Article No. mb981843 J. Mol. Biol. (1998) 280, 1–9

JMB

COMMUNICATION

Anatomy of Hot Spots in Protein Interfaces

Andrew A. Bogan* and Kurt S. Thorn

Graduate Group in Biophysics
University of California
San Francisco, CA 94143, USA

**Corresponding author*

Binding of one protein to another is involved in nearly all biological functions, yet the principles governing the interaction of proteins are not fully understood. To analyze the contributions of individual amino acid residues in protein–protein binding we have compiled a database of 2325 alanine mutants for which the change in free energy of binding upon mutation to alanine has been measured (available at http://motorhead.ucsf.edu/~thorn/hotspot). Our analysis shows that at the level of side-chains there is little correlation between buried surface area and free energy of binding. We find that the free energy of binding is not evenly distributed across interfaces; instead, there are hot spots of binding energy made up of a small subset of residues in the dimer interface. These hot spots are enriched in tryptophan, tyrosine and arginine, and are surrounded by energetically less important residues that most likely serve to occlude bulk solvent from the hot spot. Occlusion of solvent is found to be a necessary condition for highly energetic interactions.

Keywords: protein–protein interaction; molecular recognition; alanine scanning mutagenesis; hot spots; drug design

The vast majority of proteins bind to other proteins at some time in their existence in order to perform various functions. Processes as varied as cytoskeletal remodeling, vesicle transport and signal transduction are all dependent on physical interactions between proteins, yet the principles that govern the interaction of two protein surfaces remain elusive (Janin, 1995). With the recent explosion in co-crystal structures of protein heterodimers that are relevant therapeutic targets, the opportunity exists to design small molecules with high specificity and affinity that block or alter protein–protein interactions. However, this task has been extraordinarily difficult, with the few limited successes based primarily on luck and accident, not on rational design (Tilley *et al.*, 1997). Intelligent design of compounds that are capable of interfering with protein–protein interactions requires an understanding of the sources of affinity and specificity in each interface.

Most interfaces are composed of two relatively large protein surfaces with good shape and electrostatic complementarity for one another (Janin, 1995, 1997; Jones & Thornton, 1996; Lawrence & Colman, 1993; McCoy *et al.*, 1997). Almost all of these interfaces bury more than 600 Å^2 of total surface area and it is often assumed that the energy of protein–protein binding is directly related to buried hydrophobic surface area (Chothia & Janin, 1975; Horton & Lewis, 1992; Jones & Thornton, 1996; Privalov, 1979). While this assumption holds up when whole surfaces are studied, it gives little insight into the details of the contributions of individual residues to binding. Previously, these contributions have been studied using computational techniques (Krystek *et al.*, 1993; Novotny *et al.*, 1989). To experimentally probe the energetic contributions of individual side-chains to protein binding, alanine scanning mutagenesis (Wells, 1991) has been used to selectively remove individual side-chains from an interface. By combining alanine scanning with kinetic and thermodynamic measurements, it has been shown that, despite the large size of binding interfaces, single residues can contribute a large fraction of the binding free energy in an interface (Clackson & Wells, 1995). However, the detailed anatomy of these hot spots remains unclear.

In order to study protein–protein binding at the resolution of individual amino acid residues in an interface (rather than making generalizations about whole interfaces), we have compiled and analyzed a database of alanine mutants of heterodimeric protein–protein complexes for which affinity measurements and structural data are available. We have used this database to study the energetic contributions of individual amino acid side-chains to protein–protein binding, and to determine both the composition and common features of heterodi-

0022–2836/98/260001–09 $30.00/0

mer interfaces. At the resolution of individual side-chains, we do not find a correlation between buried hydrophobic surface area and energies, contrary to the results for whole surfaces (Chothia & Janin, 1975; Horton & Lewis, 1992; Jones & Thornton, 1996). Instead, we find a highly uneven distribution of energetic contributions of individual residues across each interface, with certain residues responsible for the bulk of the binding energy. This is generally true for all of the protein heterodimer interfaces we have studied. We also find that hot spots of binding energy are surrounded by a shell of energetically unimportant contacts in all of the different protein interfaces we analyzed.

The interfaces included in our database are shown in Table 1. The typical error of the measurements of change in free energy of binding ($\Delta\Delta G$) upon mutation to alanine is $\sim \pm 0.5$ kcal/mol. Mutants that result in non-specific destabilization of the complex (by unfolding of the monomer, for instance) have been successfully excluded, as evidenced by the relatively few mutants with significant energy contributions that are not at the interface. The distribution of amino acid types in the database (Table 2) is in good agreement with the distribution of amino acid types on protein surfaces (Janin *et al.*, 1988). For residues that are in dimer interfaces, our database samples evenly

Table 1. Alanine-scanning mutational analyses of interfaces included in our database

Interface	Ala scan references	PDB code of structure
A. *Dimer structures*		
Angiogenin–RNase inhibitor	Chen & Shapiro (1997) Shapiro & Vallee (1992) Shapiro & Vallee (1989)	1ang (angiogenin monomer) 1dfj (RNase monomer) Papageorgiou *et al.* (1997)
Barnase–barstar	Schreiber & Fersht (1993) Schreiber & Fersht (1995)	1brs
BPTI–trypsin	Castro & Anderson (1996)	2ptc
BPTI–chymotrypsin	Castro & Anderson (1996)	1cbw
D1.3–E5.2	Dall'Acqua *et al.* (1996) Goldman *et al.* (1997)	1dvf
D1.3–hen egg lysozyme	Dall'Acqua *et al.* (1996) Hawkins *et al.* (1993)	1vfb
Hen egg lysozyme–HyHEL-10	Kam-Morgan *et al.* (1993) Lavoie *et al.* (1990) Tsumoto *et al.* (1995)	3hfm
hGH–hGHbp	Cunningham *et al.* (1993) Cunningham & Wells (1989) Bass *et al.* (1991) Clackson & Wells (1995)	3hhr
hGH–hPRLbp	Cunningham & Wells (1991)	3hhr (hGH monomer) Somers *et al.* (1994)
Protein A–IgG1	Cedergren *et al.* (1993)	1fc2
RNase inhibitor–RNase A	Chen & Shapiro (1997)	1dfj
Tissue factor–factor VIIa	Gibbs *et al.* (1994) Ruf *et al.* (1995) Ruf *et al.* (1994) Kelley *et al.* (1995) Dickinson *et al.* (1996)	1dan
B. *Monomer structures*		
Ab4D5-5–p185^{HER2}	Kelley & O'Connell (1993)	1fvc
bFGF–FGFR1b	Zhu *et al.* (1995) Springer *et al.* (1994)	4fgf
CD4–gp120	Ashkenazi *et al.* (1990)	3cd4
hG-CSF–hG-CSFbp	Young *et al.* (1997) Reidhaar-Olson *et al.* (1996)	1rhg
hGH–21 Mabs	Jin *et al.* (1992)	3hhr
hIL2–hIL2Rαβ	Sauvé *et al.* (1991)	3ink
IL4–IL4bp	Wang *et al.* (1997)	1rcb
IL8–IL8R	Hebert *et al.* (1991)	3il8
VEGF–KDR	Muller *et al.* (1997)	Muller *et al.* (1997)
VEGF–2 Mabs	Muller *et al.* (1997)	Muller *et al.* (1997)

Those mutations for which K_d or EC_{50} had been measured were used. Mutations that were judged by the authors of the paper to have gross effects on protein conformation were excluded.

Structures for either the protein monomer or the dimer studied were obtained from the PDB (Bernstein *et al.*, 1977) and solvent-accessible surface areas were calculated for the mutated side-chains. The program NACCESS (Hubbard & Thornton, 1993) was used for calculations of solvent-accessible surface areas and all atoms beyond the C^{α} atom were included. Standard van der Waals radii and a probe radius of 1.4 Å were used in the computation. In cases where the dimer structure was available, the solvent-accessible surface area of the side-chains was calculated for the separated monomer and for the complex.

The database contains 2325 residues that have structural data for the monomer or dimer and 402 that have structural data for the dimer. In general, residues with less than 10 Å^2 of solvent-accessible surface area in the monomer were considered to be buried and were not included in the analysis.

Table 2. Amino acid preferences in hot spots

Residue	In database (Number)	(%)	Contribute ⩾ 2 kcal/mol (Number)	(%)	Enrichment in hot spots
Arg	218	9.38	29	13.30	2.47
Asn	99	4.26	5	5.05	0.93
Asp	177	7.61	16	9.04	1.67
Cys	3	0.13	0	0	0
Gln	160	6.88	5	3.13	0.58
Glu	220	9.46	8	3.64	0.68
Gly	28	1.20	1	3.57	0.45
His	50	2.15	4	8.00	1.49
Ile	104	4.47	10	9.62	1.79
Leu	242	10.41	2	0.83	0.01
Lys	143	6.15	9	6.29	1.17
Met	69	2.97	2	2.90	0.54
Phe	166	7.14	5	3.01	0.56
Pro	89	3.83	6	6.74	1.25
Ser	178	7.66	2	1.12	0.21
Thr	131	5.63	2	1.53	0.28
Trp	19	0.82	4	21.05	3.91
Tyr	122	5.25	15	12.30	2.29
Val	107	4.60	0	0	0

The number and percentage of amino acids in the database of 2325 alanine mutations studied are shown. The composition of the database is in rough agreement with the distribution of amino acids on protein surfaces (Janin *et al.*, 1988). The number and percentage of each residue type with $\Delta\Delta G \geqslant 2$ kcal/mol is also given. Percentage of amino acid type with $\Delta\Delta G \geqslant 2$ kcal/mol is calculated by dividing the number of a given residue type with $\Delta\Delta G \geqslant 2$ kcal/mol by the number of that type of residue in the database. Enrichment in hot spots gives the fold enrichment of that residue type in hot spots ($\Delta\Delta G \geqslant 2$ kcal/mol) over the database as a whole (e.g. a value of 2 indicates that the residue is twice as frequent in hot spots as in the database as a whole). The distribution of these energetically important residues is very non-random, with certain amino acid types (Trp, Tyr and Arg) much more likely to be in hot spots than others (Val, Leu and Ser).

across those interfaces. Of the 168 residues in our database that are not in the heterodimer interface, only 16 have a significant $\Delta\Delta G$ (data not shown). The largest of these is 1.22 kcal/mol, a relatively small energetic contribution. Interestingly, nine of these 16 residues are charged amino acids and only 69 of the 152 energetically insignificant ones are charged. While this is not a statistically significant difference, the charged residues may contribute to binding through electrostatic steering (Janin, 1997) even though they are not in the dimer interface.

To determine if the buried surface area of a side-chain in a complex was responsible for its energetic contribution to binding as determined by alanine mutagenesis, we looked at all residues for which the complex structure was available (calculations are described in the legend to Table 1) . We find a poor correlation ($r = 0.44$) between the energetic contributions of individual side-chains in protein–protein binding interfaces and the buried surface areas of those side-chains (Figure 1). Neither the change in total side-chain solvent-accessible surface area on complex formation (ΔASA) nor the side-chain ΔASA of hydrophobic atoms is well correlated to the change in free energy on mutation of that side-chain to alanine ($\Delta\Delta G$).

Hot spots are protected from bulk solvent

Hot spots of binding energy located near the center of the interface are a general property of the interfaces we examined (Figure 2), as was previously shown for human growth hormone bound to its receptor (Clackson & Wells, 1995). The residues that make up each hot spot tend to cluster together near the center of the interface; very few residues that contribute a large amount of binding energy (>3.5 kcal/mol) are at the edge of an interface. There is no purely geometric reason that hot spot residues should be in the center of interfaces; they could just as easily be distributed around the perimeter. The hot spots in Figure 2 are self-complementary across the interfaces; that is, the hot spot of one face packs against the hot spot of its binding partner.

To determine why the hot spot residues we observe tend to be clustered together at the center of an interface, we examined the role that occlusion of bulk solvent from a side-chain plays in determining the energetic contribution of that residue to protein–protein binding. The loss of binding energy of a particular side-chain upon mutation to alanine was plotted against the accessible surface area of that side-chain in the complex (Figure 3). We find that for a residue to have a large impact on the free energy of binding, it must be largely protected from contact with bulk solvent (low or zero *ASA*). However, we also find that many of the residues that are occluded from solvent do not make large contributions to binding energy. Surprisingly, we find no residue with high solvent accessibility that makes a large contribution to

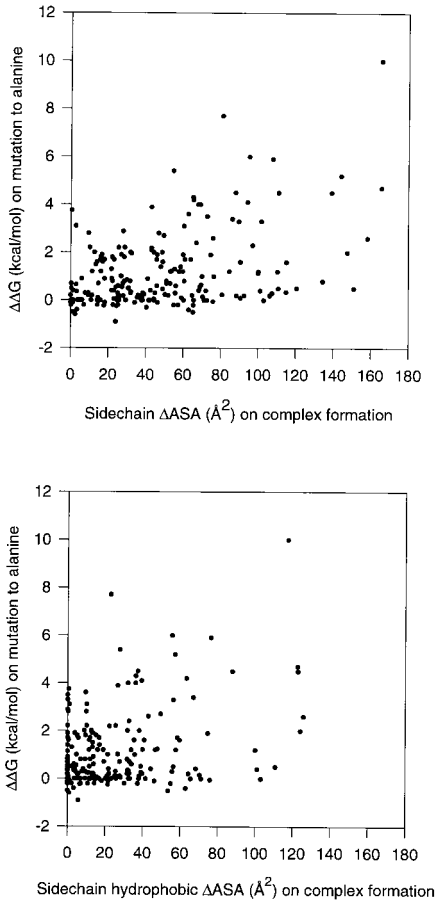

Figure 1. Relationship between change in solvent-accessible surface area and free energies of binding. ΔASA was calculated as the difference between the solvent-accessible surface area of the side-chain in the complex and in the separated monomer. Side-chains that were largely buried in the monomer (<10 Å^2 exposed in the separated monomer) were excluded, as were those for which mutation to alanine was reported to result in perturbation of the protein structure. A poor correlation is observed between change in solvent-accessible surface area and change in free energy upon mutation to alanine. The complexes used are listed in Table 1.

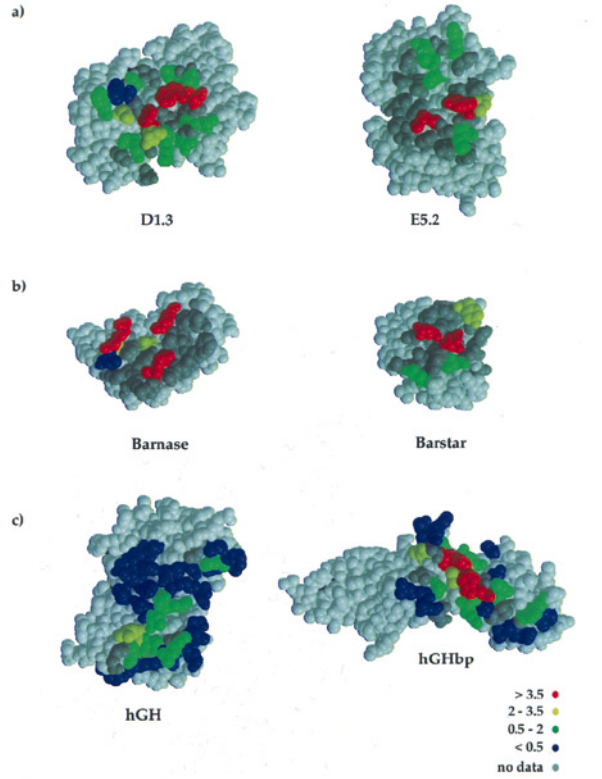

Figure 2. Mapping of $\Delta\Delta G$ of individual residues onto their location in the complexes. The complexes have been opened by 180° rotation around the x axis and water molecules have been omitted. Residues are colored by their energetic contribution to the interface: <0.5 kcal/mol, blue; 0.5 to 2.0 kcal/mol, green; 2.0 to 3.5 kcal/mol, yellow; >3.5 kcal/mol, red; no, data dark gray; not in the interface, white. The energetic contribution of specific residues for (a) the D1.3-E5.2 antibody complex (PBD 1dvf), (b) the barnase-barstar complex (PDB 1brs), and (c) the human growth hormone-receptor site 1 complex (PDB 3hhr). These complexes were selected on the basis of thorough alanine-scanning sampling of the binding interface. The hot spots of binding energy are near the center of the interfaces and are sealed from bulk solvent by surrounding residues of less energetic significance. While a few hot spot residues appear to be at the edge of the interface, they are largely occluded from solvent by residues on the complementary face of their partner. There is no purely geometric reason why the hot spots should be in the center of the interfaces and not at the edges. The Figure was prepared using MOLSCRIPT (Kraulis, 1991) and Raster3D (Bacon & Anderson, 1988; Merritt & Murphy, 1994).

$\Delta\Delta G$. This suggests that, while exposure to bulk solvent prevents a residue from contributing greatly to the overall binding free energy, solvent occlusion is a necessary, but not sufficient, condition for a residue to be in a hot spot. It is not surprising that this condition is insufficient, since other factors, like complementarity of the van der Waals surface in the interface, surely contribute to the free energy of binding.

The vast majority of protein heterodimer interfaces are larger than 600 Å^2 (Jones & Thornton, 1996). We suggest that this cutoff corresponds to a minimum area required to make a water-tight seal around a critical set of energetically favorable interactions. For example, the buried surface area of the human growth hormone receptor hot spot plus the neighboring residues (within 4 Å) is 621 Å^2. This area is remarkably similar to the minimum interface size observed in previous studies (Jones & Thornton, 1996), yet is less than half of the area buried on the receptor in complex with human growth hormone (1300 Å^2).

A necessary condition for high-affinity interactions is the exclusion of bulk solvent from the interacting residues (burial of the interacting residues). In many protein-small molecule interactions, this exclusion of solvent is achieved by burying the small molecule ligand in a deep pocket on the pro-

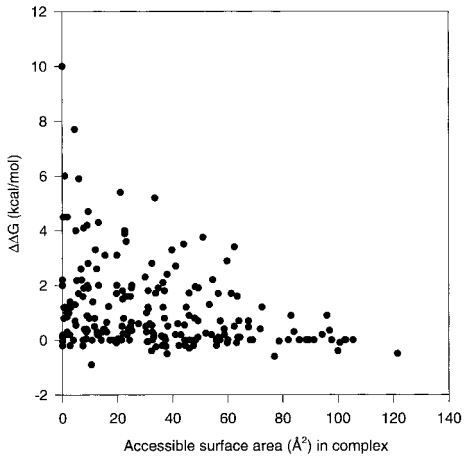

Figure 3. Correlation between solvent accessibility of interface residues in a complex and side-chain contribution to free energy of binding. The same set of residues as that in Figure 1 was used. Solvent accessibility of the interface residues was calculated for the complexed structure. The left side of the graph (0 $Å^2$ solvent accessibility) corresponds to completely buried; the right to highly exposed. Solvent inaccessibility in the complex (left side of the graph) can occur either if the side-chain is largely buried by its neighbors in the monomer or by burial due to the neighboring residues from its binding partner. This shows that a high degree of burial is necessary, but not sufficient, for a large energetic contribution to binding. Some residues with a high degree of solvent accessibility in the complex (>80 $Å^2$, right side of the graph) are still capable of burying a significant amount of surface area in the complex, yet none of these occurs in hot spots.

tein surface. However, since protein–protein interfaces are often flat, deep pockets are not usually available for burying interacting residues. In these interactions exclusion of solvent from hot spot residues is achieved by a surrounding set of contacts that are energetically unimportant (analogous to an O-ring in a pipe fitting). This O-ring is required because it occludes bulk solvent, thereby generating suitable effective dielectric and solvation conditions for a hot spot of interaction energy. For electrostatic and hydrogen bonding interactions, a lower effective dielectric increases the interaction strength. The energetic benefit of occluding bulk solvent from hydrophobic interactions is less clear. In simulations of protein unfolding it has been shown that solvation of hydrophobic residues plays a critical role in the unfolding process (Karplus & Sali, 1995). We believe that a similar mechanism operates in the dissociation of protein–protein interfaces. Occlusion of bulk solvent by the O-ring may retard attack of water molecules on hydrophobic residues in the hot spot and increase interaction affinity by slowing dissociation. As the primary role of the O-ring is to occlude bulk solvent from the hot spot, an alanine mutation in it might have little or no effect on binding affinity as long as bulk solvent was still blocked by the contact of the methyl side-chain and protein backbone.

Alanine-shaving, which is the process of making multiple simultaneous alanine mutations, has been used to experimentally test the role of affinity inert side-chains. One study (Jin & Wells, 1994) found a less than tenfold drop in affinity when 16 energetically unimportant residues were simultaneously mutated to alanine in the interface between human growth hormone and a monoclonal antibody. However, without a structure of the complex it is not known if these mutations disrupt the O-ring. We have calculated the change in solvent-accessible surface area for the hot spot of the human growth hormone (hGH) receptor binding site 1 (PDB 3hhr) using the alanine shaves reported in a recent study (Clackson *et al.*, 1998). We removed atoms beyond the C^{β} atom from the PDB coordinates for each shaved residue and calculated the solvent-accessible surface area of the hGH receptor hot spot (E44, W104, P106, E127 and D164) following the method in Table 1. The largest change in hot spot solvent-accessible surface area we find is 4.2 $Å^2$ for the shaved mutant hGH receptor [B + C + D] (S98A, S102A, E120A, K121A, Q166A, K167A and V171A). Shave [A′] (N72A, T73A, Q74A, E75A, W76A and W80A) increases the solvent accessibility of the hot spot by only 3.3 $Å^2$, whereas shave [E′] (Q216A, R217A and N218A) has no effect on the solvent accessibility of the hot spot. These small changes in solvent accessibility of the hot spot are in agreement with the high affinities for hormone reported for these shaved mutant receptors. Our results show that alanine shaving often fails to significantly increase the solvent accessibility of the hot spot in a complex. This may explain why alanine shaved mutants often have binding affinities similar to those of wild-type proteins.

Much to our surprise, we found that if one assumes no side-chain rearrangement, it is extraordinarily difficult to expose a hot spot to bulk solvent by alanine shaving, even if both sides of the interface are simultaneously shaved. Truncation to alanine of all 46 affinity-inert side-chains ($\Delta\Delta G < 1.5$ kcal/mol) in the hGH and hGH receptor interface resulted in a calculated increase in solvent-accessible surface of only 56% (60.6 $Å^2$) for the hGH receptor hot spot. In other words, alanine side-chains and the protein backbone appear to be largely sufficient to exclude solvent from a neighboring hot spot. It is possible that disrupting an O-ring by prying it open (for example, by tryptophan-scanning mutagenesis) would be easier than cutting a hole in it with alanine shaving.

Added to the difficulty of breaking an O-ring with alanine shaving is the apparent structural plasticity of heavily mutated interfaces. A recent structure (Atwell *et al.*, 1997) of a single hot spot mutant of hGH receptor bound to a high-affinity pentamutant hGH shows massive structural rearrangements in the interface. Main-chain and side-chain atoms that are 15 Å from the epicenter

of these mutations moved by as much as 3 Å relative to their wild-type neighbors (Atwell *et al.*, 1997). However, for our database of alanine mutants, these results are not a problem, since single alanine mutants are unlikely to generate any such plasticity. In fact, the same report has also shown that for a triple alanine mutant of energetically inert residues, there was no major structural change from the wild-type complex.

Perhaps the most compelling data in support of the O-ring hypothesis are those in Figure 3, which show that occlusion of bulk solvent is a necessary condition for residues to be in hot spots. Since all of the hot spot residues are mostly shielded from bulk solvent (>70% burial on average), we suggest that the O-ring serves to establish a non-aqueous environment with an effective dielectric constant at the hot spot significantly decreased from the dielectric of water. The O-ring also shields hot spot residues from attack by bulk solvent, presumably decreasing the dissociation rate. The resulting hot spot environment should therefore favor residues capable of both hydrogen bonding and hydrophobic interactions. This is precisely what we observe in Table 2, with tryptophan, arginine and tyrosine as the preferred residues in hot spots.

Amino acid preferences in hot spots

The distribution of percentages of different amino acid types that occur in hot spots (contribute more than 2 kcal/mol to a binding interaction) in our database is strikingly non-random (Table 2). Only three amino acids appear in hot spots with a frequency of more than 10%; 21% of tryptophan, 13.3% of arginine and 12.3% of tyrosine residues are in hot spots. However, many amino acids are found in hot spots very rarely. Less than 3% of the leucine, methionine, serine, threonine and valine residues in our database are in hot spots. It has been shown that most hydrophobic residues are predominantly found in the interior of proteins, while polar and charged residues are preferred on surfaces (Tsai *et al.*, 1997). Interestingly, tryptophan and tyrosine residues are found on surfaces and in interiors with nearly identical frequencies, the only two hydrophobic residues for which this is true (Tsai *et al.*, 1997). These hydrophobic residues may be found frequently on protein surfaces in order to form the interior cores of certain protein dimer interfaces. Tryptophan, tyrosine and arginine residues are also found more frequently in interfaces than they are in proteins in general; however, this is true of many amino acids (Tsai *et al.*, 1997). An enrichment of tyrosine and tryptophan as well as a discrimination against valine, isoleucine and leucine in antibody complementarity-determining region (CDR) sequences has been noted (Padlan, 1990). This was proposed to be due to aromatic residues contributing binding energy through the hydrophobic effect without a large entropic penalty because they have few rotatable bonds. Another study on several protease/inhibitor complexes (Krystek *et al.*, 1993) found that ''non-entropic'' amino acids (alanine, proline, glycine and cysteine) accounted for a large percentage of residues at the contact surface. Because these residues are not usually mutated in alanine scans, we cannot comment on the energies they contribute to binding.

The non-random composition of hot spots (Table 2) demonstrates that certain amino acids are preferred in the high-energy interactions between proteins in a heterodimer. We do not see a preference for a single type of amino acid, such as hydrophobic or charged residues. In fact, the three most common amino acids in the hot spots analyzed here (Trp, Arg and Tyr) include two reasonably hydrophobic aromatic residues and one positively charged amino acid. One explanation is that amino acids capable of making multiple types of favorable interactions are preferred in the lowered effective dielectric environment of hot spots. Tryptophan, for example, can contribute aromatic π-interactions, a hydrogen bonding donor, and a large hydrophobic surface. Likewise, tyrosine offers hydrophobic surface, and both aromatic π-interactions and the hydrogen bonding ability of its 4-hydroxyl group. Presumably, the ability of tyrosine to hydrogen bond explains why it is more than three times more likely to be in hot spots (on a percentage basis) as phenylalanine (Table 2). This finding excludes the possibility that our results reflect free energy changes due to cavity formation, since tyrosine and phenylalanine have nearly identical volumes. Arginine, too, is capable of multiple types of favorable interaction. It has the ability to form a hydrogen bond network with up to five H-bonds and a salt-bridge with its positively charged guanidinium motif. The electron delocalization of the guanidinium π-system has a pseudo-aromatic character. Also, arginine has three methylene carbon atoms, which are all hydrophobic in character.

It is also interesting to note that aspartate is favored over glutamate, appearing in hot spots more than twice as often (Table 2). Similarly, asparagine is more prevalent in hot spots than glutamine. These differences are presumably due to differences in side-chain conformational entropy (Lee *et al.*, 1994). Curiously, we find that isoleucine, which appears in hot spots with a frequency of 9.62%, is more than ten times as common in hot spots as leucine (0.83%), despite the fact that they are isomers with essentially identical chemistry. Based on these principles of amino acid preferences in hot spots, it may be possible to predict hot spots in an interface without exhaustive mutagenesis and thermodynamic analysis and thus design minimized mimics of these interfaces much more rapidly. An evolutionary trace method based on conservation of important residues has proven useful in finding interaction surfaces common to protein families (Lichtarge *et al.*, 1996). Our approach might allow this method to be extended to proteins where evolutionary data is limited.

Implications for drug design

In contrast to protein–protein targets, the design of small-molecule ligands for various enzymes has been quite successful. Unlike the reasonably large and flat interfaces seen in many protein heterodimers, most enzymes have deep pockets on their surface in which ligands can bind. The pocket is often the enzyme active site, so the rational design of inhibitors is possible without including an O-ring in the designed ligand, because the deep pocket that is provided by the enzyme presumably performs the same function (occlusion of bulk solvent). Many examples exist of successfully designed ligands for enzymes. Agouron Pharmaceutical's nelfinavir (Viracept or AG-1343), for example, is a potent inhibitor of HIV-1 protease that was recently approved for clinical use (Patick *et al.*, 1996). Likewise, the Hoffman-La Roche, Ltd compound Ro 46-6240 was rationally designed as an inhibitor of thrombin and is currently in clinical development as an antithrombotic agent (Hilpert *et al.*, 1994). Numerous other examples of high-affinity, rationally designed small-molecule ligands exist, essentially all of which have enzyme targets (Charifson & Kuntz, 1997). The observation that ligand design has been most successful for proteins with deep pockets supports the concept that occlusion of bulk solvent is a necessary condition for tight binding.

Recently, some peptides and small molecules have been designed to bind to protein interface surfaces. These mimetics include peptide agonists of the erythropoietin (EPO) and thrombopoietin (TPO) receptors that were selected by phage display (Cwirla *et al.*, 1997; Livnah *et al.*, 1996). An analysis of the complex of the EPO receptor with its peptide agonist shows that the features of the peptide-protein interface are generally similar to the protein–protein interfaces we have studied. Of the eight residues conserved in the selected peptides, seven have less than 25 $Å^2$ of surface area accessible to solvent in the complex. The one exception, proline 10, is probably important for stabilizing a β-turn. The EPO receptor and the nonconserved regions of the peptide appear to occlude bulk solvent from the presumed hot spot (the conserved residues) in a manner similar to that of the other interfaces we have studied. Non-peptidic, small-molecule mimics have proven more difficult to design. Recent attempts to design small molecules that block protein–protein interactions have resulted in compounds with micromolar affinities (Judice, 1997; Li *et al.*, 1998; Tilley *et al.*, 1997); we propose that one possible difficulty is the failure to sufficiently exclude solvent from the hot spot mimics.

As drug design moves toward both mimicking and modulating protein–protein interactions, it becomes increasingly important to understand the forces that govern those interactions. Our results indicate that the affinity of a protein interface depends on both an energetically critical hot spot located near the center of the interface, and a surrounding seal of contacting residues that may establish the correct solvation environment by occluding bulk solvent from the hot spot. Much of the solvent occlusion is provided by the main-chain and C^{β} atoms adjacent to the hot spot, which explains why alanine shaving has little effect on affinity. Small molecules designed with a built-in O-ring may be able to modulate or even mimic an entire protein surface in a functionally important protein–protein interaction. Furthermore, the rapid determination of hot spot residues in new structures based on the principles we have outlined may speed development of small-molecule competitive inhibitors of protein–protein interactions.

Acknowledgments

A.A.B. is supported by a National Defense Science and Engineering Graduate Fellowship from the United States Department of Defense. K.S.T. is supported by a Howard Hughes Medical Institute Predoctoral Fellowship. We thank Fred Cohen, Wendell Lim and Jim Wells for helpful comments and discussion. We appreciate the comments of Warren DeLano, Marcin Joachimiak, Dave Miller and Dirk Walther. Also, we thank C. Schutt for pointing out the importance of tryptophan.

References

Ashkenazi, A., Presta, L. G., Marsters, S. A., Camerato, T. R., Rosenthal, K. A., Fendly, B. M. & Capon, D. J. (1990). Mapping the CD4 binding site for human immunodeficiency virus by alanine-scanning mutagenesis. *Proc. Natl Acad. Sci. USA*, **87**, 7150–7154.

Atwell, S., Ultsch, M., De Vos, A. M. & Wells, J. A. (1997). Structural plasticity in a remodeled protein–protein interface. *Science*, **278**, 1125–1128.

Bacon, D. J. & Anderson, W. F. (1988). A fast algorithm for rendering space-filling molecule pictures. *J. Mol. Graph.* **6**, 219–220.

Bass, S. H., Mulkerrin, M. G. & Wells, J. A. (1991). A systematic mutational analysis of hormone-binding determinants in the human growth hormone receptor. *Proc. Natl Acad. Sci. USA*, **88**, 4498–4502.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). The Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535–542.

Castro, M. J. M. & Anderson, S. (1996). Alanine point-mutations in the reactive region of bovine pancreatic trypsin inhibitor: effects on the kinetics and thermodynamics of binding to β-trypsin and α-chymotrypsin. *Biochemistry*, **35**, 11435–11446.

Cedergren, L., Andersson, R., Jansson, B., Uhlen, M. & Nilsson, B. (1993). Mutational analysis of the interaction between staphylococcal protein A and human IgG1. *Protein Eng.* **6**, 441–448.

Charifson, P. S. & Kuntz, I. D. (1997). Recent successes and continuing limitations in computer-aided drug design. In *Practical Application of Computer-Aided*

Drug Design. (Charifson, P. S., ed.), pp. 1–37, Marcel Dekker, Inc., New York.
Chen, C. Z. & Shapiro, R. (1997). Site-specific mutagenesis reveals differences in the structural bases for tight binding of RNase inhibitor to angiogenin and RNase A. *Proc. Natl Acad. Sci. USA,* **94**, 1761–1766.
Chothia, C. & Janin, J. (1975). Principles of protein–protein recognition. *Nature,* **256**, 705–708.
Clackson, T. & Wells, J. A. (1995). A hot spot of binding energy in a hormone-receptor interface. *Science,* **267**, 383–386.
Clackson, T., Ultsch, M. H., Wells, J. A. & de Vos, A. M. (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.
Cunningham, B. C. & Wells, J. A. (1989). High-resolution epitope mapping of hGH-receptor interactions by alanine scanning mutagenesis. *Science,* **244**, 1081–1085.
Cunningham, B. C. & Wells, J. A. (1991). Rational design of receptor-specific variants of human growth hormone. *Proc. Natl Acad. Sci. USA,* **88**, 3407–3411.
Cunningham, B. C. & Wells, J. A. (1993). Comparison of a structural and a functional epitope. *J. Mol. Biol.* **234**, 554–563.
Cwirla, S. E., Balasubramanian, P., Duffin, D. J., Wagstrom, C. R., Gates, C. M., Singer, S. C., Davis, A. M., Tansik, R. L., Mattheakis, L. C., Boyotos, C. M., Schatz, P. J., Baccanari, D. P., Wrighton, N. C., Barrett, R. W. & Dower, W. J. (1997). Peptide agonist of the thrombopoietin receptor as potent as the natural cytokine. *Science,* **276**, 1696–1699.
Dall'Acqua, W., Goldman, E. R., Eisenstein, E. & Mariuzza, R. A. (1996). A mutational analysis of the binding of two different proteins to the same antibody. *Biochemistry,* **35**, 9667–9676.
Dickinson, C. D., Kelly, C. R. & Ruf, W. (1996). Identification of surface residues mediating tissue factor binding and catalytic function of the serine protease factor VIIa. *Proc. Natl Acad. Sci. USA,* **93**, 14379–14383.
Gibbs, C. S., McCurday, S. N., Leung, L. K. & Paborsky, L. R. (1994). Identification of the factor VIIa binding site of tissue factor by homologous loop swap and alanine scanning mutatgenesis. *Biochemistry,* **33**, 14003–14010.
Goldman, E. R., Dall'Acqua, W., Braden, B. C. & Mariuzza, R. A. (1997). Analysis of binding interactions in an idiotope-antiidiotope protein–protein complex by double mutant cycles. *Biochemistry,* **36**, 49–56.
Hawkins, R. E., Russell, S. J., Baier, M. & Winter, G. (1993). The contribution of contact and non-contact residues of antibody in the affinity of binding to antigen. *J. Mol. Biol.* **234**, 958–964.
Hebert, C. A., Vitangcol, R. V. & Baker, J. B. (1991). Scanning mutagenesis of interleukin-8 identifies a cluster of residues required for receptor binding. *J. Biol. Chem.* **266**, 18989–18994.
Hilpert, K., Ackermann, J., Banner, D. W., Gast, A., Gubernator, K., Hadvary, P., Labler, L., Muller, K., Schmid, G., Tschopp, T. B. & van de Waterbeemd, H. (1994). Design and synthesis of potent and highly selective thrombin inhibitors. *J. Med. Chem.* **37**, 3889–3901.
Horton, N. & Lewis, M. (1992). Calculation of the free energy of association for protein complexes. *Protein Sci.* **1**, 169–181.
Hubbard, S. J. & Thornton, J. M. (1993). *Naccess. Computer Program*, Department of Biochemistry and Molecular Biology, University College, London.
Janin, J. (1995). Elusive Affinities. *Proteins: Struct. Funct. Genet.* **21**, 30–39.
Janin, J. (1997). The kinetics of protein–protein recognition. *Proteins: Struct. Funct. Genet.* **28**, 153–161.
Janin, J., Miller, S. & Chothia, C. (1988). Surface, subunit interfaces and interior of oligomeric proteins. *J. Mol. Biol.* **204**, 155–164.
Jin, L. & Wells, J. A. (1994). Dissecting the energetics of an antibody-antigen interface by alanine shaving and molecular grafting. *Protein Sci.* **3**, 2351–2357.
Jin, L., Fendly, B. M. & Wells, J. A. (1992). High resolution functional analysis of antibody-antigen interactions. *J. Mol. Biol.* **226**, 851–865.
Jones, S. & Thornton, J. M. (1996). Principles of protein–protein interactions. *Proc. Natl Acad. Sci. USA,* **93**, 13–20.
Judice, K. (1997). Small molecule inhibitors of human growth hormone/receptor binding. *FASEB J.* **11**, A839.
Kam-Morgan, L. N. W., Smith-Gill, S. J., Taylor, M. G., Zhang, L., Wilson, A. C. & Kirsch, J. F. (1993). High-resolution mapping of the HyHEL-10 epitope of chicken lysozyme by site-directed mutagenesis. *Proc. Natl Acad. Sci. USA,* **90**, 3958–3962.
Karplus, M. & Sali, A. (1995). Theoretical studies of protein folding and unfolding. *Curr. Opin. Struct. Biol.* **5**, 58–73.
Kelley, R. F. & O'Connell, M. P. (1993). Thermodynamic analysis of an antibody functional epitope. *Biochemistry,* **32**, 6828–6835.
Kelley, R. F., Costas, K. E., O'Connell, M. P. & Lazarus, R. A. (1995). Analysis of the factor VIIa binding site on human tissue factor: effects of tissue factor mutations on the kinetics and thermodynamics of binding. *Biochemistry,* **34**, 10383–10392.
Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
Krystek, S., Stouch, T. & Novotny, J. (1993). Affinity and specificity of serine endopeptidase-protein inhibitor interactions. *J. Mol. Biol.* **234**, 661–679.
Lavoie, T. B., Kam-Morgan, L. N. W., Mallett, C. P., Schilling, J. W., Prager, E. M., Wilson, A. C. & Smith-Gill, S. J. (1990). Analysis of antibody-protein interactions using site-directed mutagenesis and a new evolutionary variant of lysozyme. In *Use of X-ray Crystallography in the Design of Antiviral Agents* (Laver, W. G. & Air, G. M., eds), pp. 213–232, Academic Press, San Diego.
Lawrence, M. C. & Colman, P. M. (1993). Shape complementarity at protein/protein interfaces. *J. Mol. Biol.* **234**, 946–950.
Lee, K. H., Xie, D., Freire, E. & 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.
Li, R., Dowd, V., Stewart, D. J., Burton, S. J. & Lowe, C. R. (1998). Design, synthesis, and application of a Protein A mimetic. *Nature Biotech.* **16**, 190–195.
Lichtarge, O., Bourne, H. R. & Cohen, F. E. (1996). An evolutionary trace method defines binding surfaces common to protein families. *J. Mol. Biol.* **257**, 342–358.
Livnah, O., Stura, E. A., Johnson, D. L., Middleton, S. A., Mulcahy, L. S., Wrighton, N. C., Dower, W. J.,

Joliffe, L. K. & Wilson, I. A. (1996). Functional mimicry of a protein hormone by a peptide agonist: the EPO receptor complex at 2.8 Å. *Science,* **273**, 464–471.

McCoy, A. J., Epa, V. A. & Colman, P. M. (1997). Electrostatic complementarity at protein/protein interfaces. *J. Mol. Biol.* **268**, 570–584.

Merritt, E. A. & Murphy, M. E. P. (1994). Raster3D version 2.0 – a program for photorealistic molecular graphics. *Acta Crystallog.* **50**, 869–873.

Muller, Y. A., Li, B., Christinger, H. W., Wells, J. A., Cunningham, B. C. & de Vos, A. M. (1997). Vascular endothelial growth factor: crystal structure and functional mapping of the kinase domain receptor binding site. *Proc. Natl Acad. Sci. USA,* **94**, 7192–7197.

Novotny, J., Bruccoleri, R. E. & Saul, F. A. (1989). On the attribution of binding energy in antigen-antibody complexes McPC 603, D1.3, and HyHEL-5. *Biochemistry*, **28**, 4735–4749.

Padlan, E. A. (1990). On the nature of antibody combining sites: unusual structural features that may confer on these sites an enhanced capacity for binding ligands. *Proteins: Struct. Funct. Genet.* **7**, 112–124.

Papageorgiou, A. C., Shapiro, R. & Acharya, K. R. (1997). Molecular recognition of human angiogenin by placental ribonuclease inibitor – an X-ray crystallographic study at 2.0 Å resolution. *EMBO J.* **16**, 5162–5177.

Patick, A. K., Mo, H., Markowitz, M., Appelt, K., Wu, B., Musick, L., Kalish, V., Kaldor, S., Reich, S. & Ho, D. (1996). Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrobr. Agents Chemother.* **40**, 292–297.

Privalov, P. L. (1979). Stability of proteins: small globular proteins. *Advan. Protein Chem.* **33**, 167–241.

Reidhaar-Olson, J. F., DeSouza-Hart, J. A. & Selick, H. E. (1996). Identification of residues critical to the activity of human granulocyte colony-stimulating factor. *Biochemistry,* **35**, 9034–9041.

Ruf, W., Schullek, J. R., Stone, M. J. & Edgington, T. S. (1994). Mutational mapping of function residues in tissue factor: identification of factor VII recognition determinants in both structural modules of the predicted cytokine receptor homology domain. *Biochemistry,* **33**, 1565–1572.

Ruf, W., Kelley, C. R., Sculleck, J. R., Martin, D. M. A., Polikarpov, I., Boys, C. W. G., Tuddenham, E. G. D. & Edgington, T. S. (1995). Energetic contributions and topographical organization of ligand binding residues of tissue factor. *Biochemistry,* **34**, 6310–6315.

Sauvé, K., Nachman, M., Spence, C., Bailon, P., Campbel, E., Tsien, W.-H., Kondas, J. A., Hakimi, J. & Ju, G. (1991). Localization in human interleukin 2 of the binding site to the alpha chain (p55) of the interleukin 2 receptor. *Proc. Natl Acad. Sci. USA,* **88**, 4636–4640.

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

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

Shapiro, R. & Vallee, B. L. (1989). Site-directed mutagenesis of histidine-13 and histidine-114 of human angiogenin. Alanine derivatives inhibit angiogenin-induced angiogenesis. *Biochemistry,* **28**, 7401–7408.

Shapiro, R. & Vallee, B. L. (1992). Identification of functional arginines in human angiogenin by site-directed mutagenesis. *Biochemistry,* **31**, 12477–12485.

Somers, W., Ultsch, M., DeVos, A. M. & Kossiakoff, A. A. (1994). The X-ray structure of a growth hormone-prolactin receptor complex. *Nature,* **372**, 478–481.

Springer, B. A., Pantoliano, M. W., Barbera, F. A., Gunyuzlu, P. L., Thompson, L. D., Herblin, W. F., Rosenfeld, S. A. & Book, G. W. (1994). Identification and concerted function of two receptor binding surfaces on basic fibroblast growth factor required for mitogenesis. *J. Biol. Chem.* **269**, 26879–26884.

Tilley, J. W., Chen, L., Fry, D. C., Emerson, S. D., Powers, G. D., Biondi, D., Varnell, T., Trilles, R., Guthrie, R., Mennoma, F., Kaplan, G., LeMahieu, R. A., Carson, M., Han, R.-J., Liu, C.-M., Palermo, R. & Ju, G. (1997). Identification of a small molecule inhibitor of the IL-2/IL-2Rα receptor interaction which binds to IL-2. *J. Am. Chem. Soc.* **119**, 7589–7590.

Tsai, C.-J., Lin, S. L., Wolfson, H. J. & Nussinov, R. (1997). Studies of protein–protein interfaces: a statistical analysis of the hydrophobic effect. *Protein Sci.* **6**, 53–64.

Tsumoto, K., Ogasahara, K., Ueda, Y., Watanabe, K., Yutani, K. & Kumagai, I. (1995). Role of Tyr residues in the contact region of antilysozyme monoclonal antibody HyHEL10 for antigen binding. *J. Biol. Chem.* **270**, 18551–18557.

Wang, Y., Shen, B.-J. & Sebald, W. (1997). A mixed charge pair in human interleukin 4 dominates high-affinity interaction with the receptor alpha chain. *Proc. Natl Acad. Sci. USA,* **94**, 1657–1662.

Wells, J. A. (1991). Systematic mutational analyses of protein–protein interfaces. *Methods Enzymol.* **202**, 390–411.

Young, D. C., Zhan, H., Cheng, Q.-L., Hou, J. & Matthews, D. J. (1997). Characterization of the receptor binding determinants of granulocyte colony stimulating factor. *Protein Sci.* **6**, 1228–1236.

Zhu, H., Ramnarayan, K., Anchin, J., Miao, W. Y., Sereno, A., Millman, L., Zheng, J., Balaji, V. N. & Wolff, M. E. (1995). Glu-96 of basic fibroblast growth factor is essential for high affinity receptor binding. *J. Biol. Chem.* **270**, 21869–21874.

Edited by J. Wells

(Received 12 January 1998; received in revised form 1 April 1998; accepted 3 April 1998)