Side-chain conformational entropy at protein–protein interfaces

CHRISTIAN COLE AND JIM WARWICKER

Department of Biomolecular Sciences, UMIST, Manchester M60 1QD, UK

(RECEIVED July 3, 2002; FINAL REVISION September 13, 2002; ACCEPTED September 23, 2002)

Abstract

Protein–protein interactions are the key to many biological processes. How proteins selectively and correctly associate with their required protein partner(s) is still unclear. Previous studies of this “protein-docking problem” have found that shape complementarity is a major determinant of interaction, but the detailed balance of energy contributions to association remains unclear. This study estimates side-chain conformational entropy (per unit solvent accessible area) for various protein surface regions, using a self-consistent mean field calculation of rotamer probabilities. Interfacial surface regions were less flexible than the rest of the protein surface for calculations with monomers extracted from homodimer datasets in 21 of 25 cases, and in 8 of 9 for the large protomer from heterodimer datasets. In surface patch analysis, based on side-chain conformational entropy, 68% of true interfaces were ranked top for the homodimer set and 66% for the large protomer/heterodimer set. The results indicate that addition of a side-chain entropic term could significantly improve empirical calculations of protein–protein association.

Keywords: Conformational entropy; rotamers; dimerization; protein–protein interactions

Protein–protein interactions play a key role in many cellular mechanisms. Proteomics has made possible the large-scale investigation of the protein “interactome” via yeast two-hybrid methodologies (Ito et al. 2001). Other studies probe protein–protein interactions in vivo, for example, using bioluminescence optical imaging (Ray et al. 2002) or positron-emission tomography via an inducible reporter gene (Luker et al. 2002). The large amount of data generated by these types of studies will inform on the specificity of protein–protein interactions, but not on the underlying molecular mechanisms that code for such selectivity.

An understanding of the molecular details of protein recognition sites would allow for the automated modeling of protein complexes from known monomer structures. This is pursued in the computational field of protein docking, whereby relative conformational space is searched and the resulting conformers ranked according to some force field

or scoring function (Smith and Sternberg 2002). Surface shape complementarity is the most accurate predictor of protein–protein complexes at present (Ritchie and Kemp 2000). However, this is most effective with protein components extracted from a known complex, rather than component structures that have been experimentally determined outside of the complex.

It is intriguing that promising complexes can be docked with the relatively simple potentials that describe shape complementarity, and that surface plasticity is a critical factor beyond this (Brady and Sharp 1997; Kimura et al. 2001). Finding the most discriminating potential, including shape variability, is an important goal if accurate prediction of protein complexes is to be achieved. One approach to identifying elements of an effective potential for ranking docked configurations is to characterize the properties of known interfaces. This can be achieved by examining the chemical characteristics and residue propensities of the interfacial region in the context of the rest of the protein surface (Lo Conte et al. 1999; Valdar and Thornton 2001a, 2001b), or by carrying out different ranking methodologies on arbitrary surface regions (Jones and Thornton 1997a, 1997b). In the latter method key features of protein–protein interactions

Reprint requests to: Jim Warwicker, Department of Biomolecular Sciences, UMIST, PO Box 88, Manchester M60 1QD, UK; e-mail: jim.warwicker@umist.ac.uk; fax: 44 (0)161 236 0409.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.0222702>.

are determined using "surface patches," whereby equal-sized regions of the protein surface are characterized in terms of a variety of factors, such as planarity, hydrophobicity, and residue propensity. Another useful approach is to use evolutionary data to identify conserved residues on the protein surface, as these tend to be involved in binding and/or recognition (Lichtarge et al. 1997; Hu et al. 2000; Elcock and McCammon 2001). This method is less useful for structures without an adequate number of close homologs.

The current study extends analysis of interfacial properties (Ponstingl et al. 2000; Valdar and Thornton 2001a, 2001b) with the addition of a quantity that is the ratio of the estimated change in side-chain conformational entropy (S) upon complexation and the solvent accessible surface area (A). This quantity was calculated for total surface area (S/A), and was examined both for individual residues and over surface patches (Jones and Thornton 1997a). The entropic and accessible area components are likely to be major contributors to binding/docking. While precise two-component docking will be determined by complementarity, of interest in this initial study is the intrinsic propensity of a surface, in terms of minimal side-chain entropy loss for a maximal surface area burial.

The loss of side-chain conformational entropy upon complexation has been discussed in the context of protein folding (Lee et al. 1994; Creamer 2000) and protein-protein complexation (Doig and Sternberg 1995; Brady and Sharp 1997). These discussions have given rise to values of about $1.5R$ per residue equivalent to 3.74 kJ mol^{-1} at 300 K, upon folding. Estimates of side-chain conformational entropy were made with a modification of the mean field algorithm of Koehl and Delarue (Koehl and Delarue 1994). In scanning residues or areas at an interface, it is assumed that monomer side-chain conformational entropy and solvent accessible surface of each interfacial residue would be lost in the complex. It may be expected that more favorable binding surfaces correlate with a smaller loss of side-chain conformational entropy and a larger burial of (nonpolar) surface area. The hypothesis that the magnitude of S/A may be smaller for interfacial regions is tested for a set of 25 homodimers and 14 heterodimers (Jones and Thornton 1997a).

Results

Side-chain conformational entropy at interfaces

Initially, side-chain conformational entropy (S) was estimated for each residue in both the extracted monomer and dimeric states of each protein. Residues with a change in S (ΔS) were found to be either interfacial or in direct contact with interfacial residues. Interfacial residues were assessed as those with changes in solvent accessible area on dimerization. Residues with only one rotamer (Pro, Gly and Ala)

as well as disulphide bridges (Cys) were fixed in the mean field calculations.

Some interfacial residues were found to have zero or very small ΔS upon complexation. These were mainly residues with relatively few allowed rotamers (e.g., Thr, Ser, Val), for which a small ΔS would be expected. One or two more flexible residues (e.g., Glu, Gln, His) in each interface were also found to have a relatively small ΔS . This lack of change upon complexation was in conjunction with a low monomeric S value, suggesting that they are in a conformation that is favorable for dimer formation.

To investigate differences between interfacial and noninterfacial surfaces, separate $\Sigma S/\Sigma A$ values were calculated, where these quantities are given in units of R per $100 \text{ \AA}^2$. As seen in Table 1, the side chains of 21 out of the 25 homodimers were less flexible (up to 34.9%) at the interface than elsewhere. The heterodimer data (Table 2) shows a similar trend for the large protomer set, but the small protomer set has interfaces that are only slightly less flexible than the overall surface. Overall, surfaces in the small protomer set exhibit less flexibility (per unit area) than those in the large protomer or homodimer sets.

Calculated S/A can be viewed graphically with color-coded surfaces. Figure 1 compares this surface (using 1pp2, a good, but not anomalous example: the color-coded surfaces tend to follow qualitatively with S/A) with one color coded by crystallographic B-factor, representing disorder in the crystal. Figure 1A and B, originating from the side-chain conformational entropy calculations, show a better graphical correlation of low flexibility and interface than does B-factor (Fig. 1C), indicating that $\Sigma S/\Sigma A$ provides information beyond that found in the crystallographic B-factor. In terms of a flexibility measure, $\Sigma S/\Sigma A$ will be less dependent on crystal contacts than the B-factor.

To test whether differential flexibility between interfacial and noninterfacial residues for the homodimers was encoded simply in residue types (in particular small, less flexible residues) the frequency of residue occurrence was plotted (Fig. 2). It is apparent that nonpolar residues (e.g., Leu, Met, Phe, Val) are generally more prevalent at the interface and that polar residues (i.e., Asp, Glu, Lys) are more common at the noninterfacial surface, in agreement with previous studies (Jones and Thornton 1996; Lo Conte et al. 1999; Glaser et al. 2001).

However, it is not clear from this figure whether there are intrinsically more inflexible residues at the interface. Thus, the difference in the average number of (unrestricted) rotamers per residue was determined for each region of the protein surface. For each residue type at the interface, leucine for example, the total number of allowed rotamers was determined (nine, as defined by Tuffery et al., 1997) and then multiplied by the number of leucines at the interface, four for 1pp2, to yield 36 available rotamers. Over all residue types at interfacial and noninterfacial regions average

Table 1. Side-chain conformational entropy, intrinsic flexibility, and patch analysis data for 25 homodimer structures

Structure	$\Sigma S/\Sigma A^a$			% Difference	Rotamer difference ^b	Patch analysis rank
	Interface	Noninterface	Difference			
1cdt	1.54	1.73	0.19	10.8	4.55	4
1g6n	1.89	1.98	0.09	4.7	-2.12	2
1il8	1.40	1.73	0.33	19.2	5.49	1
1msb	1.28	1.50	0.22	15.2	2.28	3
1pp2	1.54	2.04	0.50	24.6	1.52	1
1utg	1.43	1.62	0.19	11.8	3.94	1
1ypi	1.82	2.14	0.32	15.0	-0.73	1
2ccy ^c	2.33	1.55	-0.78	-50.7	-4.20	8
2cts	1.65	2.46	0.81	33.2	2.56	1
2gn5	1.90	2.08	0.18	8.7	-0.94	1
2rus	1.85	2.16	0.31	14.3	-0.90	1
2rve ^d	1.72	2.40	0.68	28.4	2.28	1
2sod	1.73	2.04	0.31	15.1	0.50	1
2ts1 ^c	1.83	2.28	0.45	19.6	2.47	1
2tsc	1.99	2.68	0.69	25.8	-1.04	1
2wrp ^c	1.70	1.50	-0.20	-13.5	-0.97	6
3aat	1.88	2.54	0.66	26.2	-0.95	1
3grs	1.68	2.22	0.54	24.4	-0.23	1
3sdh	2.20	2.17	-0.03	-1.4	-6.93	3
3sdp	1.63	1.80	0.17	9.5	-1.87	2
3ssi	1.32	1.31	-0.01	-0.4	-1.11	4
4mdh	1.92	2.44	0.52	21.0	0.58	1
5adh	1.85	2.33	0.48	20.7	2.69	1
5hvp ^d	1.52	2.33	0.81	34.9	6.83	1
8tim	2.07	2.19	0.12	5.6	-2.55	1
Mean	1.75	2.05	0.30	12.9	0.45	

^a $\Sigma S/\Sigma A$ units are R/100 Å².

^b Difference in the average number of rotamers per residue between regions, positive giving more intrinsic flexibility at the noninterfacial surface and negative more intrinsic flexibility at the interface.

^c Missing side-chain atoms, rebuilt with QUANTA.

^d Structure includes a ligand that binds at the dimer interface.

numbers of rotamers per residue were calculated for each region of the surface such that, for 1pp2, there are 7.89 rotamers per residue at the interface and 9.42 elsewhere. The difference between the average values (interface versus noninterface surfaces) relates to differences in the intrinsic flexibility of side chains, without consideration for the rotameric restriction that the mean field algorithm estimates. A negative value is returned where the interface has more intrinsic flexibility and a positive value (e.g., 1.53 for 1pp2) where the rest of the surface has more intrinsic flexibility (Tables 1, 2). Figure 3 shows only a weak correlation between this intrinsic flexibility (without 3D restrictions) and the difference in $\Sigma S/\Sigma A$. This result suggests that although simple occurrence of residue type contributes to conformational flexibility at the homodimer interface, a large part of the decreased flexibility generally observed at homodimer interfaces (Table 1) is likely to arise from the rotameric restrictions that are derived from the mean field calculations.

Patch analysis using side-chain conformational entropy

Having found that the interface is generally less flexible than the rest of the surface for the set of homodimers, it is of interest to establish whether this property could, in principle, be used in a predictive approach. Patch analysis searches the whole surface of a monomer with, initially, equal-sized patches, and all the patches can be ranked in terms of a particular property (Jones and Thornton 1997a). In this study, overlap with the experimental interface and $\Sigma S/\Sigma A$ were determined for each patch.

Initial calculations showed that 68% of the interfacial patches over the 25 homodimers, 66% over the nine larger protomer heterodimers, but only 8% over the 13 smaller protomer heterodimers appeared in the first 10 percentile (i.e., ranked first) for patch analysis (Table 1, Fig. 4). As previously demonstrated, the difference in $\Sigma S/\Sigma A$ at the interface and the rest of the protein surface is not mainly due to any intrinsic inflexibility of the side chains at the inter-

Table 2. Side-chain conformational entropy, intrinsic flexibility, and patch analysis data for 14 heterodimer structures

	$\Sigma S/\Sigma A^a$					
Structure	Interface	Noninterface	Difference	% Difference	Rotamer difference ^b	Patch analysis rank
<i>Larger protomer</i>						
1acb	1.87	2.35	0.48	20.5	2.55	1
1bgs	1.69	1.24	−0.45	−36.0	−5.86	6
1cse	1.35	1.49	0.14	9.4	3.08	3
1fss ^c	1.97	2.19	0.22	10.0	0.47	1
1gla ^c	1.80	2.44	0.64	26.1	−3.46	1
1smp ^c	1.74	1.75	0.01	0.5	−0.20	3
1udi	1.40	2.00	0.60	30.0	1.98	1
2btf	2.02	2.18	0.16	7.5	−4.69	1
2pcb	1.83	2.66	0.83	31.3	−0.12	1
Mean	1.74	2.03	0.29	11.0	−0.69	
<i>Smaller protomer</i>						
1acb	1.67	1.86	0.19	10.2	−1.22	2
1bgs	1.11	1.90	0.79	41.4	7.75	2
1cho	1.67	1.55	−0.12	−7.8	−6.72	3
1fss ^c	1.58	1.59	0.01	0.4	−8.28	4
1gla ^c	1.85	1.94	0.09	4.6	−0.48	3
1mct	1.24	0.98	−0.26	−27.4	−6.54	9
1smp ^c	1.44	1.74	0.30	17.3	−0.19	2
1tab	1.40	1.25	−0.15	−12.5	−5.62	9
1udi	1.65	2.05	0.40	19.4	4.11	1
2btf	2.07	1.85	−0.22	−11.6	−2.54	3
2pcb	2.51	2.27	−0.24	−10.5	−11.45	6
2ptc	1.38	1.57	0.19	12.6	−1.73	4
2sic	1.67	1.34	−0.33	−24.5	−4.07	4
Mean	1.63	1.68	0.05	0.9	−2.84	

^a $\Sigma S/\Sigma A$ units are R/100 Å².^b Difference in the average number of rotamers per residue between regions, positive giving more intrinsic flexibility at the noninterfacial surface region and negative more intrinsic flexibility at the interface.^c Missing side-chain atoms, rebuilt with QUANTA.

face, but is largely due to side-chain restriction with respect to the rest of the protein surface. Therefore, patch analysis was carried out over a range of patch sizes in an attempt to analyze the interface and, in particular, any specific residues which dominate the interface.

The variable patch size analysis output could rank each patch for an individual protein either by $\Sigma S/\Sigma A$ or by percentage overlap with the true interface. It should be noted that in this method not one of the randomly generated patches completely overlaps with the true interface; the range for the most overlapping patch in an individual protein is 51.9%–95.2%. This is indicative of the shape difference between the automated patches and the true interface, with the automated patches being approximately circular and contiguous over the surface, which is rarely the case for the true interface. Plotting the top hit of each against patch size for a structure yields different information (Fig. 5). The top $\Sigma S/\Sigma A$ hit (most inflexible patch) line shows that after an initial decrease in $\Sigma S/\Sigma A$ over the first six patches there is a steady increase in $\Sigma S/\Sigma A$ as the patch size increases,

indicating that beyond a threshold patch size (e.g., six for 3sdh) $\Sigma S/\Sigma A$ will tend to increase with the number of residues in each patch.

The plot of the top percentage overlap hit shows a different pattern. The general trend of the data is also an increase in $\Sigma S/\Sigma A$ as the patch size increases. However, the individual patches are generally more variable with distinct peaks and troughs in the data. The troughs identify regions of low $\Sigma S/\Sigma A$ in the interfacial surface region. This allows us to determine whether the interface is uniformly inflexible or is dominated by subregions of inflexibility. For example, in the plot for 3sdh (Fig. 5) there are four troughs in the “most overlapping” data at 5, 12, 20, and 25 patch sizes. These troughs correspond to patches centered on residues 72, 69, 92, and 93, respectively. This reveals two distinct interfacial regions (69/72 and 92/93) with low conformational flexibility for 3sdh. The top, least inflexible line in Figure 5 is the upper boundary for side-chain conformational entropy on the protein surface, with the most overlapping patches found between the least and the most in-

Fig. 1. Protein surface of 1pp2 monomer complexed with a backbone representation of the other monomer, indicating the interfacial region. The surface is colored by side-chain conformational entropy per residue (A), by side-chain conformational entropy per $\text{\AA}^2$ (B), and by B-factor (C). In all panels, red indicates regions of low conformational flexibility; blue, regions of high conformational flexibility; and orange/yellow/green, the range in between.

flexible patches, with the troughs approaching the most inflexible patches.

Additionally, the interface can be analyzed on a per residue basis to see if there are residues that dominate (Fig. 6). Generally, only two or three residues per interface stand out. They tend to be either fixed residues which are relatively solvent exposed (e.g., Val or Pro), thereby contributing low side-chain flexibility, or large residues (e.g., Lys), which confer higher flexibility to the interface. However, these stand-out residues still only alter the average $\Sigma S/\Sigma A$ for the interfacial patch by at most $\pm 8\%$.

The calculations can also be used to determine the difference in the overall side-chain rotamer conformational entropy of the protein between the monomer and the dimer conformations (ΔS_{sc}). Buried nonpolar area can also be determined for the structures, and an associated free energy estimated with an effective surface tension of 0.1 kJ mol^{-1} per $\text{\AA}^2$ of buried nonpolar area, that is within the generally used range (Raschke et al. 2001). At 300 K these contributions to protein dimerization (ΔG_{np} , ΔG_{sc}) are estimated in Table 3. Nonpolar burial determines protein folded state stability, and is generally a major factor in protein–protein complexation. Our calculations indicate that ΔG_{sc} can also make a significant (unfavorable) contribution to binding, consistent with the view that evolved modulation of ΔG_{sc} will impact on affinity.

Discussion

Here, we have shown that the interface of protein–protein complexes differs somewhat from the remaining protein surface in terms of the side-chain flexibility per unit area. On average, the interface was found to be less flexible for the 21 out of 25 homodimer complexes analyzed. Graphics examination revealed that $\Sigma S/\Sigma A$ gives insight beyond that from crystallographic B-factor or intrinsic side-chain flexibility at the interface.

Of the four homodimers that are more flexible at the interface than elsewhere, two (3sdh and 3ssi) have very small differences of -0.03 and -0.01 , respectively. Of all the homodimers, 3ssi has the lowest overall $\Sigma S/\Sigma A$ (Table 1), and ΔS_{sc} (dimerization) is small (Table 3). These features may relate to the extensive exposed β -sheet structure, which is partially used to form the dimer interface. Several water molecules are involved in a hydrogen-bonding network across the interface for 3sdh. Such interactions are not included in our analysis, but they may be indicative of a degree of functional flexibility in this dimeric haemoglobin interface (Royer 1994). Indeed, this flexibility could underlie our observation of relatively high $\Sigma S/\Sigma A$ for the interface in a monomer context.

Structures 2wrp and 2ccy have relatively large negative differences between the interfacial and noninterfacial regions (Table 1). Tryptophan repressor (2wrp) is an intertwined dimer, so that reference to an extracted monomer state is probably inappropriate in this case. The very large negative $\Sigma S/\Sigma A$ difference value for the cytochrome *c'* 2ccy (Finzel et al. 1985) is due to the highest conformational entropy at the interface for the whole dataset and a relatively low conformational entropy at the noninterface (Table 1). In an attempt to determine whether this was an isolated example, homologous cytochrome *c'* proteins from four sources were also examined *Chromatium vinosum*: 1bbh (Ren et al. 1993), *Alcaligenes denitrificans*: 1cgo (Dobbs et

Fig. 2. Cumulative frequency of residues occurring at the interface and noninterface, adjusted such that both sets are comparable for the homodimer set.

al. 1996), *Alcaligenes xylosoxidans*: 1e83 (Lawson et al. 2000), and *Rhodocyclus gelatinosus*: 1jaf (Archer et al. 1997) (Table 4). All the structures are four-helical bundle homodimers with a monomeric C_{α} RMSD of ~ 1.9 Å from 2ccy using the combinatorial extension method (Shindyalov and Bourne 1998). Despite the structural similarity, these structures yield a $\Sigma S/\Sigma A$ difference between -10.7% and 14.5% (Table 4), indicating that the underlying helical framework does not determine the original result with 2ccy. The range is most likely due to substantial sequence variation on top of a well-conserved dimeric structural framework.

Table 3 demonstrates that ΔS_{sc} , the loss of conformational entropy upon complexation, will make a significant contribution to calculations of binding affinity, which is borne out by our observation that the interface is generally less flexible than the rest of the protein surface suggesting that an interface can be preconditioned through restriction of side-chain rotamers. One might expect such an effect to be highlighted for systems with particularly tight binding. Table 5 details the $\Sigma S/\Sigma A$ differences for three endonuclease colicin structures and their associated immunity proteins: 1emv and 7cei are both DNases and 1e44 an RNase. DNase immunity proteins have dissociation constants in the

Fig. 3. Correlation between the difference in side-chain flexibility ($\Sigma S/\Sigma A$) and the difference in rotamers per residue for the homodimer set. $R^2 = 0.34$.

Fig. 4. Histogram detailing the ranking of experimental interfaces by patch analysis for a set of 25 homodimers (filled), 9 large protomer heterodimers (diagonal stripes), and 12 small protomer heterodimers (empty).

Fig. 5. Variable patch size analysis for 3sdh. Generated patches can be ranked by one of three ways in terms of side-chain conformational entropy ($\Sigma S/\Sigma A$): least inflexible patch (open squares), patch most overlapping with the true interface (open circles), and most inflexible patch (open triangles). The least inflexible patches for patch sizes <5 are curtailed at $6R/100 \text{ \AA}^2$.

femtomolar region (Wallis et al. 1995), and have significant structural and sequence similarities (Kühlmann et al. 2000). The RNases are structurally dissimilar to the DNases, and their binding is less well characterized.

Table 5 shows that 1emv and 7cei immunity proteins have the largest $\Sigma S/\Sigma A$ % difference of all the dimers here analyzed (38.1 and 43.3%, respectively). In addition, they have the lowest interfacial $\Sigma S/\Sigma A$, suggesting that the large % difference is due to the interface being extremely inflexible. The endonuclease partners to the immunity proteins for 1emv and 7cei do not show this level of interfacial inflexibility, reflected also in relatively low rankings in patch analysis (Table 5). Interestingly, the RNase structure (1e44), which has a different mode of binding, shows the

opposite trend with the endonuclease having the lower interfacial flexibility and top ranking patch. In each of the colicin/immunity protein systems studied, one or other of

Table 3. Estimated contributions to complexation for nonpolar burial and side-chain entropy loss

Structure	ΔA_{np} ($\text{\AA}^2$)	$\Delta S_{sc}/R$	ΔG_{np}^a	ΔG_{sc}^b
1cdt	-295	-3.7	-29.5	9.3
1g6n	-1260	-10.7	-126.0	26.8
1il8	-478	-5.2	-47.8	13.0
1msb	-481	-5.1	-48.1	12.8
1pp2	-982	-2.6	-98.2	6.5
1utg	-1202	-11.9	-120.2	29.8
1ypi	-958	-15.1	-95.8	37.8
2ccy	-601	-5.9	-60.1	14.8
2cts	-3330	-26.6	-333.0	66.5
2gn5	-542	-1.4	-54.2	3.5
2rus	-1967	-20.9	-196.7	45.75
2rve	-887	-4.3	-88.7	10.8
2sod	-513	-3.9	-51.3	9.8
2ts1	-1215	-7.8	-121.5	19.5
2tsc	-1412	-20.1	-141.2	50.3
2wvp	-1790	-18.8	-179.0	47
3aat	-1941	-34.2	-194.1	85.5
3grs	-2303	-24.0	-230.3	60
3shd	-524	-9.3	-52.4	23.3
3sdp	-591	-4.0	-59.1	10
3ssi	-618	-2.2	-61.8	5.5
4mdh	-1098	-13.6	-109.8	34.0
5adh	-1095	-11.7	-109.5	29.3
5hvp	-1125	-11.9	-112.5	29.8
8tim	-1043	-13.0	-104.3	32.5

^a $\Delta G_{np} = 0.1 \times \Delta A_{np}$, kJmol^{-1} units.

^b $\Delta G_{sc} = -T\Delta S_{sc}$, $T = 300 \text{ K}$, kJmol^{-1} units.

Fig. 6. Histogram detailing the relative contributions to the overall inflexibility of the interface for 1msb. Each interfacial residue is removed, and the side-chain conformational entropy ($\Sigma S/\Sigma A$) for the interface is recalculated. A positive change in $\Sigma S/\Sigma A$ indicates residues that add flexibility to the interface, and a negative change indicates residues that add rigidity to the interface.

Table 4. Sequence, structural, and side-chain entropy comparisons of four cytochrome *c'* proteins with cytochrome *c'* structure 2ccy

Structure	RMSD from 2ccy ^a C _α (Å)	Sequence identity to 2ccy (%) ^b	$\Sigma S/\Sigma A^c$		
			Interface	Noninterface	Difference (%)
1bbh	1.8	22.4 (20.4)	1.74	1.97	14.5
1cgo	1.9	34.1 (32.8)	1.68	1.74	4.3
1e83	2.0	33.9 (32.8)	2.13	1.95	-10.7
1jaf	1.8	38.2 (36.2)	2.01	1.91	-6.4

^a Determined using combinatorial extension (Shindyalov and Bourne 1998).^b Brackets, sequence identity of only the interfacial region (residues 1–60).^c $\Sigma S/\Sigma A$ units are R/100 Å².

the interacting partners gives a clear indication of significantly reduced interfacial flexibility, consistent with tight binding.

In an attempt to explore further the relationship between the strength of dimer association and side-chain conformational entropy a set of known structures with experimentally determined association free energies (Horton and Lewis 1992) were studied (Table 6). While the interfacial regions generally exhibit lower $\Sigma S/\Sigma A$ than noninterfacial regions, there is no correlation apparent between this property and the association free energy. This result is consistent with a view of interfacial energetics in which binding energy results from a complex combination of multiple factors.

For the colicin/immunity protein complexes and the heterodimer dataset, one of the interacting partners generally exhibits a significantly lower $\Sigma S/\Sigma A$ at the interface than elsewhere. For the heterodimers, this is mostly the larger protomer. Whereas $\Sigma S/\Sigma A$ is about equal for interfacial regions of the larger and smaller protomers, it tends to be larger for noninterfacial regions in the larger protomer (Table 2). This could relate to considerations of monomer size with respect to the number of surface accessible residues. Larger protomers tend to have a more defined core region that has no solvent accessibility and has different

conformational entropy properties than the surface region. Smaller protomers have a less well-defined core region, which has partially exposed core residues that also contribute to the surface region. Thus, the more conformationally restricted core region in the smaller protomers will be included into the surface region for the calculation of $\Sigma S/\Sigma A$, thereby reducing surface conformational entropy relative to the larger protomer set (Table 2).

The fact that computational protein docking is so much easier with separated components from the complex, than with structures solved outside of the complex, points to a degree of induced fit mechanism for interface formation. Generally, a mean field algorithm (Koehl and Delarue 1994) for side-chain placement can play a role in refining computed dockings, but the lower resolution question that we have asked is whether there is a detectable interfacial signature in terms of side-chain flexibility per unit surface area. Our results indicate that this is the case, although the size of the signal is not constant over the systems studied, and will also be coupled to properties such as individual monomer (folded) stability and other binding energy components. Clearly, proteins that can adapt their binding sites to bind several different ligands will probably not have a low conformational entropy at the interface, and their binding en-

Table 5. Side-chain conformational freedom and patch analysis data for three endonuclease colicin proteins and their cognate immunity proteins

Structure	$\Sigma S/\Sigma A^a$			% Difference	Patch analysis rank
	Interface	Noninterface	Difference		
<i>Endonuclease</i>					
1emv	1.52	1.73	0.21	12.5	3
7cei	2.00	2.03	0.03	1.3	3
1e44	1.39	2.00	0.61	30.4	1
<i>Immunity protein</i>					
1emv	0.90	1.45	0.55	38.1	1
7cei	1.10	1.94	0.84	43.3	1
1e44	1.73	1.51	−0.22	−14.3	6

1emv and 7cei are DNases and 1e44 is an RNase.

^a $\Sigma S/\Sigma A$ units are R/100 Å².

Table 6. Comparison between experimentally determined free energies of association (ΔG_{obs}) and calculated side-chain conformational entropy ($\Sigma S/\Sigma A$) for each of the interacting proteins for a set of dimers

Dimer	Monomer	$\Sigma S/\Sigma A^a$		% Difference	ΔG_{obs}^c
		Interface	Noninterface		
lcho	α -chymotrypsin	1.52	2.01	24.5	-65.6
	OMTKY3 ^b	1.67	1.55	-7.8	
lcse	Subtilisin Carlsberg	1.35	1.49	9.4	-54.8
	Eglin-C	1.27	1.53	17.0	
lhbs	Sickle cell	2.16	2.94	26.7	-20.1
	Deoxyhaemoglobin	1.70	2.93	41.8	
ltpa	Anhydrotrypsin	2.08	2.13	2.1	-74.4
	BPTI ^b	1.38	1.55	10.7	
2kai	Kallikrein A	1.93	1.65	-17.1	-51.8
	BPTI ^b	1.25	1.44	12.7	
2ptc	β -trypsin	2.17	2.09	-3.8	-75.7
	BPTI ^b	1.38	1.57	12.6	
2sec	Subtilisin Carlsberg	1.44	1.47	1.7	-54.8
	N-acetyl Eglin C	1.53	1.61	5.0	
2ssi	Streptomyces subtilisin inhibitor dimer	1.32	1.31	-0.4	-66.9
2tpi	Trypsinogen	1.07	1.14	6.5	-75.7
	BPTI ^b	1.69	1.73	2.4	
3hfl	Fab fragment	1.58	1.85	14.8	-59.4
	Lysozyme	1.48	1.84	19.3	
3sgb	Proteinase B	1.45	1.98	26.7	-61.4
	OMTKY3 ^b	1.89	1.54	-22.5	
4cpa	Carboxypeptidase A	2.82	2.59	-8.6	-41.8
	Potato inhibitor	4.88	4.90	0.3	
4ins	Insulin	1.03	2.58	60.2	-30.9
	Dimer	0.84	0.91	8.3	

^a $\Sigma S/\Sigma A$ units are R/100 Å².^b BPTI: bovine pancreatic trypsin inhibitor, OMTKY3: turkey ovomucoid inhibitor third domain.^c Values taken from Horton and Lewis (1992) and have units of kJ mol⁻¹.

ergy is likely to be dominated by other components (DeLano et al. 2000).

This work lends itself to progression in at least two directions. First, addition of the $\Sigma S/\Sigma A$ quantity to the patch analysis approach to prediction of potential interface regions (Jones and Thornton 1997b) is promising, particularly if combined with clustering techniques to improve overlap between computationally generated patches and true interfaces. Second, the scale of calculated ΔS_{sc} suggests that such a term should be included in empirical estimates of binding affinities. Indeed, the range of ΔS_{sc} in Table 3 (about 80 kJ/mole) is equivalent to a change of $\sim 10^{14}$ in association constant, although these mean field values are probably overestimates considering that side-chain/side-chain correlations and other potential terms could lead to a narrowing of the rotamer probability distribution (Koehl and Delarue 1994). The first suggested direction relates to a monomer-based method for estimating interfacial propensity, while the second (ΔS_{sc}) method is relevant for analysis of complexes (either experimental or computed).

Materials and methods

Coordinates

Table 1 details the 25 homodimer structures (Jones and Thornton 1997a), obtained from the PDB (Berman et al. 2000). Obsolete structures were replaced by the current available structure. All water molecules were removed and alternate atom positions reduced to a single copy. Protein monomers were extracted from any higher order occurrences in the asymmetric unit. Residues not present in the coordinates were ignored, but missing side-chain atoms were rebuilt using QUANTA (Accelrys), which was also used for general graphics and surface visualization and manipulation.

Definition of the interface

Surface residues were classed as those with an accessible surface area (A) in the monomer of ≥ 0.1 Å². The interface was defined as the set of residues for which A decreased by ≥ 0.1 Å² upon dimerization. The probe radius was 1.4 Å. All HETATM records were retained except for nonfunctional ligands (e.g., glycol) and water oxygen atoms.

Side-chain conformational entropy

A mean field method for calculating side-chain/side-chain interactions in the context of a rotamer set (Koehl and Delarue 1994) has previously been adapted to look at the packing and maximum possible solvent accessibility of ionisable side-chains in proteins (J. Warwicker, unpubl.). Koehl and Delarue (1994) calculated the effective potential (E) on rotamer k of side-chain i as:

$$E(i,k) = U(x_{ik}) + U(x_{ik}, x_0) + \sum_{j=1, j \neq i}^N \sum_{l=1}^{K_j} CM(j,l) U(x_{ik}, x_{jl}) \quad (1)$$

where x_{ik} are the coordinates of the atoms of side-chain i in rotamer k ; $U(x_{ik})$ is the potential for this rotamer alone; $U(x_{ik}, x_0)$ is the potential for this rotamer in the context of fixed coordinates (main chain, C_β atoms, disulphides), $CM(j,l)$ is a conformational matrix element, giving the probability that the conformation of side-chain j is described by rotamer l ; $U(x_{ik}, x_{jl})$ is the potential between the k rotamer of side-chain i and the l rotamer of side-chain j . The double sum is over all side-chains j (1 to N , not equal to i), and over all rotamers (1 to K_j) for each side-chain j .

In looking at side-chain packing, the Lennard-Jones VdW interaction parameters used to describe the potential in Equation 1 (Koehl and Delarue 1994) were replaced with hard sphere collisions, such that VdW overlap is disallowed subject to an overall relaxation of VdW radii that is incremented until a packing solution is found. In this adaptation of the method, Equation 1 reduces to:

$$CM(i,k) = \left\{ \begin{array}{l} 1, \text{ no clash with} \\ \text{fixed atoms} \\ 0, \text{ clash with} \\ \text{fixed atoms} \end{array} \right\} \times \prod_{j=1, j \neq i}^N \sum_{l=1}^{K_j} \left\{ \begin{array}{l} 1, \text{ no clash} \\ \text{with } (j,l) \\ 0, \text{ clash} \\ \text{with } (j,l) \end{array} \right\} CM(j,l) \quad (2)$$

where rotamer probabilities for a single residue sum to 1, and the conformational matrix of rotamer probabilities is iterated to convergence, starting from a uniform distribution of probabilities for the rotamers of each residue. It is clear from Equation 2 that there will be no packing solution unless each residue possesses at least one rotamer with nonzero probability (i.e., not clashing with the fixed atoms and with at least one nonclashing rotamer for each other residue). The relaxation of VdW radii is incremented (typically in 0.1 Å steps) until a packing solution is found. With united atom VdW radii, a solution is generally found at around 0.7 Å relaxation, a value that is determined largely by side-chain/main-chain contacts.

The backbone-independent rotamer set of Tuffery (Tuffery et al. 1997) was used. In this adaptation of the algorithm, which was designed to survey reasonable alternative packing solutions for side-chains in known structures, the experimental rotamers were also included, thereby minimizing the required VdW relaxation. Where studying configurations for which experimental side-chain rotamers are not available, the Lennard-Jones potential formalism (Koehl and Delarue 1994) is preferable because the current algorithm would impose a uniform VdW relaxation across all interactions, leading to wide-scale clashes.

The conformational matrix was used to estimate the conformational entropy of side chains (Hill 1956; Koehl and Delarue 1994):

$$S_i = -R \sum_{k=1}^{K_i} CM(i,k) \ln[CM(i,k)] \quad (3)$$

where R is the universal gas constant. Although the experimental side-chain rotamers have been used in the CM evaluation (Equation 2), they are neglected for the sum over rotamers in Equation 3, to be consistent with the number of rotamers in the database.

Surface and interface properties

In the context of protein-protein interactions, the ratio of conformational entropy to solvent accessible area (A) was calculated for each residue (i), S_i/A_i , and this measure for a patch of residues (constituting either the true interface or a computed test patch) as $\Sigma S_i/\Sigma A_i$. Such sums over residues were calculated for various surface regions, and the i subscript dropped for simplicity. A_{np} is used to denote the nonpolar contribution to the total solvent accessible area, A . Percentage difference between the interfacial and the noninterfacial surface regions was calculated as:

$$\left(\frac{(\Sigma S/\Sigma A)^{\text{non-int}} - (\Sigma S/\Sigma A)^{\text{int}}}{(\Sigma S/\Sigma A)^{\text{non-int}}} \right) \times 100 \quad (4)$$

Patch analysis

As previously described (Jones and Thornton 1997a), each patch was defined with a central surface-accessible residue and built using a defined number of nearest-neighbor residues (based on C_α - C_α atom distances). Variable patch sizes were created by starting with $n = 1$ and continuing up to the size of the experimental interface. Solvent vectors (Jones and Thornton 1997a) were not applied to these patches. The number of patches generated per structure is a function of the patch size and the protein's surface area, but generally did not exceed 400.

Percentage overlap of the patches with the experimental interface was determined as:

$$\frac{N_{\text{overlap}}}{N_{\text{int}}} \times 100$$

where N_{overlap} is the number of residues present both at the experimental interface and in the patch, and N_{int} is the number of residues at the experimental interface.

Acknowledgments

The authors acknowledge funding from the United Kingdom Biotechnology and Biological Sciences Research Council and from the EU (grant reference FAIR-CT98-7020).

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

References

- Archer, M., Banci, L., Dikaya, E., and Romão, M.J. 1997. Crystal structure of cytochrome c' from *Rhodocyclus gelatinosus* and comparison with other cytochromes c' . *J. Biol. Inorganic Chem.* **2**: 611-622.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., and Bourne, P.E. 2000. The protein data bank. *Nucleic Acids Res.* **28**: 235-242.
- Brady, G.P. and Sharp, K.A. 1997. Entropy in protein folding and in protein-protein interactions. *Curr. Opin. Struct. Biol.* **7**: 215-221.
- Creamer, T.P. 2000. Side-chain conformational entropy in protein unfolded states. *Proteins Struct. Funct. Genet.* **40**: 443-450.

- DeLano, W.L., Ultsch, M.H., de Vos, A.M., and Wells, J.A. 2000. Convergent solutions to binding at a protein-protein interface. *Science* **287**: 1279–1283.
- Dobbs, A.J., Anderson, B.F., Faber, H.R., and Baker, E.N. 1996. Three-dimensional structure of cytochrome *c'* from two *Alcaligenes* species and the implications for four-helix bundle structures. *Acta Crystallogr. D* **D52**: 356–368.
- Doig, A.J. and Sternberg, M.J.E. 1995. Side-chain conformational entropy in protein-folding. *Protein Sci.* **4**: 2247–2251.
- Elcock, A.H. and McCammon, J.A. 2001. Identification of protein oligomerization states by analysis of interface conservation. *Proc. Natl. Acad. Sci.* **98**: 2990–2994.
- Finzel, B.C., Weber, P.C., Hardman, K.D., and Salemme, F.R. 1985. Structure of ferricytochrome *c'* from *Rhodospirillum rubrum* at 1.67 Å resolution. *J. Mol. Biol.* **186**: 627–643.
- Glaser, F., Steinberg, D.M., Vakser, I.A., and Ben-Tal, N. 2001. Residue frequencies and pairing preferences at protein-protein interfaces. *Proteins Struct. Funct. Genet.* **43**: 89–102.
- Hill, T.L. 1956. *Statistical mechanics*. McGraw Hill, New York.
- Horton, N. and Lewis, M. 1992. Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**: 169–181.
- Hu, Z., Ma, B., Wolfson, H., and Nussinov, R. 2000. Conservation of polar residues as hot spots at protein interfaces. *Proteins Struct. Funct. Genet.* **39**: 331–342.
- Ito, T., Chiba, T., and Yoshida, M. 2001. Exploring the protein interactome using comprehensive two-hybrid projects. *Trends Biotechnol.* **19**: S23–S27.
- Jones, S. and Thornton, J.M. 1996. Principles of protein-protein interactions. *Proc. Natl. Acad. Sci.* **93**: 13–20.
- . 1997a. Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**: 121–132.
- . 1997b. Prediction of protein-protein interaction sites using patch analysis. *J. Mol. Biol.* **272**: 133–143.
- Kimura, S.R., Brower, R.C., Vajda, S., and Camacho, C.J. 2001. Dynamical view of the positions of key side chains in protein-protein recognition. *Biophys. J.* **80**: 635–642.
- Koehl, P. and Delarue, M. 1994. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J. Mol. Biol.* **239**: 249–275.
- Kühlmann, U.C., Pommer, A.J., Moore, G.R., James, R., and Kleanthous, C. 2000. Specificity in protein-protein interactions: The structural basis for dual recognition in endonuclease colicin-immunity protein complexes. *J. Mol. Biol.* **301**: 1163–1178.
- Lawson, D.M., Stevenson, C.E.M., Andrew, C.R., and Eady, R.R. 2000. Unprecedented proximal binding of nitric oxide to heme: Implications for guanylate cyclase. *EMBO J.* **19**: 5661–5671.
- Lee, K.H., Xie, D., Freire, E., and Amzel, L.M. 1994. Estimation of changes in side-chain configurational entropy in binding and folding—General methods and application to helix formation. *Proteins Struct. Funct. Genet.* **20**: 68–84.
- Lichtarge, O., Yamamoto, K.R., and Cohen, F.E. 1997. Identification of functional surface of the zinc binding domains of intracellular receptors. *J. Mol. Biol.* **274**: 325–337.
- Lo Conte, L., Chothia, C., and Janin, J. 1999. The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**: 2177–2198.
- Luker, G.D., Sharma, V., Pica, C.M., Dahlheimer, J.L., Li, W., Ochesky, J., Ryan, C.E., Piwnica-Worms, H., and Piwnica-Worms, D. 2002. Noninvasive imaging of protein-protein interactions in living animals. *Proc. Natl. Acad. Sci.* **99**: 6961–6966.
- Ponstingl, H., Henrick, K., and Thornton, J. M. 2000. Discriminating between homodimeric and monomeric proteins in the crystalline state. *Proteins Struct. Funct. Genet.* **41**: 47–57.
- Raschke, T.M., Tsai, J., and Levitt, M. 2001. Quantification of the hydrophobic interaction by simulations of the aggregation of small hydrophobic solutes in water. *Proc. Natl. Acad. Sci.* **98**: 5965–5969.
- Ray, P., Pimenta, H., Paulmurugan, R., Berger, F., Phelps, M.E., Iyer, M., and Gambhir, S.S. 2002. Noninvasive quantitative imaging of protein-protein interactions in living subjects. *Proc. Natl. Acad. Sci.* **99**: 3105–3110.
- Ren, Z., Meyer, T., and McRee, D.E. 1993. Atomic structure of a cytochrome *c'* with an unusual ligand-controlled dimer dissociation at 1.8 Å resolution. *J. Mol. Biol.* **234**: 433–445.
- Ritchie, D.W. and Kemp, G.J.L. 2000. Protein docking using spherical polar Fourier correlations. *Proteins Struct. Funct. Genet.* **39**: 178–194.
- Royer, W.E. 1994. High-resolution crystallographic analysis of a co-operative dimer hemoglobin. *J. Mol. Biol.* **235**: 657–681.
- Shindyalov, I.N. and Bourne, P.E. 1998. Protein structure alignment by incremental combinatorial extension (CE) of the optimal path. *Protein Eng.* **11**: 739–747.
- Smith, G.R. and Sternberg, M.J.E. 2002. Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**: 28–35.
- Tuffery, P., Etchebest, C., and Hazout, S. 1997. Prediction of protein side chain conformations: A study on the influence of backbone accuracy on conformation stability in the rotamer space. *Protein Eng.* **10**: 361–372.
- Valdar, W.S.J. and Thornton, J.M. 2001a. Conservation helps to identify biologically relevant crystal contacts. *J. Mol. Biol.* **313**: 399–416.
- . 2001b. Protein-protein interfaces: Analysis of amino acid conservation in homodimers. *Proteins Struct. Funct. Genet.* **42**: 108–124.
- Wallis, R., Moore, G.R., James, R., and Kleanthous, C. 1995. Protein-protein interactions in colicin E9 DNase-immunity protein complexes. Diffusion-controlled association and femtomolar binding for the cognate complex. *Biochemistry* **34**: 13743–13750.