ELSEVIER

The molecular architecture of protein–protein binding sites

Dana Reichmann, Ofer Rahat, Mati Cohen, Hani Neuvirth and Gideon Schreiber

The formation of specific protein interactions plays a crucial role in most, if not all, biological processes, including signal transduction, cell regulation, the immune response and others. Recent advances in our understanding of the molecular architecture of protein–protein binding sites, which facilitates such diversity in binding affinity and specificity, are enabling us to address key questions. What is the amino acid composition of binding sites? What are interface hotspots? How are binding sites organized? What are the differences between tight and weak interacting complexes? How does water contribute to binding? Can the knowledge gained be translated into protein design? And does a universal code for binding exist, or is it the architecture and chemistry of the interface that enable diverse but specific binding solutions?

Addresses
Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel

Corresponding author: Schreiber, Gideon (bcges@weizmann.ac.il)

Current Opinion in Structural Biology 2007, **17**:67–76

This review comes from a themed issue on
Folding and binding
Edited by Valerie Daggett and Alan R Fersht

Available online 18th January 2007

0959-440X/$ – see front matter

DOI 10.1016/j.sbi.2007.01.004

Introduction

Proteins are polymers comprising 20 chemically and structurally different building blocks (amino acids) that fold into a highly specific tertiary structure. This complexity enables proteins to interact with almost any other type of molecule, from small organic compounds, inorganic salts and metals, to sugars, fatty acids, nucleotides, peptides and other proteins. In this review, we focus our attention specifically on interactions between proteins.

The classification of protein–protein interactions has received much attention in recent years, because these interactions are central to the emerging field of systems biology, in which various networks of associations are mapped and their physiological role is scrutinized [1]. Proteins interact with other proteins with varying affinities, ranging from as low as millimolar to as high as femtomolar. Despite this large difference in affinity, all protein–protein interactions maintain a high degree of specificity for their partners. It is now clear that many proteins interact with multiple partners, either simultaneously or separately, depending on availability and environment. The ability to bind multiple partners forms the basis of network complexity, which in turn is related to the extent of evolution of a particular organism.

In this review, we highlight recent advances in our understanding of the molecular architecture of protein–protein binding sites, which facilitates such diversity in binding in response to biological requirements. The complexity of the quest to find a universal code for binding is emphasized throughout the review.

Potential binding sites are imprinted in the unbound protein

Is the architecture of a binding site already printed in the unbound state of a protein? Several bioinformatics studies have been initiated to answer this question, with the goal of identifying the location of potential binding sites on the unbound structure of the protein. The feasibility of such a quest will, in itself, demonstrate that binding is a characteristic of the monomer, similar to active sites in enzymes, and thus that some locations on the protein surface can be designated as active ('warm') sites for interactions.

For interactions between small molecules and proteins, it has been found that the small molecules bind most readily to the deepest clefts on the surface [2]. For protein–protein interaction sites, however, the situation is more complicated because the surface area involved is rather large (700–1500 $Å^2$ per protein) and the binding surfaces are relatively flat [3,4]. Nevertheless, a success rate of ~70% correct prediction has been independently achieved by several different groups [5–8]. These predictions have been achieved by analyzing a wide range of parameters, including solvation potential, amino acid composition, conservation, electrostatics and hydrophobicity. No parameters have high predictive power on their own, but in combination they seem to be relatively successful. Protein-binding interfaces are clearly not homogenous. Some interface residues contribute much to binding (called 'hotspots'), whereas many others make only a marginal contribution [9]. Keskin *et al.* [10•] have suggested that residues in hotspots are often pre-organized in the unbound protein state, strengthening the view that much of the surface does not accommodate binding and that binding sites are predefined. Accordingly, we can designate

some areas on the protein surface as potential binding sites and the rest of the surface as 'cold' with respect to binding.

Specificity of binding

Dictated by their diverse functional requirements, proteins bind to each other with a large range of affinities and rate constants. Nevertheless, by the definition of an interaction, some degree of specificity is maintained in all cases. Perhaps the best-identified family of protein–protein interactions in terms of their specificity and diversity are antibodies. In this family, the same framework (albeit different sequence and length) of six complementarity-determining regions (CDRs) binds specifically to most other proteins, as demonstrated many times by biologists who use antibodies to detect their proteins of interest by western blotting on whole-cell protein extracts. Thus, binding is not a case-specific feature of the protein architecture, but is related to the shape and chemistry of the binding surface.

Specificity is not less important than affinity and is present even in weak binding protein complexes. But specificity is fragile and it can be altered even by introducing a single point mutation. In the EphB4 receptor, for example, Leu95 (as indicated by its substitution with another amino acid) plays a particularly important part in defining the structural features that confer the ligand selectivity of EphB4 [11].

The β-lactamase inhibitor protein (BLIP) has the ability to bind a wide range of β-lactamase proteins from different species at the same site, but with affinities ranging from micromolar to nanomolar [12]. Alanine scanning analysis of residues at the BLIP interface binding to four different β-lactamase proteins has shown that some hotspot locations are shared by all β-lactamase proteins, whereas others are specific to a particular one [13]. This variation in hotspots is not related to structural differences in the complex, but to differences in amino acid composition among the β-lactamase proteins. Accordingly, it is not surprising that a seemingly neutral Asp104Glu mutation in SHV β-lactamase confers an increase in binding affinity for BLIP from the micromolar to nanomolar range, similar to the binding affinity of TEM1 β-lactamase for BLIP [14•].

Partial sharing of hotspots has been also found for placental lactogen binding to the extracellular domain of the human prolactin receptor [15], and for the binding of interferon-α2 (IFN-α2) versus IFN-β to the interferon receptor IFNA-R2 [16,17].

Knowledge of specificity-determining sites is important in drug design because it enables the selection of one activity over another through the same binding partners. For example, mutagenesis and structural analyses of the complex between bone morphogenetic protein-2 (BMP-2) and its receptors have shown that Leu51 and Asp53 of BMP-2 represent a hotspot of binding to the type I receptor. A variant of BMP-2, Leu51Pro, is deficient in type I receptor binding, although its overall structure and its binding to type II receptors and inhibitor proteins, such as noggin, are unchanged. Thus, the Leu51Pro substitution converts BMP-2 into a receptor-inactive inhibitor of noggin [18].

A classical example of proteins with high discriminative specificities is the endonuclease (DNase) domains of colicins, which bind to their specific immunity proteins (Im2, Im7, Im8 and Im9). Comparative alanine scanning has revealed significant differences in the amino acid composition of the specificity-determining site — differences that contribute to cognate binding with variations in affinity of up to ten orders of magnitude [19,20].

In the cytokine superfamily, the so-called 'redundant function' describes the sharing of one or more receptor subunits, even though sequence similarity is frequently below 25%. For example, the ligands interleukin-4 (IL4) and IL13 both bind to the IL13-Rα1 receptor subunit, but each bind to a different second subunit, namely, IL13-Rγc and IL13-Rα2, respectively. Nevertheless, their mode of binding to the shared IL13-Rα1 receptor is very different. Kraich *et al.* [21•] have suggested that the modular architecture of this binding site permits a mechanism by which proteins generate binding affinity and specificity independently within an interface formed by several interaction clusters, facilitating a broad range of affinities by selecting the clusters used for binding. A similar mechanism has also been proposed by these authors for the interaction between BMP-2 and its type I receptor [18].

A more subtle way to dictate specificity is used by the Src homology 3 (SH3) domain to interact with type I and type II polyproline ligands. The structural basis for type I and II binding specificity of SH3 domains has been found to relate to a conserved tryptophan residue in the SH3 binding pocket that adopts two different orientations. These orientations determine the type of ligand (I or II) that is able to bind to the domain [22]. Thus, a conformational change has profound effects on protein–protein interactions, highlighting the importance of structural details for the prediction of protein–protein interactions.

Amino acid composition of binding sites

How is a high degree of specificity and affinity achieved in protein–protein interactions? In one study, Sidhu and coworkers [23•] assessed whether the use of all 20 natural amino acids is important to obtain tight and specific binding. They derived an antigen-binding fragment, Fab-YADS2, that recognizes vascular endothelial growth factor (VEGF) from a library in which the diversity is

restricted to four amino acids (tyrosine, serine, alanine and aspartate). The structure of the Fab–antigen complex revealed that the structural paratope is dominated by sidechains of tyrosines, which account for 11 of the 15 functionally important residues. Isothermal titration calorimetry and cell-based assays showed that restricted chemical diversity does not limit the affinity or specificity of Fab-YADS2 relative to natural antibodies [23•]. The abundance of tyrosine in the interface is not necessarily surprising, because tyrosine has been found to be the most common amino acid in binding sites in general [4,5].

In a second study, Sidhu and co-workers [24••] went on to assess the extent of exchangeability of amino acids at the binding site. This is a rather unusual approach, because mutational analysis is most often restricted to alanine substitution, which does not provide a comprehensive view of the allowed amino acid space at any specific position. Using the complex between human growth hormone (hGH) and its receptor (hGHR) as their experimental platform, they introduced any one of the 20 natural amino acids at all 35 interface positions by implementing the shotgun approach, in which binding affinity is related to the abundance of specific sequences after limited rounds of panning. The reliability of this technique has been verified by *in vitro* affinity measurements of selected mutants [25]. The results of this study are rather surprising. First, Sidhu and co-workers [24••] verified that, from a structural point of view, the interface was highly adaptable to mutations. Functionally (with function determined as binding), the interface was also adaptable to mutations, but counter-intuitively the tolerated mutations were neither chemically nor evolutionarily conserved. In fact, neither chemical nor evolutionary conservation was a good indicator of allowed mutations, which seemed to be very context dependent. Whereas some of the alanine scanning hotspot positions showed high specificity against substitution, others did not, and some highly specific positions were not hotspots at all. Negative specificity was rare, however, arguing against the hypothesis that negative specificity is required for species specificity.

Hotspots within the protein–protein binding site

Clackson and Wells [9] introduced the concept of binding hotspots with their basic observation that most of the binding energy is contributed by a only few residues, which in turn are surrounded by supporting interactions of lesser importance. This work has been extended to other systems by Bogan and Thorn [26], who, using a bioinformatics approach, found that hotspots are surrounded by energetically less important residues that most probably serve to occlude bulk solvent from the hotspots. Occlusion of solvent has been found to be a necessary condition for highly energetic interactions; however, not all binding interfaces seem to have hotspots. In a recent study, Roisman *et al.* [27] dissected, using alanine mutagenesis, the binding site of IFN-α*2* with its receptor IFNA-R1 (dissociation constant, K_d = 1.5 μM). None of the ~30 mutations probed reduced binding by more than fivefold. Interestingly, three of the nine alanine mutations mapped to the binding site caused an increase in binding affinity. Overlaying the 13 different IFN-α*2* sequences showed that these three residues (histidine, glutamate, glutamine) are conserved throughout this family, which suggests that weak binding of IFN-α*2* to IFNA-R1 is functionally important. The only tight binding interferon is IFN-β, which is known to posses unique cellular activities [27,28•].

An absence of hotspots has also been found in the binding of protein L to the immunoglobulin-κ light chain. Svensson *et al.* [29] suggest that the low abundance of hotspots within this nanomolar affinity binding complex is due to the large number of mainchain interactions present between the proteins. This explanation might also be applicable to interferon, because mutational analysis of IFNA-R1 did identify hotspots of binding to interferon.

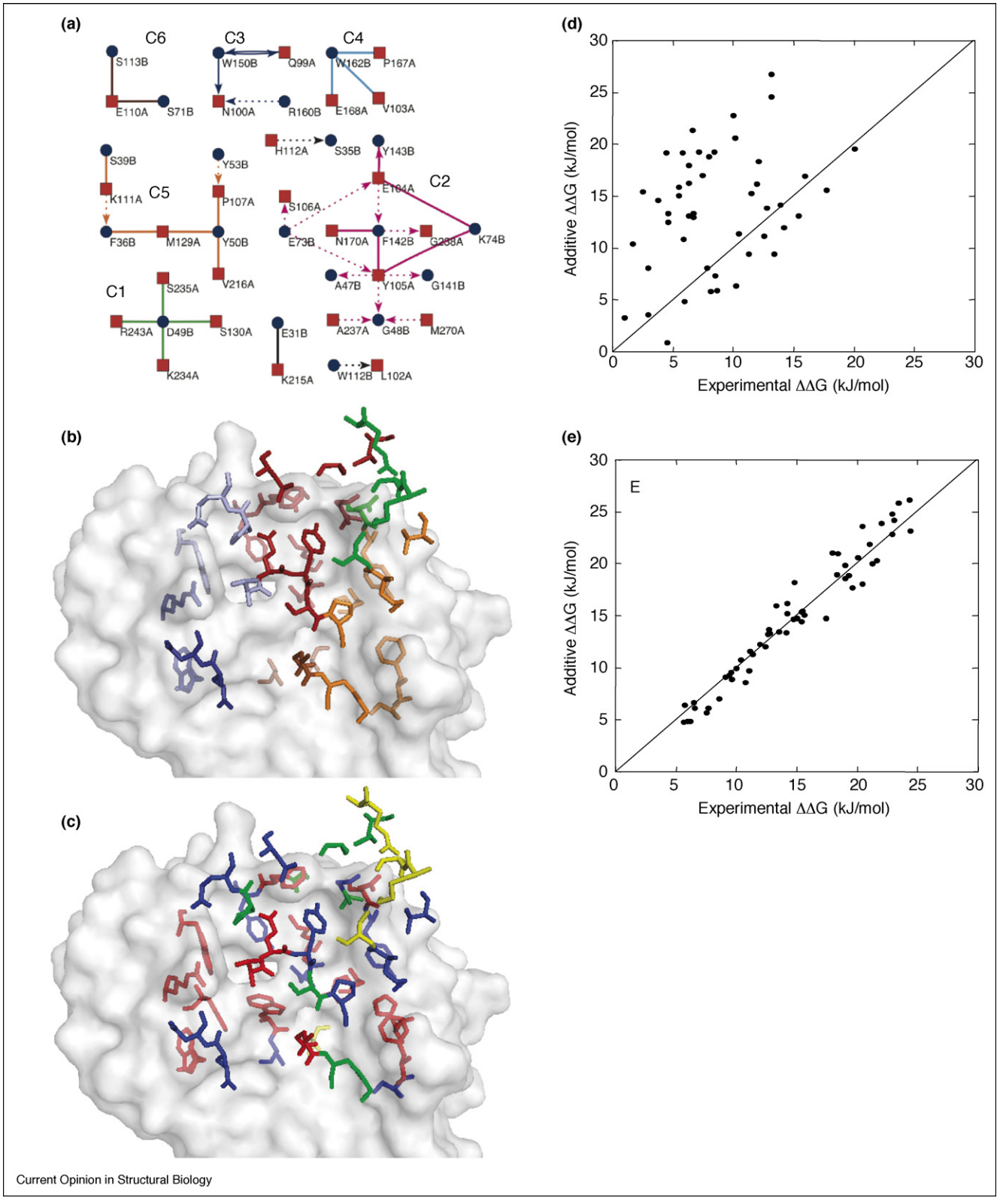
In analyzing the composition and nature of binding hotspots, one should be cautious of indirect effects of mutations, for example, structural perturbations in the monomer that might affect binding. This is not an easy task. For the binding of monoclonal antibody mAb4E11 to gpE of the DEN1 virus, Bedouelle *et al.* [30] compared the change in free energy of binding upon mutation with the probabilities of the residues in the CDRs to form topological contacts with an antigen — contacts that have been intensively studied in past years [31,32]. This study was done to investigate whether large changes in binding affinity are due to direct effects on the interface or might be attributed to conformational changes in the antibody [30]. In most cases, however, this issue remains unresolved, and can lead to considerable inconsistencies between computational and experimental results.

Binding sites can be viewed as an assembly of individual modules

Single-point mutation analysis and its potential indirect effects on binding, such as structural perturbations, raise the following question: is a protein–protein binding site simply a conglomerate of multiple interacting residues, whereby the affinity is dictated by the number of optimized pairwise interactions, or are the interactions highly cooperative?

To answer this question, we have constructed contact maps of the protein–protein interface in the complex between TEM1 and BLIP, and compared them to their respective structural maps (Figure 1a–c). A contact map is built from the physical interactions between the proteins, such as hydrogen bonds, van der Waals interactions and so on. The map shows that the TEM1–BLIP complex can be divided into six individual clusters: each cluster comprises several

Figure 1



Three different views of the protein–protein interaction between TEM1 and its inhibitor, BLIP. **(a)** Connectivity map of the TEM1–BLIP complex, with TEM1 and BLIP residues being nodes in the graph (red squares and blue circles, respectively). Edges are colored in line with their cluster. Three interaction types are shown in the map: sidechain–sidechain (solid lines), backbone–sidechain (dotted lines with the arrow pointing towards

closely interacting residues, with only few interactions being present between the clusters (Figure 1a) [33••]. The change in binding free energy of mutations located on different clusters is found to be additive, whereas mutations within the same cluster cause complex energetic and structural consequences (Figure 1d,e). As a result, deleting complete clusters from the interface causes structural and energetic consequences that are smaller than expected compared with the additive values of the single mutations constituting the clusters [33••].

Constructing interaction maps of many other complexes has shown that this modular architecture is a general design criterion of binding sites. Moreover, it seems that large, highly evolved clusters can result in tight binding interfaces. By contrast, weak interfaces seem to be characterized by a low-complexity interaction map. This view is supported by bioinformatics analyses of hotspot residues performed by the groups of Nussinov [10•] and Vishveshware [34], who propose that hotspots are located in densely packed 'hot regions'. Within these regions, the hotspots form networks of interactions, contributing cooperatively to the stability of the complex. The contributions of separate, independent hot regions are, however, additive.

Because hotspot residues are also conserved by evolution, proteins that bind multiple partners at the same sites are expected to use all or some combination of these regions [35]. Furthermore, strong and weak interfaces are identified on the basis of the interaction strength between amino acid residues and the sizes of the interface clusters. The interface strengths evaluated on the basis of the interface clusters and hubs also correlate well with experimentally determined dissociation constants of known complexes [34].

The modularity of binding sites suggests a roughness of the energy landscape of the interaction surface. Indeed, this roughness has been measured by single-molecule dynamic force spectroscopy for a complex consisting of the small GTPase Ran protein and the nuclear transport receptor importin-β. These measurements indicate a bumpy energy surface, which is consistent with the ability of importin-β to accommodate multiple conformations and to interact with different, structurally distinct ligands [36•].

A key challenge in studying proteins is their diversity. Whereas a simple modular architecture is found for some interfaces, long-range cooperative effects have been detected for others. Binding of the variable domain of the T-cell receptor to a bacterial superantigen, for example, has been shown to involve cooperative binding energies between distinct hot regions that are separated by more than 20 Å [37••]. The propagation of these cooperative effects through a dynamic structural network is a complex evolutionary phenomenon reminiscent of the allosteric effects in enzymes. By contrast, the simple additivity found between remote hotspots in the TEM1–BLIP complex can be viewed as the typical mechanism. Long-range cooperativity can be predicted by statistical coupling analysis pioneered by the group of Ranganathan [38]. Using this method, these authors have identified a network of energetically coupled residues that link the functional surfaces of nuclear receptor ligand-binding domains, facilitating long-range cooperativity between the residues.

Are multiple mutations located on one side of a protein interface additive? Using the shotgun method to monitor intracooperativity between hGH and its receptor, hGHR, or between the Nogo receptor and Nogo, has demonstrated that binding site residues located on one chain are essentially additive towards other residues on the same chain [39,40]. This arrangement contrasts with the interfaces between TEM1 and BLIP, or between the monoclonal antibody mAb4E11 and gpE of the DEN1 virus, for which cooperativity among residues on the same protein chain has been detected [31,33••].

Enthalpic and entropic effects on binding

Thorough thermodynamic analysis of binding complexes has held much hope for providing a more detailed understanding of the binding process and the reasons underlying the architecture of binding sites. Time after time, however, the results of such studies point to the disappointing conclusion that large enthalpy–entropy compensations dominate binding [41]. Indeed, this has been found for 16 different interactions measured between cognate and non-cognate partners in colicin–immunoprotein complexes [19], Ras–effector interactions [42] and the complex between ankyrin and the transcription factor GABP [43].

It seems that the changes in enthalpies and entropies introduced by mutations are much larger than the overall differences in free energy. As a result, the predicted change in free energy upon mutation can be calculated with higher precision by using software such as FOLD-X [44] than by evaluating the enthalpic and entropic effects of the same mutations.

(Figure 1 Legend Continued) the backbone) and interactions of both sidechain–sidechain and backbone–sidechain between the same pair of residues (solid lines with arrows pointing to the backbone donor atom). **(b)** Structural view of the connectivity map in (a). Color coding as in (a); BLIP is shown as a transparent surface. **(c)** Same view as in (b), but the residues are colored according to their contribution to binding, as measured by mutagenesis of the respective residues to alanine. Red represents hotspots (>2 kcal/mol), orange represents intermediate binders (0.5–2 kcal/mol), blue represents inert mutations and green represents residues for which no data are available. **(d,e)** The degree of additivity of free energy of binding between mutations of TEM1 and BLIP (d) within or (e) between the six clusters shown in (a). Additive ΔΔG is defined as $\Delta\Delta G_{mut1} + \Delta\Delta G_{mut2}$, and is plotted against the experimentally determined values of the double mutant ($\Delta\Delta G_{mut1,mut2}$).

The dynamic nature of protein binding sites

Structural rearrangements of the unbound proteins are often observed to be part of the binding process. Sometimes these changes involve only small movements of loop regions and sidechain rearrangements, but significant backbone movements are observed in some proteins. A computational study of several complexes undergoing structural rearrangement on binding has suggested that the motions calculated for the monomer correlate with the experimentally observed structural changes that occur on binding. This emphasizes the pre-existing equilibrium of different conformations as a selection mechanism for protein–protein interactions [45•,46•].

Greater flexibility of the protein monomer might be related to higher free energy [47], which in turn might be relevant to understanding the binding mechanism of partially unfolded proteins that gain structure during binding — a subject of increasing popularity in recent years. A good example of how the affinity of a complex is increased by increasing the disorder of the unbound state has been demonstrated for a multiple hGH mutant selected by phage display to bind hGHR with a 400-fold higher affinity. This affinity is far higher than that promoting optimal function [48], which suggests that complexation of the extracellular domains is not the only factor dictating signal transduction [49•]. Analysis of the binding mechanism of this multiple mutant shows that much of the increased binding is due to weakening of the intraprotein interactions, particularly those of helix-1 in the unbound state. On complexation, hydrogen-deuterium exchange studies have shown that helix-1 is gaining back the lost stability of the unbound state [50••]. Horn *et al.* [50••] suggest that the increased binding energy of the hGH multiple mutant can be attributed to an increase in free energy of the unbound state and not to a reduced absolute energy of the bound state.

Low-affinity protein complexes

Effectors of Ras and Rap constitute a large family of proteins that have low sequence and structural homology, but bind to the same Ras or Rap epitope. The specific interactions in this family are characterized by strong charge complementarity and low binding affinity, resulting from fast rates of association and dissociation. These fast rate constants are seemingly important for proper signal transduction, because they are conserved throughout the family of Ras and Rap effector interactions [51]. Thus, it might not be surprising that changing the electrostatic surface potential of the effectors does alter the specificity of interactions [51,52].

Ras–effector interactions belong to a group of protein complexes that have evolved to bind weakly. The same phenomenon has been documented for the binding of some cytokines to their receptors [27] and for electron transfer protein complexes [53]. Whereas strong electrostatic complementarity dominates both Ras/Rap–effector and electron transfer protein interactions, the same electrostatic characteristics are also present in some of the tightest known protein–protein interactions, such as those between RNase A and ribonuclease inhibitor, between barnase and barstar, and between immunoproteins and DNase [20,54,55]. Moreover, weak binding is not an intrinsic property of an interaction, but can be changed by mutagenesis.

The best-characterized complexes evolving from weak to tight binding are found in the immune system, either between antibodies and protein antigens, or between T-cell receptors and their ligands. In a recent study of the maturation of the Vβ domain of the murine T-cell receptor binding to the superantigen staphylococcal enterotoxin C3 (SEC3), an increase in affinity of 1500-fold was measured. A detailed structural and energetic analysis of the changes that occur during this process has shown that the devil is in the detail: several small perturbations in structure, hydrogen bonding, buried surface area, shape complementarity and cooperativity each have an effect on binding, resulting in a large change overall [56•]. Essentially, this is what the structure–function relationship is all about: in other words, many exact details come together into one. This possibly represents one of the main difficulties in energy simulations: we lack a detailed, high-resolution understanding of the forces governing interactions, which are further complicated by the non-additivity of many interactions.

Another interesting example of weak binding evolving to tight binding is provided by a comparison of the TEM1 and SHV-1 β-lactamases binding to BLIP. The difference in affinity is 1000-fold; however, a single, seemingly conserved mutation in SHV (Asp104Glu) increases the binding affinity of SHV to almost the value measured for TEM1 [14•]. Similarly, the weak affinity between IFNAR1 and IFN-α2 (K_d = 1.5 μM) can be increased by 50-fold (K_d = 30 nM) through the introduction of three single mutations. These mutations do not change the basic architecture of the binding site, but add other specific interactions [27,28•]. Thus, it seems that, at least in some cases, the difference between tight and weak binding might be quantitative and not qualitative in nature.

The role of interface water in binding

The function of water in protein interfaces has been reviewed recently [57]. Lawrence and Colman [58] previously noted in 1993 that, on average, 30% of an interface is filled with water molecules that bridge the two binding partners. More recently, Rodier *et al.* [59] compared the water content of crystal packing with that of specific protein–protein interactions. They found that interfaces resulting from crystal packing are more hydrated than those in specific homo- and hetero-complexes.

In specific protein–protein interactions, water molecules often form a ring around the center of the interface that remains 'dry'. Interestingly, the hydration of weak protein–protein interactions, such as those between energy transfer partners, is reminiscent of crystal packing [60]. Water molecules at interfaces form hydrogen bonds with protein groups, and have a preference for the mainchain carbonyl and the charged sidechains of glutamate, aspartate and arginine. Water-mediated polar interactions are as abundant at interfaces as direct protein–protein hydrogen bonds and might contribute to the stability of the assembly [59,61]. Hotspot residues tend to be located in dry areas of the interface, suggesting that the contribution of water to binding might be small. However, the HINT forcefield (a non-Newtonian forcefield based on experimentally determined log Poctonal/water values) suggests that interface water molecules make a significant contribution to the stability of a complex [62]. In a very recent study on the energetics of water-mediated hydrogen bonds, determined using double-mutant cycles, we have measured a neutral effect of these interactions on protein–protein binding (G Schreiber *et al*.., unpublished). Thus, the contribution of water to binding remains under debate.

Engineering and design

One of the main goals in studying protein–protein interfaces is to generate methods facilitating the recognition and engineering of protein complexes. Because binding forms a basis for all living processes, the ability to engineer interactions will open up endless possibilities to interfere with these processes, both for the sake of knowledge and to cure diseases. Today, the protein engineer can choose to use either a rational, computer-based design or the selection of good binders from large libraries of different compounds and/or mutants. Although computer-based design is conceptually more satisfying because it tests our understanding of the process, selection methods are the way forward if we want to achieve successful results.

Antibodies are most often used to generate new binding functions; over the years, however, other systems have been developed to serve as scaffolds for the design. This issue has been recently reviewed [63•] and thus is not discussed here. What can be learned from the various selection methods is that, first, it is not that difficult to achieve nanomolar affinity using a good scaffold, a large library and a good selection method; and, second, many of the scaffolds used are proteins that naturally bind to various different target proteins, such as SH2 and SH3 domains, PTZ domains, T-cell receptors, antibodies and some repeat proteins, which have been recently found to be part of an adaptive immune response in vertebrates and plants [63•].

To generate or alter binding sites through rational design is a much more complex task. So far, there is no example of a binding site that has been designed from scratch to achieve reasonable affinity. In several cases, however, tighter binding or new specificities for binding have been successfully achieved [14•,52,64•,65]. Solving the X-ray structures of these designed interfaces teaches us that some design features are implemented as predicted, but others are not [65]. Thus, the main problem in the design of new interfaces might be our inability to produce high-resolution models, particularly if backbone movements are involved. This problem, coupled with the approximated nature of our energy functions in predicting even a relatively simple task (such as mutation to alanine) [44,66,67], suggests that the path to the routine design of new binding interfaces might still be long.

Conclusions

After many years of intense scrutiny of the composition and architecture of protein–protein binding sites, we still do not fully appreciate their complexity and how proteins are able to fulfill the requirements of specificity and affinity of binding. In our view, there is no universal track to achieve this. On the contrary, it seems that different proteins adopt various solutions to this problem. In this sense, the problem is even more complex than protein folding, in which the formation of a hydrophobic core with polar groups ordered on the surface presents the basic constraint governing protein structure.

Much of the complexity of protein–protein interactions stems from the requirement of these proteins to be stable in both their unbound and bound state, to not aggregate and to maintain a high degree of specificity for second-order reacting molecules at low cellular concentrations. To fulfill these requirements, specific polar interactions (hydrogen bonds, salt bridges) have to be formed. These interactions require a high degree of surface complementarily and precision in binding site docking. The complexity of this process is enhanced by the shape of the interacting surfaces, which is at least partly pre-defined in the monomer structures.

On the up-side, protein–protein interactions are more additive in nature than the process of protein folding (where the protein is either folded or not). Proteins can interact weakly (down to millimolar affinities) or tightly (up to femtomolar affinities), and the difference in the range between the two extremes is more quantitative than qualitative. In other words, affinities can be changed (through mutagenesis) by many orders of magnitude without changing the basic structure of the interface. Thus, achieving good binding is a meter of evolution and not revolution. This observation gives hope to the protein engineer, because higher affinity always seems to be achievable.

Acknowledgements

This work was funded by MINERVA (grant 8525) and the Israeli Ministry of Science and Technology (grant 0263).

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Cusick ME, Klitgord N, Vidal M, Hill DE: **Interactome: gateway into systems biology**. *Hum Mol Genet* 2005, **14**:R171-R181.

2. Vajda S, Guarnieri F: **Characterization of protein–ligand interaction sites using experimental and computational methods**. *Curr Opin Drug Discov Devel* 2006, **9**:354-362.

3. Bahadur RP, Chakrabarti P, Rodier F, Janin J: **A dissection of specific and non-specific protein–protein interfaces**. *J Mol Biol* 2004, **336**:943-955.

4. Nooren IM, Thornton JM: **Structural characterisation and functional significance of transient protein–protein interactions**. *J Mol Biol* 2003, **325**:991-1018.

5. Neuvirth H, Raz R, Schreiber G: **ProMate: a structure based prediction program to identify the location of protein–protein binding sites**. *J Mol Biol* 2004, **338**:181-199.

6. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES: **Are protein–protein interfaces more conserved in sequence than the rest of the protein surface?** *Protein Sci* 2004, **13**:190-202.

7. Bordner AJ, Abagyan R: **Statistical analysis and prediction of protein–protein interfaces**. *Proteins* 2005, **60**:353-366.

8. Zhou HX, Shan Y: **Prediction of protein interaction sites from sequence profile and residue neighbor list**. *Proteins* 2001, **44**:336-343.

9. Clackson T, Wells JA: **A hot spot of binding energy in a hormone–receptor interface**. *Science* 1995, **267**:383-386.

10. • Keskin O, Ma B, Rogale K, Gunasekaran K, Nussinov R: **Protein–protein interactions: organization, cooperativity and mapping in a bottom-up systems biology approach**. *Phys Biol* 2005, **2**:S24-S35.

In this paper, the authors show that hotspot clusters are within densely packed hot regions, where they form networks of interactions contributing cooperatively to the stability of the complex. The contributions of separate, independent hot regions are, however, additive.

11. Chrencik JE, Brooun A, Kraus ML, Recht MI, Kolatkar AR, Han GW, Seifert JM, Widmer H, Auer M, Kuhn P: **Structural and biophysical characterization of the EPHB4–EPHRINB2 protein protein interaction and receptor specificity**. *J Biol Chem* 2006, **281**:28182-28185.

12. Strynadka NC, Jensen SE, Alzari PM, James MN: **A potent new mode of β-lactamase inhibition revealed by the 1.7 Å X-ray crystallographic structure of the TEM-1–BLIP complex**. *Nat Struct Biol* 1996, **3**:290-297.

13. Zhang Z, Palzkill T: **Dissecting the protein–protein interface between β-lactamase inhibitory protein and class A β-lactamases**. *J Biol Chem* 2004, **279**:42860-42866.

14. • Reynolds KA, Thomson JM, Corbett KD, Bethel CR, Berger JM, Kirsch JF, Bonomo RA, Handel TM: **Structural and computational characterization of the SHV-1 β-lactamase/β-lactamase inhibitor protein (BLIP) interface**. *J Biol Chem* 2006, **281**:26745-26753.

BLIP binds various β-lactamases with affinities ranging from micromolar to picomolar. Although the TEM1 and SHV-1 β-lactamases are almost structurally identical, BLIP binds TEM1 1000-fold tighter than SHV-1. Structural and computational analysis pinpoints a mutation, from aspartate to glutamate, that increases SHV–BLIP binding from micromolar to nanomolar affinity, which is comparable to that found for the TEM1–BLIP complex. Thus, small, chemically conserved mutations can have a big effect on binding.

15. Walsh ST, Kossiakoff AA: **Crystal structure and site 1 binding energetics of human placental lactogen**. *J Mol Biol* 2006, **358**:773-784.

16. Piehler J, Schreiber G: **Mutational and structural analysis of the binding interface between type I interferons and their receptor ifnar2**. *J Mol Biol* 1999, **294**:223-237.

17. Slutzki M, Jaitin DA, Yehezkel TB, Schreiber G: **Variations in the unstructured C-terminal tail of interferons contribute to differential receptor binding and biological activity**. *J Mol Biol* 2006, **360**:1019-1030.

18. Keller S, Nickel J, Zhang JL, Sebald W, Mueller TD: **Molecular recognition of BMP-2 and BMP receptor IA**. *Nat Struct Mol Biol* 2004, **11**:481-488.

19. Keeble AH, Kirkpatrick N, Shimizu S, Kleanthous C: **Calorimetric dissection of colicin DNase–immunity protein complex specificity**. *Biochemistry* 2006, **45**:3243-3254.

20. Li W, Keeble AH, Giffard C, James R, Moore GR, Kleanthous C: **Highly discriminating protein–protein interaction specificities in the context of a conserved binding energy hotspot**. *J Mol Biol* 2004, **337**:743-759.

21. Kraich M, Klein M, Patino E, Harrer H, Nickel J, Sebald W,
• Mueller TD: **A modular interface of IL-4 allows for scalable affinity without affecting specificity for the IL-4 receptor**. *BMC Biol* 2006, **4**:13.

IL4 and IL13 both bind to the IL-4Rα receptor, albeit with a different organization of the binding site stemming from the modular architecture of the interactions. This observation suggests a possible mechanism by which proteins might be able to generate binding affinity and specificity independently: namely, an interface formed by several interaction clusters facilitates a broad range of affinities by selecting how many of these interaction clusters will contribute to the overall binding free energy.

22. Fernandez-Ballester G, Blanes-Mira C, Serrano L: **The tryptophan switch: changing ligand-binding specificity from type I to type II in SH3 domains**. *J Mol Biol* 2004, **335**:619-629.

23. Fellouse FA, Barthelemy PA, Kelley RF, Sidhu SS: **Tyrosine plays**
• **a dominant functional role in the paratope of a synthetic antibody derived from a four amino acid code**. *J Mol Biol* 2006, **357**:100-114.

From this study, it seems that the diversity allowed by the use of 20 amino acids is more than required to form a specific, stable binding interface. Here, the authors construct a protein–protein interface using only four amino acids, of which tyrosine is the most abundant, and obtain a seemingly good binding antibody.

24. Pal G, Kouadio JL, Artis DR, Kossiakoff AA, Sidhu SS:
•• **Comprehensive and quantitative mapping of energy landscapes for protein–protein interactions by rapid combinatorial scanning**. *J Biol Chem* 2006, **281**:22378-22385.

We take it for granted that chemically conserved amino acid replacements and conservation scores are a good way to determine which mutations are allowed in specific positions within a binding interface. In this study, a holistic approach seems to shatter this conception. Many of the allowed mutations are neither chemically nor evolutionary conserved, whereas many of the conserved mutations are not allowed. Moreover, the tendency of allowed mutations is very much location specific and is not directly related to whether a position is a hotspot for binding.

25. Vajdos FF, Adams CW, Breece TN, Presta LG, de Vos AM, Sidhu SS: **Comprehensive functional maps of the antigen-binding site of an anti-ErbB2 antibody obtained with shotgun scanning mutagenesis**. *J Mol Biol* 2002, **320**:415-428.

26. Bogan AA, Thorn KS: **Anatomy of hot spots in protein interfaces**. *J Mol Biol* 1998, **280**:1-9.

27. Roisman LC, Jaitin DA, Baker DP, Schreiber G: **Mutational analysis of the IFNAR1 binding site on IFNα2 reveals the architecture of a weak ligand-receptor binding-site**. *J Mol Biol* 2005, **353**:271-281.

28. Jaitin DA, Roisman LC, Jaks E, Gavutis M, Piehler J, Van der
• Heyden J, Uze G, Schreiber G: **Inquiring into the differential action of interferons (IFNs): an IFN-α2 mutant with enhanced affinity to IFNAR1 is functionally similar to IFN-β**. *Mol Cell Biol* 2006, **26**:1888-1897.

IFN-α and IFN-β are multifunctional cytokines that show differential activities through a common receptor composed of the subunits IFNA-R1 and IFNA-R2. Engineering an IFN-α2 triple mutant with 30-fold higher binding affinity for IFNA-R1 transforms this weak, hotspot-deficient interaction into a 50 nM affinity complex that mimics the biological properties of IFN-β. Conservation of the mutated residues in IFN-α subtypes suggests that IFN-α has evolved to bind IFNA-R1 weakly, apparently to sustain levels of biological activities different to those induced by IFN-β.

29. Svensson HG, Wedemeyer WJ, Ekstrom JL, Callender DR, Kortemme T, Kim DE, Sjobring U, Baker D: **Contributions of amino acid side chains to the kinetics and thermodynamics of the bivalent binding of protein L to Igκ light chain**. *Biochemistry* 2004, **43**:2445-2457.

30. Bedouelle H, Belkadi L, England P, Guijarro JI, Lisova O, Urvoas A, Delepierre M, Thullier P: **Diversity and junction residues as hotspots of binding energy in an antibody neutralizing the dengue virus**. *FEBS J* 2006, **273**:34-46.

31. Potapov V, Sobolev V, Edelman M, Kister A, Gelfand I: **Protein–protein recognition: juxtaposition of domain and interface cores in immunoglobulins and other sandwich-like proteins**. *J Mol Biol* 2004, **342**:665-679.

32. MacCallum RM, Martin AC, Thornton JM: **Antibody–antigen interactions: contact analysis and binding site topography**. *J Mol Biol* 1996, **262**:732-745.

33. •• Reichmann D, Rahat O, Albeck S, Meged R, Dym O, Schreiber G: **The modular architecture of protein–protein binding interfaces**. *Proc Natl Acad Sci USA* 2005, **102**:57-62.
By clustering the contact map of the interface, it is shown that the interface between proteins is built from several independent clusters of interactions. Mutations in one cluster do not affect residues located in a neighboring module. As a result, the structural and energetic consequences of the deletion of whole modules are surprisingly small. By contrast, within their cluster, mutations cause complex energetic and structural consequences.

34. Brinda KV, Vishveshwara S: **Oligomeric protein structure networks: insights into protein–protein interactions**. *BMC Bioinformatics* 2005, **6**:296.

35. Sohn J, Parks JM, Buhrman G, Brown P, Kristjansdottir K, Safi A, Edelsbrunner H, Yang W, Rudolph J: **Experimental validation of the docking orientation of Cdc25 with its Cdk2–CycA protein substrate**. *Biochemistry* 2005, **44**:16563-16573.

36. • Nevo R, Brumfeld V, Kapon R, Hinterdorfer P, Reich Z: **Direct measurement of protein energy landscape roughness**. *EMBO Rep* 2005, **6**:482-486.
Combining theory with single-molecule dynamic force spectroscopy experiments demonstrates the overall energy scale of roughness for a complex between the GTPase Ran and the nuclear transport receptor importin-β. The results give a degree of roughness that is consistent with the ability of importin-β to accommodate multiple conformations and to interact with different, structurally distinct ligands.

37. •• Moza B, Buonpane RA, Zhu P, Herfst CA, Rahman AK, McCormick JK, Kranz DM, Sundberg EJ: **Long-range cooperative binding effects in a T cell receptor variable domain**. *Proc Natl Acad Sci USA* 2006, **103**:9867-9872.
In protein–protein interfaces, a subset of hotspots contributes disproportionately to the affinity of the complex. These hotspot residues can be further clustered into hot regions. Here, the authors show that, in the complex formed between a variable domain of the T-cell receptor and a bacterial superantigen, significant long-range conformational effects are driven by cooperative binding energetics between distinct hot regions that are separated by more than 20 Å. They propose that these cooperative effects are propagated through a dynamic structural network.

38. Shulman AI, Larson C, Mangelsdorf DJ, Ranganathan R: **Structural determinants of allosteric ligand activation in RXR heterodimers**. *Cell* 2004, **116**:417-429.

39. Pal G, Ultsch MH, Clark KP, Currell B, Kossiakoff AA, Sidhu SS: **Intramolecular cooperativity in a protein binding site assessed by combinatorial shotgun scanning mutagenesis**. *J Mol Biol* 2005, **347**:489-494.

40. Schimmele B, Pluckthun A: **Identification of a functional epitope of the Nogo receptor by a combinatorial approach using ribosome display**. *J Mol Biol* 2005, **352**:229-241.

41. Frisch C, Schreiber G, Johnson CM, Fersht AR: **Thermodynamics of the interaction of barnase and barstar: changes in free energy versus changes in enthalpy on mutation**. *J Mol Biol* 1997, **267**:696-706.

42. Kiel C, Serrano L, Herrmann C: **A detailed thermodynamic analysis of Ras/effector complex interfaces**. *J Mol Biol* 2004, **340**:1039-1058.

43. Desrosiers DC, Peng ZY: **A binding free energy hot spot in the ankyrin repeat protein GABPβ mediated protein–protein interaction**. *J Mol Biol* 2005, **354**:375-384.

44. Guerois R, Nielsen JE, Serrano L: **Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations**. *J Mol Biol* 2002, **320**:369-387.

45. • Tobi D, Bahar I: **Structural changes involved in protein binding correlate with intrinsic motions of proteins in the unbound state**. *Proc Natl Acad Sci USA* 2005, **102**:18908-18913.
Here, the equilibrium motions of proteins that exhibit relatively large (non-local) conformational changes upon protein binding were explored. Gaussian and anisotropic network models of protein dynamics for four complexes show that the largest conformational change in its unbound form correlates with the experimentally observed structural changes upon binding. The results emphasize the pre-existing equilibrium of different conformations as a selection mechanism for protein–protein interactions.

46. • Keeble AH, Kleanthous C: **The kinetic basis for dual recognition in colicin endonuclease–immunity protein complexes**. *J Mol Biol* 2005, **352**:656-671.
Using complexes of colicin endonuclease (DNase) and immunity proteins as a model, the authors show that cognate complexes are many orders of magnitude more stable than non-cognate complexes, largely because of a much slower rate of dissociation. Pre-steady-state kinetics show that cognate and non-cognate complexes alike form a non-specific, conformationally dynamic encounter complex, which is most probably centered on conserved interactions at the interface and allows optimal alignment of interface specificity sites. Such a mechanism of kinetic discrimination explains the ability of DNase-specific immunity proteins to show dual recognition specificity, wherein they are broadly cross-reactive and yet highly specific.

47. Grunberg R, Nilges M, Leckner J: **Flexibility and conformational entropy in protein–protein binding**. *Structure* 2006, **14**:683-693.

48. Pearce KHJ, Cunningham BC, Fuh G, Teeri T, Wells JA: **Growth hormone binding affinity for its receptor surpasses the requirements for cellular activity**. *Biochemistry* 1999, **38**:81-89.

49. • Brown RJ, Adams JJ, Pelekanos RA, Wan Y, McKinstry WJ, Palethorpe K, Seeber RM, Monks TA, Eidne KA, Parker MW, Waters MJ: **Model for growth hormone receptor activation based on subunit rotation within a receptor dimer**. *Nat Struct Mol Biol* 2005, **12**:814-821.
Dimerization alone of the hGHR is insufficient to activate the cytoplasmic JAK2 kinase. Here, it is reported that the receptor can be activated by rotation even without ligand binding by inserting a defined number of alanine residues into the transmembrane domain. The authors propose an activation mechanism involving the relative rotation of subunits within a dimeric receptor as a result of asymmetric placement of the receptor-binding sites on the ligand.

50. •• Horn JR, Kraybill B, Petro EJ, Coales SJ, Morrow JA, Hamuro Y, Kossiakoff AA: **The role of protein dynamics in increasing binding affinity for an engineered protein–protein interaction established by H/D exchange mass spectrometry**. *Biochemistry* 2006, **45**:8488-8498.
Hydrogen-deuterium exchange mass spectrometry is used to assess the role of dynamics in the binding of a high-affinity hGH variant or its wild-type counterpart and the extracellular domain of the hGHR receptor. On the basis of the stability, binding kinetics and thermodynamic data presented, the increase in the binding free energy of high-affinity hGH is attributed to an increase in the free energy of the unbound state relative to the free energy of the bound complex. This implies that destabilizing one of the interacting molecules without destabilizing the resulting bound complex is a way to achieve higher binding affinity.

51. Wohlgemuth S, Kiel C, Kramer A, Serrano L, Wittinghofer F, Herrmann C: **Recognizing and defining true Ras binding domains I: biochemical analysis**. *J Mol Biol* 2005, **348**:741-758.

52. Kiel C, Selzer T, Shaul Y, Schreiber G, Herrmann C: **Electrostatically optimized Ras-binding Ral guanine dissociation stimulator mutants increase the rate of association by stabilizing the encounter complex**. *Proc Natl Acad Sci USA* 2004, **101**:9223-9228.

53. Prudencio M, Ubbink M: **Transient complexes of redox proteins: structural and dynamic details from NMR studies**. *J Mol Recognit* 2004, **17**:524-539.

54. Hofsteenge J, Servis C, Stone SR: **Studies on the interaction of ribonuclease inhibitor with pancreatic ribonuclease involving differential labeling of cysteinyl residues**. *J Biol Chem* 1991, **266**:24198-24204.

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

56. • Cho S, Swaminathan CP, Yang J, Kerzic MC, Guan R, Kieke MC, Kranz DM, Mariuzza RA, Sundberg EJ: **Structural basis of affinity maturation and intramolecular cooperativity in a protein–protein interaction**. *Structure* 2005, **13**:1775-1787.
A detailed structural analysis of the molecular mechanism of affinity maturation for the murine T-cell receptor Vβ domain is provided in this paper. Information from nine X-ray structures of combinations of mutations accumulated along the maturation pathway and comprehensive thermodynamics reveal the fine details of high-affinity binding.

57. Raschke TM: **Water structure and interactions with protein surfaces**. *Curr Opin Struct Biol* 2006, **16**:152-159.

58. Lawrence MC, Colman PM: **Shape complementarity at protein/protein interfaces**. *J Mol Biol* 1993, **234**:946-950.

59. Rodier F, Bahadur RP, Chakrabarti P, Janin J: **Hydration of protein–protein interfaces**. *Proteins* 2005, **60**:36-45.

60. Paliwal A, Asthagiri D, Abras D, Lenhoff AM, Paulaitis ME: **Light-scattering studies of protein solutions: role of hydration in weak protein–protein interactions**. *Biophys J* 2005, **89**:1564-1573.

61. Bhat TN, Bentley GA, Boulot G, Greene MI, Tello D, Dall'Acqua W, Souchon H, Schwarz FP, Mariuzza RA, Poljak RJ: **Bound water molecules and conformational stabilization help mediate an antigen-antibody association**. *Proc Natl Acad Sci USA* 1994, **91**:1089-1093.

62. Amadasi A, Spyrakis F, Cozzini P, Abraham DJ, Kellogg GE, Mozzarelli A: **Mapping the energetics of water-protein and water–ligand interactions with the 'natural' HINT forcefield: predictive tools for characterizing the roles of water in biomolecules**. *J Mol Biol* 2006, **358**:289-309.

63. • Binz HK, Amstutz P, Pluckthun A: **Engineering novel binding proteins from nonimmunoglobulin domains**. *Nat Biotechnol* 2005, **23**:1257-1268.
This review gives an excellent summary of current experimental systems and selection methods to achieve high-affinity binding through the construction of smart combinatorial libraries.

64. • Song G, Lazar GA, Kortemme T, Shimaoka M, Desjarlais JR, Baker D, Springer TA: **Rational design of intercellular adhesion molecule-1 (ICAM-1) variants for antagonizing integrin lymphocyte function-associated antigen-1-dependent adhesion**. *J Biol Chem* 2006, **281**:5042-5049.
The interaction between LFA-1 and its ligand ICAM-1 is low affinity. Here, multiple rational design methods are used to engineer ICAM-1 mutants with enhanced affinity for LFA-1. The affinity of the most improved ICAM-1 mutant for intermediate- and high-affinity LFA-1 I domains is increased by 19-fold and 22-fold, respectively, resulting in a similar increase in potency.

65. Joachimiak LA, Kortemme T, Stoddard BL, Baker D: **Computational design of a new hydrogen bond network and at least a 300-fold specificity switch at a protein–protein interface**. *J Mol Biol* 2006, **361**:195-208.

66. Pokala N, Handel TM: **Energy functions for protein design: adjustment with protein–protein complex affinities, models for the unfolded state, and negative design of solubility and specificity**. *J Mol Biol* 2005, **347**:203-227.

67. Kortemme T, Kim DE, Baker D: **Computational alanine scanning of protein–protein interfaces**. *Sci STKE* 2004, **2004**: pl2.