Factor VII A (−)	4, 4 [4, 4]	1.0, 1.0 [1.0, 1.0]	2, 2	1.0, 1.0	17	0.76	ASEdb, KB, PINT
1dkg	GRPE (14)	DNAK (−)	9, 5	0.44, 0.8	1, 1	1.0, 1.0	5	0.8	PINT
1dn2	IgG1 Fc (3)	Peptide (2)	5, 5 [5, 5]	0.8, 0.8 [0.8, 0.8]	2, 2	1.0, 1.0	–	–	KB
1dvf	D1.3 (16)	E5.2 (9)	22, 10	0.41, 0.80	9, 7	0.67, 0.86	3	0.67	ASEdb
1emv	Im9 (20)	Colicin E9 (−)	10, 9	0.50, 0.56	7, 7	0.57, 0.57	10	1.0	PINT
1fcc	Protein G (8)	IgG Fc (−)	5, 5 [5, 5]	0.8, 0.8 [0.8, 0.8]	4, 4	0.75, 0.75	3	1.0	KB
1gc1	CD4 (17)	gp120 (−)	3, 3 [2, 2]	0.33, 0.33 [0.5, 0.5]	–	–	14	0.79	ASEdb, KB
1iar	IL-4 (12)	IL-4 receptor (−)	3, 2	0.33, 0.50	1, 1	1.0, 1.0	9	0.78	GNS, PINT
1jck	SEC3 (9)	TCRβ (−)	8, 6 [8, 6]	0.63, 0.67 [0.63, 0.67]	4, 3	0.50, 0.33	1	1.0	ASEdb, KB
1jdh	HTCF-4 (3)	β-catenin (−)	1, 1	1.0, 1.0	–	–	2	0.0	PINT
1jrh	A6 (17)	IFNγ receptor (10)	17, 13 [17, 13]	0.65, 0.69 [0.65, 0.69]	8, 8	0.75, 0.75	10	0.8	KB
1jtg	TEM1 β-lactamase (3)	BLIP (1)	2, 2	0.5, 0.5	–	–	2	1.0	AUS
1nmb	NC10 (1)	N9 neuraminidase (−)	1, 1 [1, 1]	1.0, 1.0 [1.0, 1.0]	–	–	–	–	KB
1p2j	PTI (2)	Trypsinogen (−)	2, 2	1.0, 1.0	2, 2	1.0, 1.0	–	–	PINT
1vfb	D1.3 (13)	HEL (12)	9, 7 [8, 7]	0.78, 0.86 [0.88, 0.86]	3, 2	1.0, 1.0	16	0.75	ASEdb, KB
1ycq	P53 (7)	Mdm2 (−)	4, 4	1.0, 1.0	2, 2	1.0, 1.0	3	1.0	MK
1ycr	P53 (9)	Mdm2 (−)	4, 3	0.75, 1.0	2, 2	1.0, 1.0	5	0.8	MK
2ptc	BPTI (1)	β-trypsin (−)	1, 1 [1, 1]	1.0, 1.0 [1.0, 1.0]	1, 1	1.0, 1.0	–	–	ASEdb, KB
3hfm	HYHEL-10 (12)	HEL (13)	17, 15 [16, 15]	0.71, 0.73 [0.75, 0.73]	12, 11	0.58, 0.64	8	0.63	ASEdb, KB
3hhr	hGH (40)	hGHbp (26)	18, 12	0.39, 0.42	8, 5	0.50, 0.60	48	0.73	ASEdb, PINT
Overall performance			203, 156 [104, 92]	0.68, 0.74 [0.80, 0.80]	99, 87	0.82, 0.83	256	0.73	

The prediction is assumed to be correct if both the calculated and experimental $\Delta\Delta G$ values are ≥ 1 kcal/mol (for warm/hot residues), if the experimental $\Delta\Delta G \geq 2$ kcal/mol and calculated value ≥ 1 kcal/mol (for hot residues), and, both the calculated and experimental $\Delta\Delta G$ values are <1 kcal/mol (for neutral residues)

[a] KB [14], *ASEdb* alanine scanning energetics database, http://www.asedb.org/ [34], GNS [12], *PINT* protein–protein interactions thermodynamic database, http://www.bioinfodatabase.com/pint/index.html [35], AUS [43], MK [11]

[b] The numbers in square brackets shows the performance of our algorithm only on those mutated residues that were designated as interface residues both by our method as well as by Kortemme and Baker (from Supporting information, Table 8 of [14])

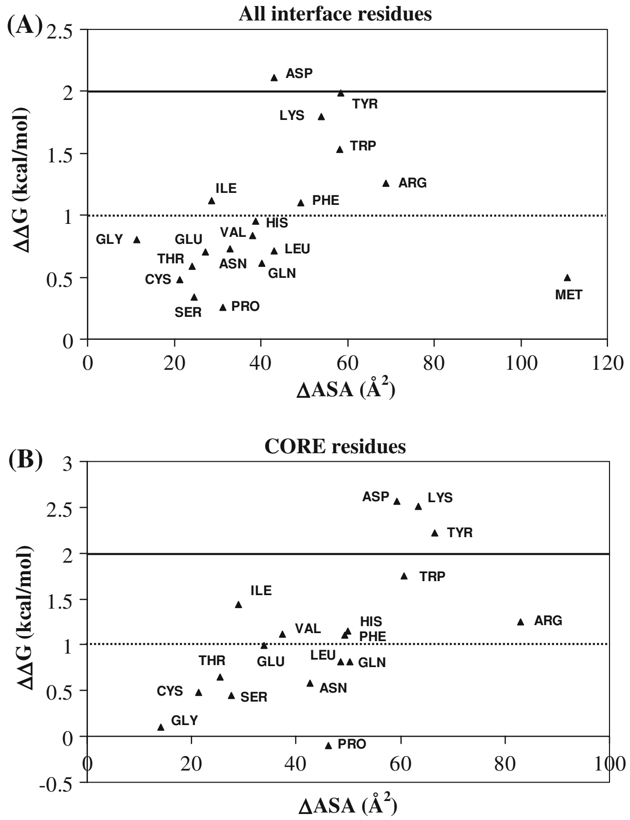

Fig. 1 Joint distribution of the change in free energy of binding ($\Delta\Delta G$) for Ala-scanning and the burial of surface area (ΔASA) for 19 amino acid residues. Residues in the whole interface and those belonging to the core are used in (**a**) and (**b**), respectively

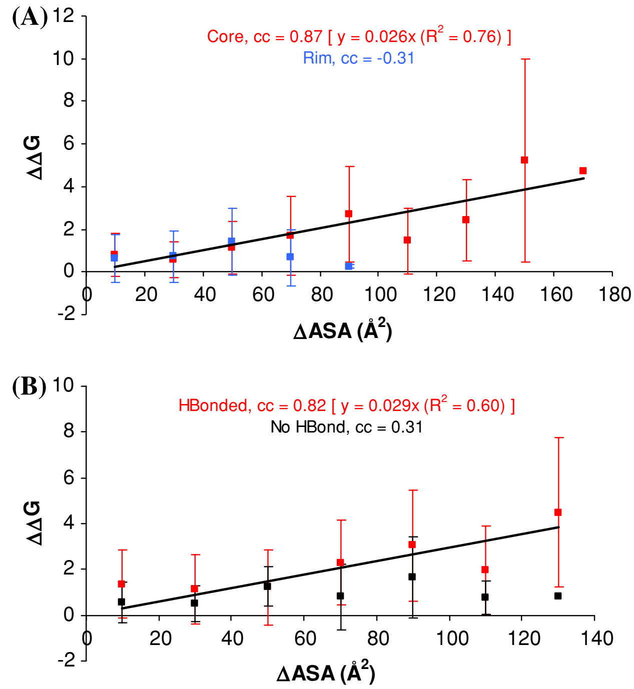

Fig. 2 Relationship between ΔASA and the change in free energy of binding ($\Delta\Delta G$) due to alanine scanning mutagenesis. Along the x-axis all of the values in a bin (size 20 Å^2) are pulled together and shown in the *middle*, while the y value corresponds to the mean of their $\Delta\Delta G$ values (the *vertical bars* representing the SDs). ΔASA corresponds to the contribution of the entire side chain (C^{β} onward) of a residue to the interface area. **a** Values for the core and rim residues are marked in *red* and *blue*, respectively. The correlation coefficient (cc) between the variables and the equation for the least-squares line (passing through the origin) for the core residues are indicated. In **b** the segregation into two categories is on the basis of the presence (*red*) or absence (*black*) of hydrogen bonding across the interface. The SDs of the slopes in **a** and **b** are 0.0026 and 0.0036, respectively. In **b** if data points for the hydrogen-bonded residues are weighted using their variance, $cc = 0.76$ and the fitted line is $y = 0.028x$, with 0.004 as the SD of the slope

interesting difference between chemically similar groups is shown by the pair Asp and Glu, with only the former occurring in hot spots, though both the residues are under-represented in interfaces [1–4]. In this connection it may be mentioned that a comparison of propensities of residues to occur in secondary structural elements in the interface relative to those in the tertiary structure has shown that Asp, but not Glu, has a high propensity to occur in interface β-strands [39], indicating its importance in the binding specificity. It is also interesting to note that while Asn and Asp bury similar amount of surface area, are of similar size [40] and both can participate in hydrogen bonds, Asn contributes less than 1 kcal/mol whereas Asp is a hot spot residue. Ser, Thr, Asn and Gln can all form hydrogen bonds, but because of their small size contribute less than 1 kcal/mol towards binding. The surprisingly large ΔASA value for Met is based on only a single interface Ala-scan. Overall, we can conclude that a large extent of burial alone or in conjunction with hydrogen bonding determine the importance of a residue to binding. Restricting the analysis to residues belonging to the core, we observe a qualitatively similar distribution (Fig. 1b). Assuming that a $\Delta\Delta G$ value of $\geq$1 kcal/mol qualifies a residue as 'warm' and a value $\geq$2 kcal/mol as 'hot', several residues have moved upwards with a concomitant increase in ΔASA. The top 5 residues bury more than 45 Å^2 in Fig. 1a, which increases to >60 Å^2 in Fig. 1b. Most of the warm and hot residues are located in the interface core and/or participate in interface hydrogen bonds [31].

Quantifying the relationship between $\Delta\Delta G$ and ΔASA for hydrogen bonded residues

Although previous studies had found little correlation between the buried surface area of the side chain and the free energy of binding [6], we noticed that the situation improves on separating the data based on the location of the residues in the core or the rim, especially because the residues that are important for binding energetics tend to occur more near the center of the interface than at its edges

[31]. The correlation between ΔASA (calculated by considering the contribution of the side chain atoms C^{β} onwards of the residue to the interface area) and ΔΔG is significant for the interface core, but not the rim, and it yielded a slope of 26 cal/mol Å^2 (the plot from [31] is reproduced in Fig. 2a). We had also found a good correlation between ΔASA and $\Delta\Delta G$ for hydrogen-bonded residues as opposed to non-hydrogen-bonded ones. However, ΔASA was calculated by using a slightly different definition of the side-chain atoms. For the sake of consistency here we use the same definition of the side chain (C^{β} onwards) to calculate ΔASA and the plot against $\Delta\Delta G$ for hydrogen-bonded residues (Fig. 2b) yields a value of the energetic contribution to be 29 cal/mol per Å^2.

Predicting the $\Delta\Delta G$ of X → Ala mutations in the interface

462 cases (from 28 complexes) with experimental $\Delta\Delta G$ data have been used, out of which 143 form hydrogen bonds through the side chain (Table S1 of supplementary material gives details about all these mutations). Three of these residues are glycines, which have been excluded from the analysis, as with no side-chain these cannot be included in our calculations. The average unsigned error (calculated as $|\Delta\Delta G_{calc} - \Delta\Delta G_{exp}|$) between predicted and experimental $\Delta\Delta G$ values (calculated over all 459 mutations) is 1.04 kcal/mol. Similar values have been reported by other studies, although carried out on a smaller number of mutations. For example, Guerois et al. [12] obtained a value of 0.88 kcal/mol (for 82 mutations), Pokala and Handel [16], 1.04 kcal/mol (436), and Kortemme and Baker [14], 1.06 kcal/mol (233). It has previously been pointed out [16] that independent experiments to determine binding energy changes often result in average discrepancy between 0.25 and 0.5 kcal/mol.

We designate interface residues with experimental $\Delta\Delta G$ values between 1 and 2 kcal/mol, ≥2 kcal/mol, and <1 kcal/mol as warm, hot and neutral, respectively. We judged the ability of the method in reproducing the experimentally derived energetic importance of interface residues from each complex by using the following criteria. (1) The model was considered to be successful in qualitatively predicting a warm/hot spot residue if both the experimental as well as calculated $\Delta\Delta G$ values were ≥1 kcal/mol. For each interface, we calculated the fraction of hot/warm residues that satisfied this criterion. (2) Separately, we also calculated the fractions of hot residues from each interface for which the calculated $\Delta\Delta G$ value was ≥1 kcal/mol. (3) Lastly, neutral residues were considered to be successfully predicted if both experiment and calculation assigned values less than 1 kcal/mol. Following this scheme, the fractions of correctly predicted hot/warm, hot and neutral residues for all 28 interfaces are tabulated (Table 1). 68% of hot/warm residues (74% if only interface core residues are considered), 82% of hot residues (83%, considering those belonging to core) and 73% of neutral residues are identified correctly. Kortemme and Baker [14] had used a similar scheme for demonstrating the performance of their computational method in predicting the energetic effect of mutation, and the two can be compared based on the success percentage. Due to differences in the two approaches, certain residues that were labeled as belonging to the interface by our method were non-interface residues according to theirs and vice versa. When we used only those hot residues (104 in number) that were identified as belonging to the interface by both methods, 80% of hot/warm residues could be identified correctly (these values are given in square brackets in Table 1). This success rate compares favorably with the 79% success rate obtained for 120 interface hot spots by Kortemme and Baker [14]. The general trend of localization of hot residues within the interface core and the match between experimental and calculated $\Delta\Delta G$ values obtained using our method are depicted for four selected interfaces (1a4y, 1dan, 1fcc and 1vfb) in Fig. 3. Detailed description of predictions for these four interfaces is provided in the supporting text of supplementary material.

Interfaces containing bridging water molecules

Water molecules are often trapped in the interface forming bridged hydrogen bonds with protein atoms on either side [41]. Although water-mediated hydrogen bonds are ignored by our method, it is likely that the participating residues also form direct hydrogen bonds or belong to the interface core, and as such are included in our calculations. The complex between barnase and barstar (1brs) is an interesting case in which seven residues are known to form water-mediated hydrogen bonds in the interface and six of them have $\Delta\Delta G_{exp} > 1$ kcal/mol. Of these, four residues (K27, R59, D35 and D39) are correctly assigned following the scheme mentioned in the last section (Fig. 4a) (detailed values are given in Table S2). The remaining two (N58 and E73) are poorly predicted—these were also underpredicted in an earlier study [14]. The complex between lysozyme and its cognate antibody (D1.3) (1vfb) is another example of an interface containing water-mediated hydrogen bonds. The predictions are accurate (Fig. 4b) as seven out of the nine experimentally determined binding hot spots are correctly identified. These results suggest that for biological interfaces, with water molecules filling in the cavities, the buried surface area and any direct hydrogen bonding involving the side chain is usually adequate in determining the degree of its importance towards the binding.

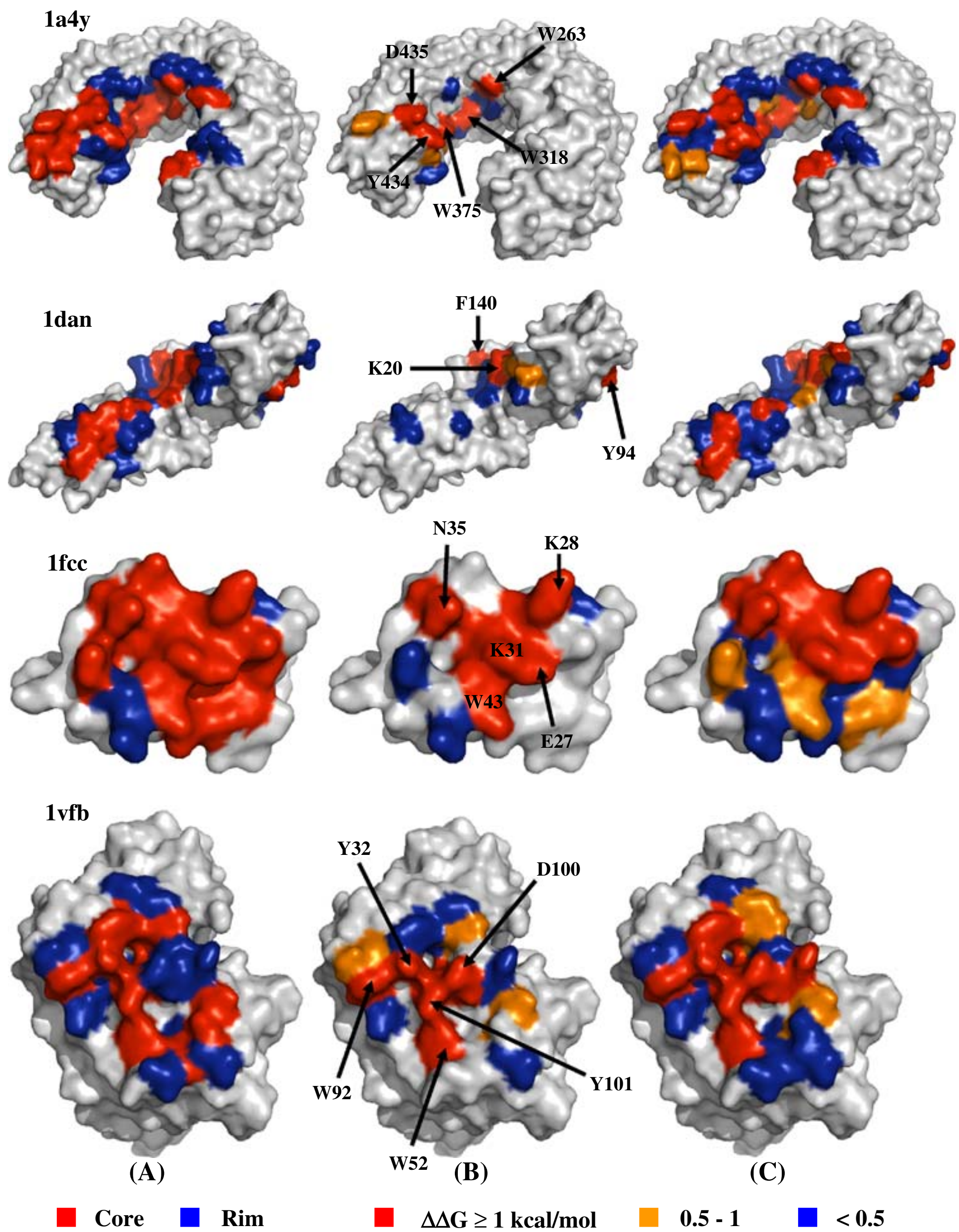

Fig. 3 Match between **a** core/rim dissection, **b** experimental $\Delta\Delta G$ and **c** calculated $\Delta\Delta G$ for four selected protein complexes (with PDB IDs) 1a4y, 1dan, 1fcc, and 1vfb

Predictions for interfaces that undergo structural rearrangements

Conformational rearrangements are not modeled in our method either. Kortemme and Baker [14] optimized rotameric side-chain conformations, but found that this additional step did not significantly improve predictions in the majority of cases. The complex between human growth hormone and its receptor (1a22) is an example of an interface showing large amount of structural plasticity. Our method correctly identified $\sim$50% of hot spots (Fig. 4c), including two Trp residues (W104 and W169 on the receptor) that have the maximum contribution to binding energy. There are several residues in this interface that have relatively large effects on the binding by helping to position the two Trp residues [42]. Such indirect effects not leading to significant intermolecular contacts are not recognized by our scoring functions and that explains the relatively low prediction accuracy for this complex. Another example is the interface between staphylococcal enterotoxin C3 and the T cell receptor β chain (1jck) which undergoes conformational rearrangements upon binding. Still five out of eight interface hot spots are correctly identified (Fig. 4d).

Cooperative binding effects

We have also investigated a situation involving cooperative binding effects caused by the multiply connected nature of

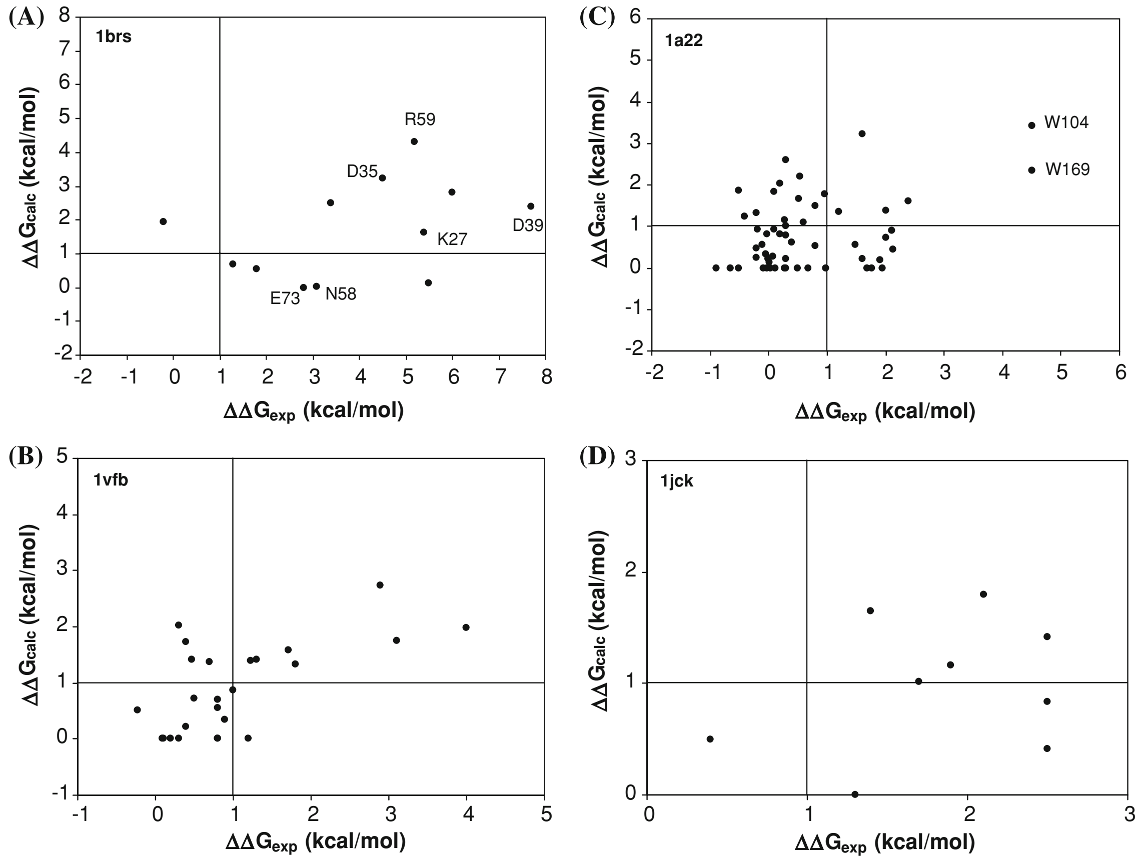

Fig. 4 $\Delta\Delta G_{\mathrm{calc}}$ versus $\Delta\Delta G_{\mathrm{exp}}$ for interfaces known to **a**, **b** form water-bridged hydrogen bonds and **c**, **d** involve conformational changes upon association. For ease of comparison two lines have been drawn cutting the axes at 1 kcal/mol. The *upper right quadrant* contains the correctly predicted hot/warm residues

the binding region. Albeck et al. [43] applied a modified multiple-mutant cycle protocol to evaluate binding free energy contributions of five residues forming a distinct and specific binding unit in the complex between TEM-1-β-lactamase and its protein inhibitor, BLIP (1jtg). Within this binding unit, Asp49 of BLIP serves as 'hub' residue forming two salt bridges (with Arg243 and Lys234) and two hydrogen bonds (with Ser235 and Ser130) with the enzyme molecule (Figure S1). Each of these four partner residues is highly conserved and lines the active site of the TEM molecule. Asp49 is the most connected residue in this interface, which is reflected in the predicted $\Delta\Delta G$ value of 3.27 kcal/mol ($\Delta\Delta G_{\mathrm{exp}} = 1.97$ kcal/mol—this residue is clearly picked up as a key hot spot residue, although our algorithm over-estimates the absolute $\Delta\Delta G$ value).

Comparison with molecular mechanics approaches

We performed computational alanine scanning on the residues forming the interface between mouse double mutant (mdm2) and a peptide derived from the tumor suppressor (p53) and compared our results with those obtained earlier using molecular mechanics Poisson Boltzmann surface area (MM-PBSA) approach [11]. We observe a strong correlation (0.92) between the $\Delta\Delta G$ results obtained using the two methods (Figure S2) showing that in principle, the simple model developed here has incorporated the principal factors that decide binding energetics and has the potential to perform well relative to computationally intensive methods.

Discussion

The $\Delta\Delta G$ is an indicator of the energetic importance of a given interface residue to the binding process. The model presented in this paper visualizes the protein–protein interface as dissected into an energetically important 'core' (containing most of the hot spots) that is sheltered from the bulk solvent by being surrounded by the less important 'rim', thus allowing a better microenvironment for interaction. Even for residues that form hydrogen bonds across the interface, solvent exclusion is an important factor which determines their energetic contribution. Indeed, exposed hydrogen bonds at the interface contribute little energy [14], but may contribute to the specificity. This is similar in philosophy to the model proposed by Bogan and Thorn [6], where inaccessibility to the solvent has been shown to be a necessary condition to define a residue as a binding hot spot, effectively forming an ‘O-ring’ structure surrounded by energetically less important residues that mainly serve to occlude bulk solvent. In general, hot spots contribute larger ΔASA relative to non-hotspot residues. The percentages of both hot and non-hotspot residues falling in different ΔASA bins are plotted in Fig. 5a. As ΔASA increases there is a sharp decrease in the percentage of non-hotspot residues, whereas a significant fraction of hot spot residues bury large amounts of surface area. In addition, there is also a trend of interface residues forming hydrogen bonds to be buried to a greater extent compared to the ones that do not participate in hydrogen bonding (Fig. 5b). The extent of burial of the residue side-chain correlates with the $\Delta\Delta G$ of binding for residues in the core and also for residues hydrogen bonded across the interface (Fig. 2). These relationships have been quantified and used

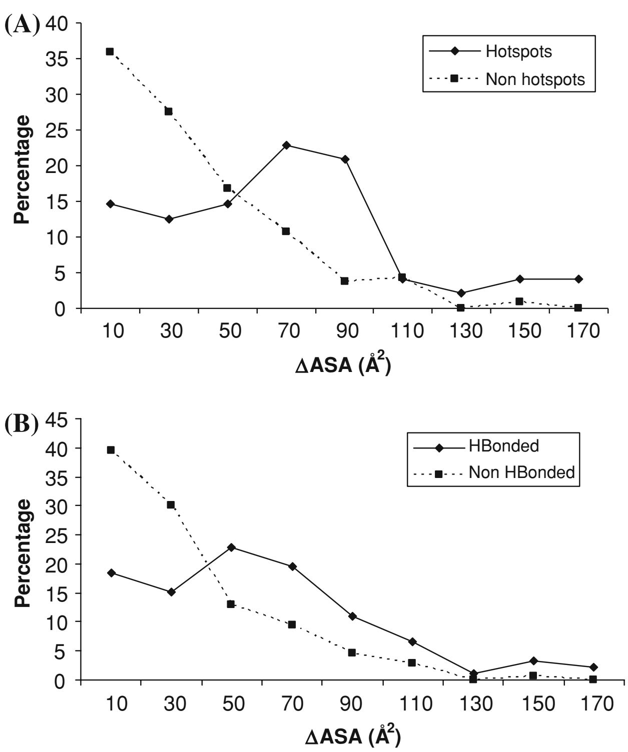

Fig. 5 Percentage of **a** hot spot ($\Delta\Delta G_{exp} \geq 2$ kcal/mol), and non-hotspot interface residues ($\Delta\Delta G_{exp} < 2$ kcal/mol), and, **b** hydrogen bonded and non-hydrogen-bonded interface residues (taken from ASEdb [34]) plotted as a function of the extent of burial in the interface. Along the *x*-axis, ΔASA (the extent of residue burial) has been divided into bins of 20 Å^2, and the percentage of residues in them belonging to each category is shown

to predict the $\Delta\Delta G$ of X to Ala mutations of interface residues (Table 1). Almost 70% and more than 80% of experimental binding warm/hot and hot spots (with $\Delta\Delta G \geq 1$ kcal/mol, and $\geq$2 kcal/mol, respectively) could be identified correctly using the method. The interplay between residue burial and the presence of polar and nonpolar components is also underlined in Fig. 1, which considers the role of individual residues to the binding. For a residue to provide substantial binding energy it must bury at least 45 Å^2 of the surface area (most likely located in the core of the interface) and also be capable of forming hydrogen bond(s).

There are some limitations to the present method, which however are common to many other approaches as well. For example, the coupling effects between different mutations cannot be taken into account explicitly, and multiple mutations are always assumed to be additive. Also, for the same reason, indirect effects on the binding energy exerted by residues not making direct interactions in the interface [8] are generally not captured (as we saw in the interface between human growth hormone and its receptor, 1a22, where only half the hot spots were correctly identified; Fig. 4c). It is also assumed that there are no conformational changes upon binding (results presented in Fig. 4c, d). However, the exclusion of interface water molecules does not seem to have an adverse effect, as a majority of the hot spots could be correctly identified (Fig. 4a, b). Compared to hydrogen bonding interactions, Coulombic effects play a negligible role in binding predictions [14] and our model performs well even without considering Coulomb energies, which however, are more important in controlling the kinetics of association [44].

In conclusion, a method is discussed that is computationally inexpensive and conceptually simple, for translating parameters (ΔASA) derived from the structure of the protein complex into energies. Predicting hot spots is not only important toward drug design and in depth understanding of protein interactions, they are also useful in the prediction of a protein’s function by modeling its interactions with other proteins. There are methods, such as PRISM [45] for predicting protein interactions based on geometric matching of surface patches with known interfaces, coupled with the identification of structurally conserved residues. The present method of computing hot spots would assist such algorithms by filtering out false positives. In addition, the knowledge of important interface residues (Fig. 1) would be useful for docking studies [46].

Acknowledgments We are grateful to Prof. Joël Janin for his comments on the manuscript. MG is a recipient of a senior research fellowship from CSIR and PC is a JC Bose National Fellow. The funding was provided by DBT, India.

References

1. Jones S, Thornton JM (1996) Principles of protein–protein interactions. Proc Natl Acad Sci USA 93:13–20. doi:10.1073/pnas.93.1.13
2. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. J Mol Biol 285:2177–2198. doi:10.1006/jmbi.1998.2439
3. Chakrabarti P, Janin J (2002) Dissecting protein–protein recognition sites. Proteins 47:334–343. doi:10.1002/prot.10085
4. Neuvirth H, Raz R, Schreiber G (2004) ProMate: a structure based prediction program to identify the location of protein–protein binding sites. J Mol Biol 338:181–199. doi:10.1016/j.jmb.2004.02.040
5. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone–receptor interface. Science 267:383–386. doi:10.1126/science.7529940
6. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9. doi:10.1006/jmbi.1998.1843
7. DeLano WL, Ultsch MH, de Vos AM, Wells JA (2000) Convergent solutions to binding at a protein–protein interface. Science 287:1279–1283. doi:10.1126/science.287.5456.1279
8. DeLano WL (2002) Unraveling hot spots in binding interfaces: progress and challenges. Curr Opin Struct Biol 12:14–20. doi:10.1016/S0959-440X(02)00283-X
9. Eisenberg D, Marcotte EM, Xenarios I, Yeates TO (2000) Protein function in the post-genomic era. Nature 405:823–826. doi:10.1038/35015694
10. Keskin O, Ma B, Nussinov R (2005) Hot regions in protein–protein interactions: the organization and contribution of structurally conserved hot spots residues. J Mol Biol 345:1281–1294. doi:10.1016/j.jmb.2004.10.077
11. Massova I, Kollman PA (1999) Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 121:8133–8143. doi:10.1021/ja990935j
12. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387. doi:10.1016/S0022-2836(02)00442-4
13. Huo S, Massova I, Kollman PA (2002) Computational alanine scanning of the 1:1 human growth hormone–receptor complex. J Comput Chem 23:15–27. doi:10.1002/jcc.1153
14. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein–protein complexes. Proc Natl Acad Sci USA 99:14116–14121. doi:10.1073/pnas.202485799
15. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein–protein interfaces. Sci STKE 219:pl2
16. Pokala N, Handel TM (2005) Energy functions for protein design: adjustment with protein–protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. J Mol Biol 347:203–227. doi:10.1016/j.jmb.2004.12.019
17. Almlöf M, Aqvist J, Smalås AO, Brandsdal BO (2006) Probing the effect of point mutations at protein–protein interfaces with free energy calculations. Biophys J 90:433–442. doi:10.1529/biophysj.105.073239
18. Moreira IS, Fernandes PA, Ramos MJ (2006) Unraveling the importance of protein–protein interaction: application of a computational alanine-scanning mutagenesis to the study of the IgG1 streptococcal protein G (C2 fragment) complex. J Phys Chem B 110:10962–10969. doi:10.1021/jp054760d
19. Ofran Y, Rost B (2007) Protein–protein interaction hotspots carved into sequences. PLoS Comput Biol 3:e119. doi:10.1371/journal.pcbi.0030119
20. Moreira IS, Fernandes PA, Ramos MJ (2007) Computational alanine scanning mutagenesis—an improved methodological approach. J Comput Chem 28:644–654. doi:10.1002/jcc.20566
21. Guney E, Tuncbag N, Keskin O, Gursoy A (2008) HotSprint: database of computational hot spots in protein interfaces. Nucleic Acids Res 36:D662–D666. doi:10.1093/nar/gkm813
22. Kauzmann W (1959) Some factors in the interpretation of protein denaturation. Adv Protein Chem 14:1–63. doi:10.1016/S0065-3233(08)60608-7
23. Chothia C, Janin J (1975) Principles of protein–protein recognition. Nature 256:705–708. doi:10.1038/256705a0
24. Janin J, Chothia C (1990) The structure of protein–protein recognition sites. J Biol Chem 265:16027–16030
25. Young L, Jernigan RL, Covell DG (1994) A role for surface hydrophobicity in protein–protein recognition. Protein Sci 3:717–729
26. Tsai C-J, Lin SL, Wolfson HJ, Nussinov R (1997) Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci 6:53–64
27. Li Y, Huang Y, Swaminathan CP, Smith-Gill SJ, Mariuzza RA (2005) Magnitude of the hydrophobic effect at central versus peripheral sites in protein–protein interfaces. Structure 13:297–307. doi:10.1016/j.str.2004.12.012
28. Xu D, Lin SL, Nussinov R (1997) Protein binding versus protein folding: the role of hydrophilic bridges in protein associations. J Mol Biol 265:68–84. doi:10.1006/jmbi.1996.0712
29. Fernández A, Scheraga HA (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc Natl Acad Sci USA 100:113–118. doi:10.1073/pnas.0136888100
30. Hu Z, Ma B, Wolfson H, Nussinov R (2000) Conservation of polar residues as hot spots at protein interfaces. Proteins 39:331–342. doi:10.1002/(SICI)1097-0134(20000601)39:4<331::AID-PROT60>3.0.CO;2-A
31. Guharoy M, Chakrabarti P (2005) Conservation and relative importance of residues across protein–protein interfaces. Proc Natl Acad Sci USA 102:15447–15452. doi:10.1073/pnas.0505425102
32. Chothia C (1974) Hydrophobic bonding and accessible surface area in protein. Nature 248:338–339. doi:10.1038/248338a0
33. Eisenhaber F (1996) Hydrophobic regions on protein surfaces. Derivation of the solvation energy from their area distribution in crystallographic protein structures. Protein Sci 5:1676–1686. doi:10.1002/pro.5560050821
34. Thorn KS, Bogan AA (2001) ASEdb: a database of alanine mutations and their effects on the free energy of binding in protein interactions. Bioinformatics 17:284–285. doi:10.1093/bioinformatics/17.3.284
35. ShajiKumar MD, Gromiha MM (2006) PINT: protein–protein interactions thermodynamic database. Nucleic Acids Res 34:D195–D198. doi:10.1093/nar/gkj017
36. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28:235–242. doi:10.1093/nar/28.1.235
37. Saha RP, Bahadur R, Pal A, Mandal S, Chakrabarti P (2006) ProFace: a server for the analysis of the physicochemical features of protein–protein interfaces. BMC Struct Biol 6:11. doi:10.1186/1472-6807-6-11
38. McDonald IK, Thornton JM (1994) Satisfying hydrogen bonding potential in proteins. J Mol Biol 238:777–793. doi:10.1006/jmbi.1994.1334
39. Guharoy M, Chakrabarti P (2007) Secondary structure based analysis and classification of biological interfaces: identification of binding motifs in protein–protein interactions. Bioinformatics 23:1909–1918. doi:10.1093/bioinformatics/btm274

40. Chothia C (1976) The nature of the accessible and buried surfaces in proteins. J Mol Biol 105:1–12. doi:10.1016/0022-2836(76)90191-1
41. Rodier F, Bahadur RP, Chakrabarti P, Janin J (2005) Hydration of protein–protein interfaces. Proteins 60:36–45. doi:10.1002/prot.20478
42. Clackson T, Ultsch MH, Wells JA, de Vos AM (1998) Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. J Mol Biol 277:1111–1128. doi:10.1006/jmbi.1998.1669
43. Albeck S, Unger R, Schreiber G (2000) Evaluation of direct and cooperative contribution towards the strength of buried hydrogen bonds and salt bridges. J Mol Biol 298:503–520. doi:10.1006/jmbi.2000.3656
44. Schreiber G, Shaul Y, Gottschalk KE (2006) Electrostatic design of protein–protein association rates. Methods Mol Biol 340:235–249
45. Ogmen U, Keskin O, Aytuna AS, Nussinov R, Gursoy A (2005) PRISM: protein interactions by structural matching. Nucleic Acids Res 33:W331–W336. doi:10.1093/nar/gki585
46. Halperin I, Wolfson H, Nussinov R (2004) Protein–protein interactions: coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. Structure 12:1027–1038. doi:10.1016/j.str.2004.04.009