Cell. Mol. Life Sci. 65 (2008) 1059 – 1072
1420-682X/08/071059-14
DOI 10.1007/s00018-007-7451-x
© Birkhäuser Verlag, Basel, 2007

Cellular and Molecular Life Sciences

Review

The interface of protein-protein complexes: Analysis of contacts and prediction of interactions

R. P. Bahadur and M. Zacharias*

School of Engineering and Science, Jacobs University Bremen, Campus Ring 1, 28759 Bremen (Germany), Fax: +49-421-200-3249, e-mail: m.zacharias@jacobs-university.de

Received 1 October 2007; received after revision 9 November 2007; accepted 9 November 2007
Online First 15 December 2007

Abstract. Specific protein-protein interactions are essential for cellular functions. Experimentally determined three-dimensional structures of protein-protein complexes offer the possibility to characterize binding interfaces in terms of size, shape and packing density. Comparison with crystal-packing interfaces representing nonspecific protein-protein contacts gives insight into how specific binding differs from nonspecific low-affinity binding. An overview is given on empirical structural rules for specific protein-protein recognition derived from known complex structures. Although single parameters such as interface size, shape or surface complementary show clear trends for different interface types, each parameter alone is insufficient to fully distinguish between specific *versus* crystal-packing contacts. A combination of interface parameters is, however, well suited to characterize a specific interface. This knowledge provides us with the essential ingredients that make up a specific protein recognition site. It is also of great value for the prediction of protein binding sites and for the evaluation of predicted complex structures.

Keywords. Protein-protein interactions, protein-protein docking, protein-binding region, crystal packing, nonspecific interaction, protein interface description, binding site prediction.

Introduction

Almost all biological processes involve protein-protein interactions ranging from essential functions in the immune system, the regulation of enzymes, signal transduction or mediating the adhesion of cells. Furthermore, many (if not most) biological functions require not only pair-wise but multiple protein-protein interactions. These interactions form the basis of the quaternary structure of multimeric proteins, and represent one of the most complex levels of structural organization in biological molecules. To mediate a specific biological function or to form a defined multimeric structure the specificity of protein-protein interactions is of crucial importance. The immense importance of these interactions in biological systems has made the protein-protein recognition process an area of significant interest. The understanding of protein-protein interactions and its specificity at atomic detail requires the knowledge of the three-dimensional (3D) structure of protein complexes and protein-protein interfaces.

Protein X-ray crystallography, nuclear magnetic resonance (NMR) spectroscopy and cryo-electron microscopy (CryoEM) are the main experimental techniques that have provided the atomic structure of many protein-protein complexes. During the last three decades, structural biologists have provided the 3D structures of a large number of protein-protein

* Corresponding author.

Table 1. List of the web servers for the analysis of protein-protein interfaces.

Name of the servers	Web address
Protein-protein interaction server	http://www.biochem.ucl.ac.uk/bsm/PP/server/
DIP : Database of interacting proteins	http://dip.doe-mbi.ucla.edu/
CAPRI: Critical assessment of prediction of interactions	http://capri.ebi.ac.uk
PROFACE: A server for the analysis of the physicochemical features of protein-protein interfaces	http://www.boseinst.ernet.in/resources/ bioinfo.stag.html
The morphology of protein-protein interfaces	http://mgl.scripps.edu/people/goodsell/interface
PISA: Protein interfaces, surfaces and assemblies	http://www.ebi.ac.uk/msd-srv/prot_int/ pistart.html

complexes ranging from binary (homodimer, enzyme-inhibitor, antigen-antibody) to multi-subunit (oligomers, chaperones, virus capsid) complexes, which have helped immensely for the understanding of the recognition processes. Recently, large-scale studies of protein complexes in yeast have demonstrated that the majority of proteins in the cell exist as parts of multi-component assemblies, to a large degree novel and of unknown function [1]. The structural and functional analysis of these complexes could become a priority for structural biologists in the coming years. A wealth of chemical, physical and geometrical data on protein-protein complexes has already become accessible through the Protein Data Bank (PDB; www.rcsb.org) [2], and several Web servers for the structural analysis of the protein-protein complexes (Table 1). General analysis on structural aspects of protein-protein interactions has been carried out in detail by the comparison of protein-protein complex structures [3–18]. In the following, an overview is given on the interface analysis of known protein-protein complex structures and how this knowledge can be used to gain a better understanding of the mechanism of specific recognition. The focus is on observed types of contacts at protein interfaces and on the structural as well as physicochemical character of protein-protein interfaces. The analysis is restricted to those protein-protein complexes for which atomic resolution structures are available. Furthermore, interface analysis can also be helpful to predict the geometry of protein-protein binding structures and may also give hints to specifically influence protein-protein interactions. Its possible impact on identifying protein-binding regions and prediction of protein complex geometries is also discussed.

Structural diversity of protein-protein recognition

Among the protein-protein complexes that have been determined experimentally one can distinguish two major groups: binary and multi-subunit complexes. Depending on the nature of the polypeptide chain, the binary complexes can be further subdivided into two major classes: homodimers, where two interacting polypeptide chains have identical amino acid sequence; and heterodimeric complexes, where the sequence of two interacting polypeptide chains is different. The subunits that form homodimers are typically not found as stable structures inside the cell and the complex formation occurs simultaneously during the folding process. On the other hand the subunits of heterodimeric complexes are often, but not always, independently stable inside the cell and they interact with each other to carry out a specific function in the cell. Most of the 3D structures of protein-protein complexes have been determined using X-ray crystallography. Formation of such protein crystals also involves non-covalent interactions similar to those found in homodimers and heterodimers, yet they are not subject to any natural selection and, thus, they lack biological specificity [19]. In the following, interfaces formed in homodimers and heterodimers are termed specific interfaces as they occur within the living cell, whereas the interfaces formed by the crystal contacts (also between monomeric proteins) are called nonspecific interfaces as they typically do not occur in biological processes and are ‘artifacts’ of the crystallization process.

Interfaces in homodimeric and oligomeric proteins

3D atomic structures of homodimeric proteins are abundantly represented in the PDB. Generally, most of the homodimers are permanent and their assembly and folding process take place simultaneously; but some may dissociate into monomers or form larger oligomers depending on concentration, pH, ligand binding, and other parameters that can modulate protein–protein interactions. An example for an “obligate” dimer is the bacteriophage P22 Arc repressor DNA binding protein (pdb:1arr). It consists of two identical chains that associate and fold simulta-

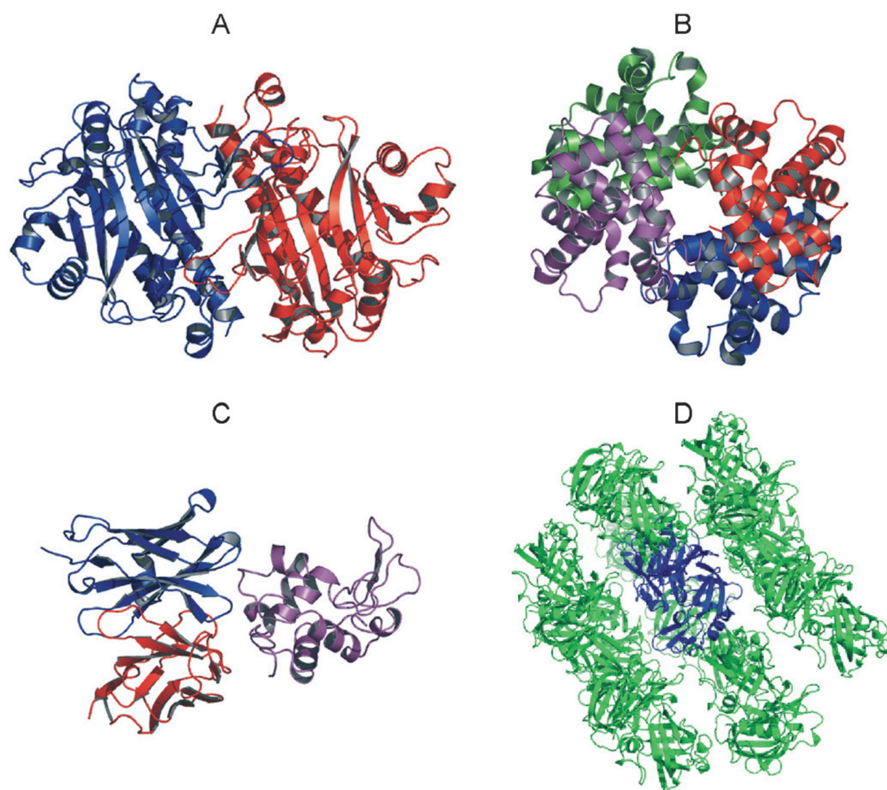

Figure 1. Assembly of protein-protein complexes. (*A*) Homodimeric assembly of asparagine synthetase (pdb:12as [95]). The two subunits are colored red and blue, respectively. (*B*) Hemoglobin tetramer (pdb:1fsx [96]). The two chains in the alpha subunit are colored in blue and green, and the other two chains in beta subunit are colored in red and magenta, respectively. (*C*) Heterocomplex assembly of antigen and antibody: Crystal structure of the FV fragment of mouse monoclonal antibody D1.3 (the light and heavy chains are colored red and blue, respectively) complexed with hen egg lysozyme (in magenta, pdb:1vfb [97]). (*D*) Crystal-packing interactions in crystal lattice: Crystal contacts in monomeric chymosin (pdb:3cms [98]). One central biologically active monomer subunit is shown in blue.

neously, and exists only as dimer in solution [20]. On the other hand, "non-obligate" dimers are active as monomers and become inactive upon dimer formation (reviewed in [20]). Figure 1A shows the contact between two subunits in dimeric asparagine synthetase. The surface involved in these contacts is more hydrophobic than the rest of the protein surface and also similar to the protein interior [15, 21].

In the higher order of the homo-oligomeric assembly, such as tetramer, hexamer, octomer, and dodecamer, etc., more than two identical subunits come in contact. Oligomerization occurs due to strong selection pressure for the evolution of monomeric proteins into oligomeric complexes, driven by benefits such as reduction of surface area, increased stability and novel function through inter-subunit communication [20–31]. As an example, the tetrameric association of the hemoglobin is shown in Figure 1B. The protein is viewed along its molecular twofold rotation axis (shown as red circular dot), which relates the two α β protomers.

Interfaces in heterodimeric complexes

Typical heterodimeric complexes are non-covalent assemblies of proteins that fold separately to carry out independent functions and associate under certain physiological conditions. These types of interactions are found in protease-inhibitor, antigen-antibody, enzyme complexes, and many signal transduction and cell cycle proteins [9, 12, 13]. Although their recognition sites share some common properties, protein hetero-complexes are extremely diverse in their cellular functions. Figure 1C shows an antigen-antibody complex (monoclonal antibody D1.3 interacting with hen egg lysozyme) as a typical example of a stable heterodimeric complex. Many cellular functions are mediated through transient protein-protein interactions with short lifetimes. The structure determination by protein crystallography, however, requires stable complex structures that can form well-ordered crystals. The present analysis is restricted to sufficiently stable protein-protein complexes and it might be possible that the recognition principles for such stable interactions differ from interfaces formed during transient interactions with a short lifetime.

Crystal-packing interfaces and nonspecific protein-protein interactions

Crystal contacts are made by molecules related by the symmetries and translations of the crystal lattice (Fig. 1D). These contacts are essential to build up a well-ordered assembly that attains macroscopic di-

mensions. Many crystal contacts are artifacts of crystallization that may not occur in solution or in the physiological state to any significant extend. Although some of these observed contacts may have biological relevance, in the majority of cases such contact interfaces represent short-lived protein-protein associations, which can occur transiently between any pair of proteins in a cell. These interfaces are called nonspecific interfaces as they do not have any biological specificity, and are abundantly available in the PDB as they govern the molecular packing of protein crystals.

In crystals the major types of interactions are found where the subunit contacts are related by a twofold symmetry (180 degree rotation about an axis called twofold axis). Besides the twofold symmetry, crystal-contact interfaces can also have higher point group symmetry, which are, however, quite rare [24]. Janin and Rodier [19] have shown that crystal interfaces that incorporate a twofold symmetry, on average, produce larger and probably more stable interfaces than those, which do not have this symmetry. It has been found that the physicochemical properties of large crystal-packing interfaces varied widely compared to the specific interfaces, which are very useful features that help to discriminate between the specific and nonspecific interfaces [16, 32–40]. The comparison of the physicochemical character of interfaces due to crystal contacts, and in particular bound protein-protein complexes, can also help to identify the key features that distinguish specific from nonspecific contacts (see below).

Structural properties of the interfaces

Size of the interfaces: Interface area

The size of the protein-protein interfaces can be quantified by calculating the interface area (given as B). It can be calculated in terms of the solvent accessible surface area (SASA) [41–43] of the protein molecules and is given by the following equation:

$$B = SASA_{subunit1} + SASA_{subunit2} - SASA_{complex} \qquad (1)$$

The first two terms on the right hand side of the equation gives the SASA of two interacting subunit and the last term represents the same for the complex (Fig. 2). The interface area B is the area of the two interacting protein surfaces that becomes buried when two molecules associate. The SASA can be calculated from the atomic coordinates of the isolated subunit by rolling a solvent probe (with the radius of a water molecule) over the surface of the protein. Table 2 gives the average size of the protein-protein interfaces in specific and nonspecific complexes (values are taken from [16]). On average, specific interfaces are larger in size compared to nonspecific interfaces. Homodimers are on average 2-fold larger than heteromeric protein-protein complexes and about 2.5-fold larger than the crystal-packing interfaces of monomeric proteins. Most of the heterodimeric protein-protein complexes bury a surface area in the range of 1200–2000 Å^2, which is a “standard size” interface according to Lo Conte et al. [12]. It should be noted that the average buried surface area in the case of crystal-packing interfaces given in Table 2 was obtained from a subset of 188 unique pair-wise crystal-packing interfaces with a significant B value of $\geq$800 Å^2 [16]. A smaller average B of 570 Å^2 has been obtained in a sample of 1320 packing interfaces found in 152 crystals of monomeric proteins [19]. There are also cases of exceptionally large crystal-packing interfaces. For example, the crystal dimers of hexokinase (pdb:1qha), creatine kinase (pdb:1cki_1) and a MHC class I homologue (1b3j) have interfaces with B near 3400 Å^2, equivalent in size to the average homodimer interface. As further illustrated in Figure 3 the distribution of interface sizes for homodimeric, heterodimeric proteins as well as for crystal contacts is quite broad, so that the corresponding distributions overlap significantly. In general, the size of the interfaces ranges from as small as ~800 Å^2 to very large with buried surface area more than 10 000 Å^2 (in some homodimeric complexes, e.g., copper amine oxidase dimer which buries 14 300 Å^2 at the dimer interface). This result indicates that the size of the buried surface area alone is not sufficient to distinguish specific contacts in protein-protein complexes from crystal contacts.

However, the analysis clearly indicates that there seems to be a lower limit of the buried surface area for specific recognition. The lack of interfaces with a size below 800 Å^2 indicates that the formation of a stable and specific complex between two proteins requires a sufficient number of contacts and removing water at least from part of the protein interface.

Shape of the interfaces

Besides of the size of a protein-protein interface the shape or curvature of an interface could be an interesting feature to distinguish specific and nonspecific recognition. Jones and Thornton [9] derived a parameter termed planarity to assess whether a protein-protein interface is flat or curved. The ‘planarity’ of the surfaces between two components of a complex is analyzed by calculating the root mean square deviation of all the interface atoms from the least-squares plane through the atoms. If all atoms would exactly fit to a plane, the planarity index would

Table 2. Structural features of the protein-protein recognition sites (from [16]).

Parameters	Homodimers	Heterodimers	Crystal packing
Number of complexes	122	70	188
Interface area BSA (Å^2)	3880	1910	1510
Number of interface			
Atoms	396	204	160
Residues	104	57	48
Fraction of			
Non-polar area (f_np)	65	58	58
Buried atoms (f_bu)	36	34	21
Core residues	77	72	–
Hydrogen bonds			
Number per interface	19	10	5
BSA per bond (Å^2)	210	190	280
Interface waters			
Number per interface	44	20	23
Number per 1000 Å^2 BSA	11	10	15
Planarity index	3.46	2.80	2.77
Circularity index	0.71	0.73	0.77
GV index	2.1	2.5	4.4
SC score	0.70	0.69	0.63
LD index	45	42	32
GD index	1.30	1.32	0.96

be zero. The average value of the planarity is 3.5 ± 1.7 Å for homodimers and 2.8 ± 0.9 Å for heterocomplexes [9]. Crystal-packing interfaces are also as planar as protein heterocomplex interfaces with an average planarity index of 2.8 ± 0.8 Å. This result indicates that protein-protein interfaces, whether homo- or hetero, are generally flat in shape [4, 9]. Exceptions are, however, some enzyme-inhibitor complexes, where the inhibitor site forms a convex surface fitting to the often concave shape of the enzyme active site. It should also be mentioned that in contrast to protein-protein interfaces binding sites for enzyme substrates of small organic ligands are usually very non-planar allowing contacts to the ligand from many different sides.

Another parameter to characterize the shape of protein binding sites is the 'circularity' of the interface, which was also calculated by Jones and Thornton [9]. The circularity is the ratio of the lengths of the principal axes of the least-squares plane through the atoms in the interface. A ratio near 1.0 indicates that an interface is approximately circular meaning that the atoms that form the buried interface are all within a circular plane. On average, the interfaces are not perfectly circular. In general, all the three types of interfaces, interfaces of homodimers (0.71 ± 0.17), heterocomplexes (0.73 ± 0.05) and crystal-packing of monomers (0.77 ± 0.09) do not significantly differ in terms of circularity.

Shape complementarity is a measure for the interfacial packing of protein complexes. One way to calculate shape complementarity is to relate the volume of the interface cavities to the interface area. Laskowski [43] defined an index called gap volume (GV) index, which is given by the following equation:

$$\text{GV (Å)} = \text{gap volume between molecules (Å}^3\text{)/ interface area (Å}^2\text{) (per complex)} \qquad (2)$$

The GV corresponds to the volume of interface cavities normalized with respect to the buried surface area [43]. The average value of the Gap index is 2.10 ± 1.2 Å for homodimers, 2.5 ± 1.0 Å for protein heterocomplexes, and 4.4 ± 1.9 Å for crystal-packing interfaces (Table 2) [9, 16]. The large value of GV in crystal-packing interfaces suggests that they contain a significantly larger cavity volume per unit area of the interface compared to homodimers and heterocomplexes. In the above example, the GV index for the dimeric alkaline phosphatase is 1.09 Å, meaning that it is very well packed compared to the average GV index of homodimeric and heterodimeric complexes. A combination of interface area and GV packing index appears to be a quite effective measure to

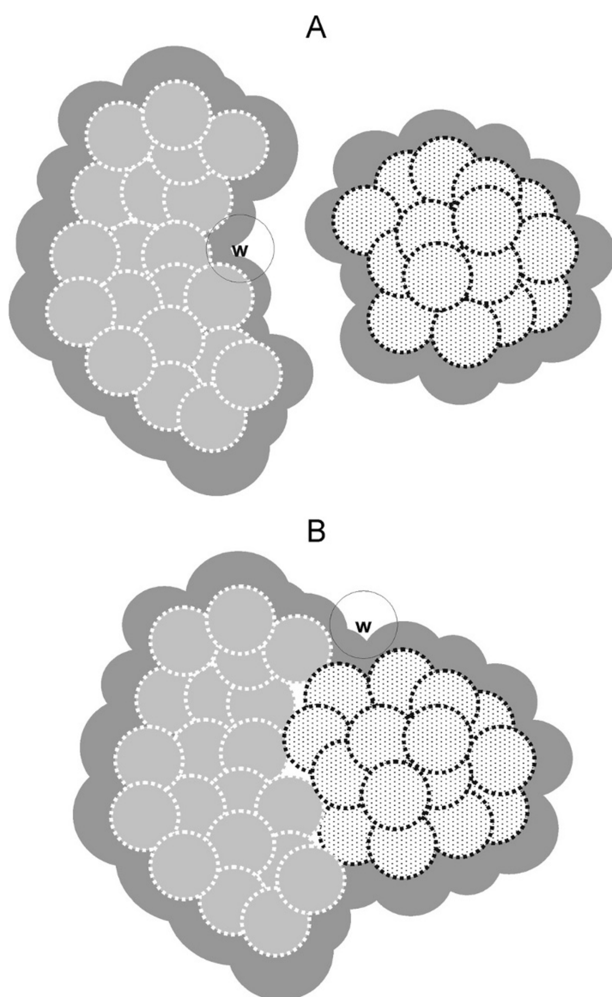

Figure 2. Change in solvent accessible surface area upon complex formation. The solvent accessible surface (shown in dark gray) is computed by rolling a water probe over the surface of molecules. In the complex, the water (shown as a sphere marked w) is excluded from the interface between two subunits and the surface area lost by them is called buried surface area or interface area.

distinguish specific interfaces from nonspecific crystal-packing interfaces (Fig. 4).

Bahadur et al. [16] have developed two packing indices that are less sensitive to errors in the atomic positions of the X-ray structures than methods that explicitly consider the size of interface cavities. The local density (LD) index is the mean number of interface atoms that are within 12 Å of another interface atom. This measures the packing density at each point of the interface. The average value of LD index is 42–45 (neighboring interface atoms) for the specific interfaces and almost 30% lower in the crystal-packing interfaces, which suggests that the specific interfaces are significantly better packed compared to the nonspecific interfaces. The other packing index parameter called the global density (GD) index measures the atomic density at the interface atoms normalized to the dimensions of the interface [16]. In Table 2 the average values of the GD index suggests that the atom density at the interface in specific protein-protein complexes is significantly higher compared to the nonspecific interfaces. The comparative analysis using different methods to define interface-packing density consistently indicates that surface complementarity (as measured by packing density) is a parameter to distinguish between specific and nonspecific (crystal) protein-protein contacts. The differences in interface packing for a nonspecific crystal contact and a specific homodimer complex are illustrated in Figure 5.

Clustering of interface atoms

Interface atoms may form one contiguous patch or they can be distributed in several patches throughout the interface. Patches are clusters of residues or atoms in contact with the partner protein separated from other patches by surface residues or atoms that are not in contact with the protein-binding partner. One possibility to identify such patches is to perform a cluster analysis of the distances between interface residues or atoms using the average-linkage clustering algorithm [44]. Residues belonging to one patch have distances smaller than a threshold distance (D) to separate residues that belong to a different patch. This method has been applied to heterodimeric and homodimeric protein complexes [13, 15]. Depending on the value of D, the interfaces may be considered to be made up of a single patch or multiple patches. In general, most of the single patch interfaces in homodimers and in heterodimers are standard size interfaces (single patch), and large interfaces contain more than one recognition patch. Figure 6A and B shows a typical example of a single patch interface of dimeric asparagine synthetase and a two-patch interface of the BTB domain of promyelocytic leukemia zinc finger protein. The larger interfaces involving several patches in protein-protein complexes are often contributed by separate structural domains [13]. In crystal-packing interfaces of monomeric proteins, the interface atoms are scattered throughout the interfaces and a single significant patch cannot be defined [16].

Spatial distribution of interface residues

Interface residues can be divided into two distinct regions, the 'core' and the 'rim' region, based on their accessibility to solvent (illustrated in Fig. 6C, D). The 'core' region contains residues that have at least one fully buried interface atom (i.e., zero SASA after complex formation) surrounded by the 'rim' region, which contains residues having accessible atoms only [13]. Core residues having at least one or more atoms with zero SASA are colored in red, whereas rim

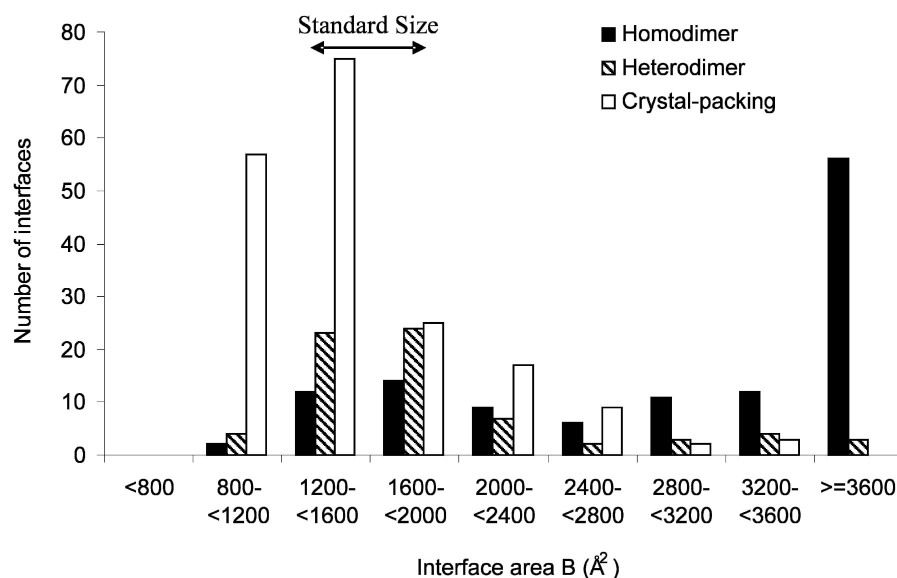

Figure 3. Size distribution of the protein-protein interfaces. Interface distributions in homodimers, heterodimers and crystal packing of monomeric proteins are shown in black, dash and white bars respectively. Data sets are the same as used for average interface characteristics presented in Table 2 (see also text for details and [15]). Note that only significant crystal contact areas above 800 Å^2 have been included.

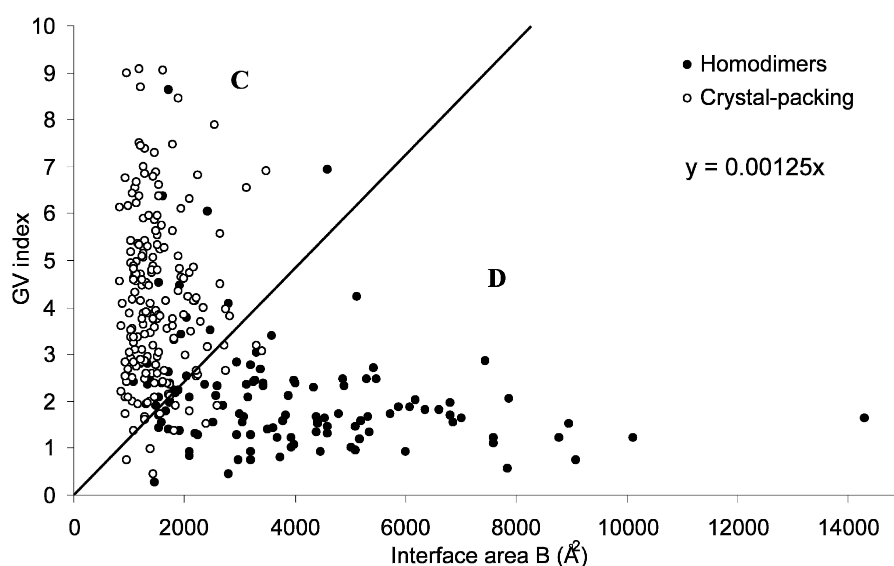

Figure 4. Discriminating homodimers and crystal-packing interfaces. In the scatter plot of the interface area (Å^2) *versus* the gap volume index (GV, in Å), region D contains mainly dimeric interfaces and region C corresponds mainly to crystal-packing interfaces. These two regions can be separated with a straight line (given by: GV = 0.00125 × interface size).

residues, which are accessible to the solvent, are colored in blue. The two regions differ in their amino acid composition: The rim is very similar to the remainder of the protein surface, whereas the core region depicts the protein interior. On average the 'core' constitutes 72–77 % of the interface area of homodimers and heterodimeric complexes [13, 15]. In crystal-packing interfaces of monomeric proteins one cannot define any 'core' and 'rim' region, since they contains very few buried atoms and the amino acid composition of the interface is similar to the rest of the protein surface [16].

Chemical composition of the protein-protein interfaces

The composition of amino acid residues at the specific interfaces differs from the rest of the protein surface. Relative to the solvent-accessible protein surface, homodimer and heterodimer complex interfaces are enriched in aliphatic (Leu, Val, Ile, Met) and aromatic (His, Phe, Tyr, Trp) residues, and depleted in charged residues (Asp, Glu, Lys) other than arginine [9, 11–15, 22, 26]. Arginine makes overall similar contributions to all the three kinds of protein-protein interfaces. Lysine is more abundant than arginine on the protein surface, but it is largely excluded from homodimer interfaces [8, 12–15]. The greater capacity of the guanidinium group in Arg to form hydrogen bonds compared to the amino group of Lys is a likely reason

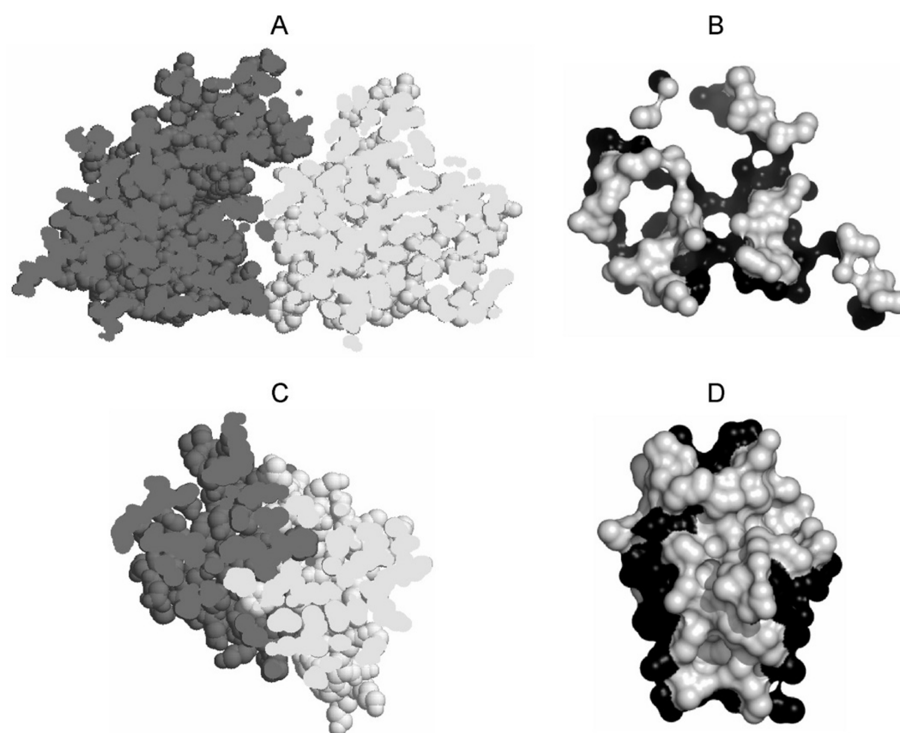

Figure 5. Shape and complementarity of interfaces. (*A*) Slab mode presentation of a crystal-packing dimer (cut in the middle, left) and the corresponding interface (perpendicular view, right) of monomeric *N*-myristoyl transferase (pdb:1nmt [99]). (*B*) Met repressor homodimer (pdb:1cmb [100]) interface (right: slab mode; left: interface atoms). In both cases, subunits are colored in white and gray, respectively. Crystal-packing interface is loosely packed with large number of cavities at the interface compared to the tightly packed homodimer interface without any interface cavity.

for the preference of Arg at interfaces over Lys. In contrast, the composition of the large crystal-packing interfaces is not significantly different from that of the accessible protein surface [16]. Polar and charged residues (Asn, Gln, Glu, Lys) are abundant and hydrophobic residues (Ile, Leu, Met, Val) are depleted in the crystal-packing interfaces as observed in rest of the protein surface.

At a higher resolution, the chemical groups at the protein surface may be divided into two types: non-polar (carbon containing) and polar (N, O, and S containing). On average non-polar chemical groups contribute 58 % of the protein surface area buried in large crystal-packing interfaces [16], which is very close to the average composition of the SASA of proteins [12]. Interfaces formed in heterodimeric and homodimeric proteins tend to be more hydrophobic: on average 65 % non-polar and 35 % polar [8, 15, 22]. These differences must relate to the modes of assembly in these three types of proteins. The protein surface that is buried at the interface of protein heterodimers is in contact with water until the complex assembles. Large hydrophobic patches on the surface would increase the tendency of aggregation, which would make the protein insoluble. In contrast, burying of hydrophobic patches in homodimers favors the formation of a permanent assembly that can form already during folding of the protein (coupled structure formation and dimerization).

Upon association, polar groups at interfaces can form hydrogen bonds (H-bond). The comparative analysis of protein complexes indicates that there are on average about 10 direct H-bonds in each protein heterodimer and 19 in each homodimer [12, 15]. In both cases, one H-bond corresponds to about 200 $Å^2$ of buried surface area (Table 2). The hydrogen-bonding interactions in crystal-packing interfaces of monomeric proteins are smaller in number (only 5 per interface and 280 $Å^2$ of buried surface area per H-bond) compared to specific interfaces [16].

Besides of hydrophobic contributions due to the desolvation of non-polar groups, complex formation can also create tight van der Waals contacts that are weaker but are more abundant at interfaces than H-bonds. Van der Waals interactions are due to the polarizability of atoms and act only over short distances. As indicated in the previous paragraph the interface of specific protein complexes is much more tightly packed with few cavities and it is expected that van der Waals interactions make a significant favorable contribution to specific protein-protein recognition.

Electrostatic interactions have much longer range than van der Waals interactions. Apart from an influence on protein-protein affinity and specificity, electrostatic interactions can also have an influence on the process (diffusional encounter) of protein association. Although high ion concentration can efficiently screen electrostatic interactions, it can influence the

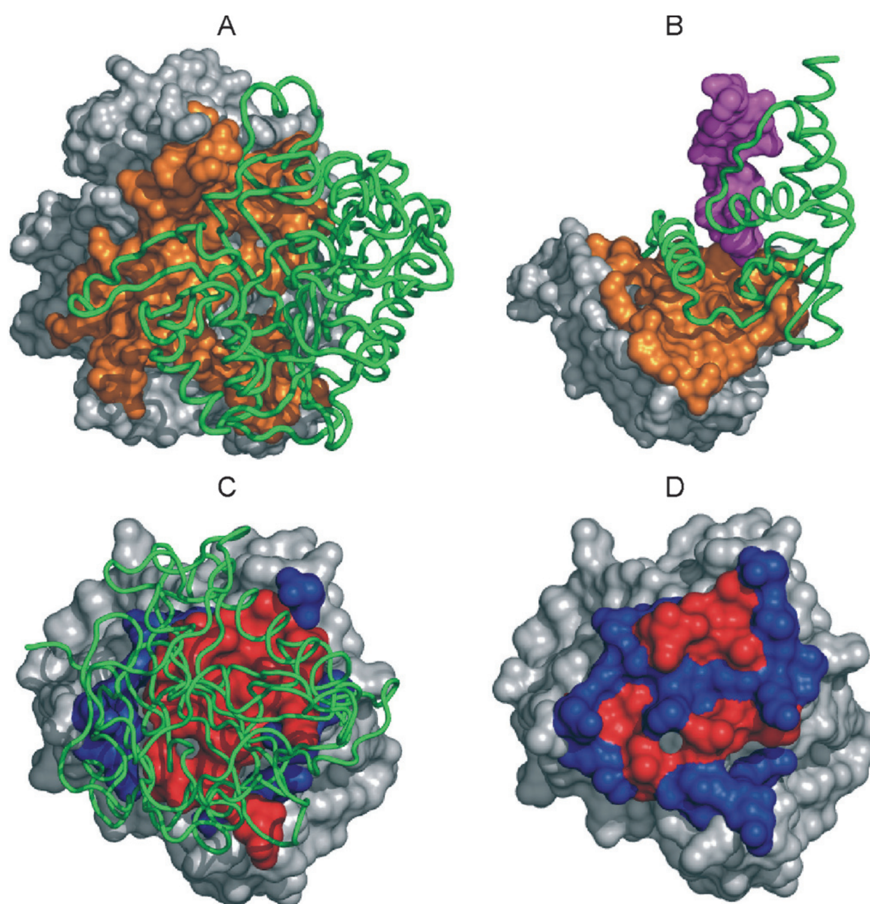

Figure 6. Architecture of protein-protein interfaces. (*A*) Single patch interface (in orange) of dimeric asparagine synthetase (pdb:12as [95]). (*B*) Two patch interface (in orange and magenta) of the BTB domain of promyelocytic leukemia zinc finger protein (pdb:1buo [101]). The two patches mainly represent the two classes of intersubunit contacts of this domain swapped protein. The patch colored in magenta corresponds to the "closed interface", involving mainly beta sheets, and that colored in orange corresponds to the central "open interface", mainly involving alpha helices [101]. In both the diagrams one subunit is shown in molecular surface representation and the other subunit is drawn as tube. (*C*) 'Core' and 'rim' interface of a protein-protein interface. One subunit is shown as molecular surface, and the other subunit is drawn as tube representation. The interface 'core' and 'rim' are colored in red and blue respectively. (*D*) Subunit interface is colored according to the hydrophobic (red) and hydrophilic (blue) surface area. Residues Ala, Gly, Ile, Leu, Met, Phe, Pro, and Val are considered as hydrophobic and the rest are hydrophilic. Note, one of the central residues is a Tyr, which although counted as hydrophilic could, however, also counted as hydrophobic.

rate of collision between two macromolecules bearing net electric charges or dipoles and can pre-orient protein partners to guide association [45]. The Coulomb attraction between charges of opposite sign can make collisions more frequent. An example is the barnase/barstar system (a protease/inhibitor complex) where the net positive charge on barnase and net negative charge on barstar results in an enhanced association rate that is more sensitive to ionic strength than the association for other enzyme inhibitor systems (e.g., trypsin-PTI) or the association of lysozyme and an antibody [46].

Role of water molecules in protein-protein association

Water molecules are present in abundance at the protein-protein interfaces, and play a major role in polar interactions that stabilize the complexes [12, 23, 47–57]. When two polypeptide chains interact with each other and form a stable specific complex or oligomeric structures, bulk solvent is excluded from the interface, which allows direct protein-protein contacts, creating a favorable hydrophobic effect as a major driving force for association. In spite of this phenomenon, structural biologists often report water molecules in the X-ray structures, which are bound to the interfaces between the polypeptide chains. These solvent molecules are supposed to contribute to the stability of the quaternary structure by forming extensive hydrogen-bonding networks with the amino acid residues at both sides of the interface. Bhat et al. [47] explored the role of interface water molecules in stabilizing the antigen-antibody association (Fig. 7). Extensive water-mediated H-bonds at

the proteinase-inhibitor interface were also identified by Huang et al. [48].

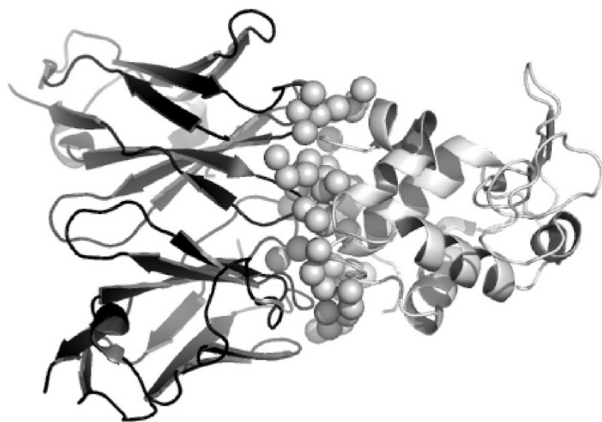

Figure 7. Distribution of water molecules at a protein-protein interface. Water molecules (shown as light gray sphere) at the interface of the Fv fragment of mouse monoclonal antibody D1.3 (the light and heavy chains are shown as black cartoon) complexed with hen egg lysozyme (gray cartoon; pdb:1vfb [47]).

For the analysis of the water–protein interactions, only high-resolution X-ray structures can be used because water positions are often not reported in medium resolution coordinate sets. A general study on the interface water molecules at the interfaces of protein hetero- and homo-complexes revealed that there is an average of one interface water molecule per 100 Å^2 (Table 2) of the interface area, suggesting that water-mediated polar interactions are even more numerous than direct hydrogen bonds (one H-bond per 200 Å^2 of interface area). Crystal-packing interfaces are relatively more hydrated with one interface water per 65 Å^2 of the interface area [12, 52]. This suggests that the water-mediated polar interactions contribute (in terms of the number of interactions) as much as direct protein-protein hydrogen bonds to the stability of protein-protein complexes. At the protein-protein interfaces, Rodier et al. [52] found that each water molecule makes on average one hydrogen bond with the interface amino acids, which is equivalent to ~1.5 kcal/mol energy contributed by each interface water molecule.

According to the spatial distribution of the interface water molecules, Janin [49] distinguished between two types of interface: One is the 'dry' interface where water molecules forming a ring around interface atoms and are almost completely excluded from the center of the interface. Another is the "wet" interface with water molecules scattered throughout the interface. Rodier et al. [52] developed a parameter that quantifies this phenomenon and have found that the specific interfaces are generally 'dry' in nature, whereas most of the crystal-packing interfaces are 'wet'. This suggests that specific interfaces are well packed and the interface residues are well shielded from the solvent molecules, thus, maximizing the hydrophobic effect during the complex formation.

Conformational changes upon protein-protein complex formation

An essential feature of protein-protein association is the way it affects the conformation of the protein components. Conformational changes may mediate signaling events or trigger allosteric effects. The extent and nature of the conformational changes can be assessed when the X-ray structures of the individual components and that of the complex are known independently. For moderate changes, a convenient measure of the extent of the changes is the root-mean-square distance (rmsd) of equivalent main chain atoms after least-squares superimposition of the bound and free components. Many protein-protein complexes display an rmsd in the range 0.5 – 1.0 Å. At the interface, the main chain can often move by 1 – 2 Å and a few surface side chains reorient [5, 12, 58]. Local movements of up to 1 Å and side chain rotations are also seen at crystal contacts when different crystal structures of the same protein are compared [59]. A majority of the protein antigens also undergo main chain movements of 1 – 2 Å as they bind antibodies [12].

There are, however, complexes where the rmsd between bound and unbound protein partners is well over 2 Å and large movements take place. The conformational change modifies the shape and chemical character of regions of the protein surface (induced fit). It may be localized to loop regions at the interface or affect the whole protein structure (e.g., the movement of whole domains upon association). Essentially all complexes with interface area larger than 2000 Å^2 for which the structures of the free components are known, undergo large changes upon association. One plausible explanation is that, for large interface sizes, the pre-formation of a stable (rigid) protein surface that exactly fits to the surface of a partner protein is more difficult than for a smaller interface. Hence, for larger interfaces a greater capacity for conformational adaptation is required to form a stable protein-protein complex.

Computational approaches for predicting protein-protein complex structures

Experimental structure determination of protein complexes for example by X-ray crystallography requires purification of large amounts of proteins

and the ability to crystallize the protein-protein complex. Ultimately, the knowledge of all protein-protein complex structures is desirable but experimental determination is costly and may not be feasible for all known interacting proteins. The realistic prediction of protein-protein complex structures (protein-protein docking) is therefore of increasing importance.

The analysis of known protein-protein complexes as described above could be helpful to guide predictive methods and to evaluate predicted protein-protein complex structures. The problem of docking a protein, rather than a small ligand, onto another protein was first considered by Wodak and Janin [60]. Since then, many algorithms for docking two macromolecules have been developed based on different approaches, such as: simplified protein models [60–62]; shape complementarity [63–69]; geometric hashing [70, 71]; or correlation methods [72–74]. An essential test of a docking procedure is its capacity to handle "unbound" molecules, that is, atomic coordinate derived from the structure of the isolated partner proteins.

The CAPRI (critical assessment of prediction of interactions; [75]) experiment for assessing blind predictions of protein-protein interactions is playing an important role in the development of docking procedures. CAPRI (http://capri.ebi.ac.uk) is a community-wide docking experiment analogous to CASP (critical assessment of structure prediction; http://predictioncenter.llnl.gov/casp1/Casp1.html), but aimed at assessing the performance of protein-protein docking procedures. The predictions are performed blindly and assessed by an independent team by comparison to X-ray structures. The structures of the complex are revealed only at the time of the evaluation [75, 76].

Docking procedures start from the atomic coordinates of the two interacting proteins, generate putative complexes, and give them a score [77]. In principle, there are computational methods such as molecular dynamics (MD) simulations available that can treat both binding partners as fully flexible. However, these techniques are very time consuming and limited to cases where the binding placement is approximately known and if experimental data can be included to restrict the search for putative protein-protein docking geometries [78]. To reduce the computational complexity most of the current systematic protein-protein docking procedures assume rigid-body association. Within the rigid-body approximation, the problem has only six degrees of freedom (three translational degrees and three rotational degrees). Conformational changes are then only taken into account in a second refinement stage (reviewed in Smith and Sternberg [79]). More recently, however, methods have also been developed that cover both local conformational changes of side chains [62, 80, 81] and loop motions [82] as well as global motions at least approximately during systematic docking searches [83–85].

Identification of protein-binding sites and scoring of predicted protein-protein complex structures

Beside of the appropriate inclusion of conformational flexibility during docking, a major problem of protein-protein docking is the realistic scoring of docked complexes. One major reason is that the understanding of the many energetic and entropic contributions to bio-molecular association is still incomplete. Secondly, even computationally demanding calculations at atomic resolution including explicit solvent molecules often fail to give quantitative agreement with experiment due to force field artifacts or sampling problems. For scoring of docked solutions, the additional difficulty is encountered that it should be computationally sufficiently fast to allow for screening thousands of solutions within a few hours of computer time. A promising alternative is the application of knowledge-based energy functions. These knowledge-based energy functions are based on the contact frequencies of amino acid residues (or atoms) at the interface of known protein-protein complexes. A "free energy"-like score can be defined for each contact by calculating the logarithm of the frequency of that observed divided by the frequency of expected contacts (expected based on the general occurrence of amino acids on protein surfaces). This approach can be extended to look also at a distance dependence of contacts and it is also frequently used to evaluate models of folded proteins (e.g., protein threading approach, [86]). Such knowledge-based scoring functions have been used frequently to score protein-protein docking solutions (e.g., [87, 88]).

A step further in this direction is to make use of the whole information extracted from the analysis of known protein-protein interfaces as presented in the previous paragraphs. The analysis could be used in two different ways to improve the prediction accuracy of protein-protein docking: It can be applied to predict putative binding regions on known protein structures and to restrict the docking search to just these regions [78]. Secondly, it could also be used to design a scoring function based on comparison of the predicted interface with interfaces in known complexes.

Indeed, the possibility to predict putative protein-binding sites (often called hotspots) on protein surfaces has been quite intensively investigated in recent years (reviewed in Moreira et al. [89]). Many methods to discover binding hotspots on protein surfaces through biochemical information, sequence

conservation or physicochemical properties have been developed [10, 90–92]. The inclusion of such information as bias during docking or to evaluate final results could be a promising route to achieve realistic predictions.

It has also been suggested that the total number of interaction types is limited (proteins sharing similar sequences tend to interact similarly). Aloy et al. [93] have estimated that most interactions in nature may conform to 1 of about 10 000 types similar to the 1000 natural protein folds that have been suggested by Chothia [94]. Such a limited set of specific protein-protein interfaces would allow the prediction of how two proteins may interact based on "homology" to a known interface type.

Conclusions

The comparative analysis of specific and nonspecific protein-protein binding interfaces (crystal-packing contacts) indicates several parameters that help to distinguish specific from nonspecific contacts. On average, the size of the buried interface differs between specific and nonspecific complexes; however, since the size distribution is for all types relatively broad, a clear classification based only on interface size is not possible. Similarly, the shape does not differ significantly between specific and nonspecific crystal-packing interfaces. However, there appears to be a clear lower limit for the size of an interface that is sufficient for specific protein-protein recognition, which is ~1000 Å^2. Specific and nonspecific interfaces differ significantly in terms of surface complementarity. Nonspecific protein crystal contacts contain many cavities that are also often water filled and the interfaces are not as well packed compared to specific protein-protein binding interfaces. It should be kept in mind that weak protein-protein contacts similar to the low-affinity contacts between proteins that are formed during crystallization at high protein concentration may in some cases have specific biological roles. Because of the transient nature many signal transduction events may involve only weak protein-protein interactions. Also, since cells are densely filled with proteins and other biomolecules even weak interactions may influence the spatial distribution of proteins.

Most specific protein binding sites have a defined structural organization consisting of a more hydrophobic central part and a hydrophilic rim region, whereas the distribution of hydrophobic and hydrophilic residues at crystal-packing interfaces is more random. The comparative analysis of protein-protein interfaces indicates that, despite the great variability of interfaces, a combination of interface properties like surface complementarity, packing and distribution of hydrophobic and hydrophilic interface residues allows for the classification of specific and nonspecific interfaces purely based on rules extracted from known structures. It is also possible to derive empirical rules for how a specific protein-protein interface has to look, in terms of limits for the size, packing, surface complementarity and amino acid composition. These rules are very valuable for understanding and evaluating specific recognition but also for the design or redesign of new specific protein-protein interactions. For structural biologists, the analysis can help to distinguish specific interactions observed in protein crystals from interactions due to crystal packing.

As also discussed, this knowledge can be very helpful for a better prediction of putative protein-binding sites on protein surfaces and to improve the realistic docking prediction of protein-protein complex geometries based on the structure of the individual partner proteins.

Acknowledgements. We thank the Deutsche Forschungsgemeinschaft (DFG) for financial support through grant Za153-11 to M.Z.

1 Janin, J. and Seraphin, B. (2003) Genome-wide studies of protein-protein interaction. Curr. Opin. Struct. Biol. 13, 383–388.
2 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. Nucleic Acids Res. 28, 235–242.
3 Chothia, C. and Janin, J. (1975) Principles of protein-protein recognition. Nature 256, 705–708.
4 Argos, P. (1988) An investigation of protein subunits and domain interfaces. Protein Eng. 2, 101–113.
5 Janin, J. and Chothia, C. (1990) The structure of protein-protein recognition sites. J. Biol. Chem. 265, 16027–16030.
6 Janin, J. (1995) Principles of protein-protein recognition from structure to thermodynamics. Biochimie 77, 497–505.
7 Janin, J. (1996) Quantifying biological specificity: The statistical mechanics of molecular recognition. Proteins 25, 438–445.
8 Jones, S. and Thornton, J. M. (1995) Protein-protein interactions: A survey of protein dimer structures. Prog. Biophys. Mol. Biol. 63, 31–65.
9 Jones, S. and Thornton, J. M. (1996) Principles of protein-protein interactions. Proc. Natl. Acad. Sci. USA 93, 13–20.
10 Jones, S. and Thornton, J. M. (1997) Analysis of protein-protein interaction sites using surface patches. J. Mol. Biol. 272, 121–132.
11 Tsai, C. J., Lin, S. L., Wolfson, H. J. and Nussinov, R. (1997) Study of protein-protein interfaces: a statistical analysis of the hydrophobic effect. Protein Sci. 6, 53–64.
12 Lo Conte, L., Chothia, C. and Janin, J. (1999) The atomic structure of protein-protein recognition sites. J. Mol. Biol. 285, 2177–2198.
13 Chakrabarti, P. and Janin, J. (2002) Dissecting protein-protein recognition sites. Proteins 47, 334–343.
14 Wodak, S. J. and Janin, J. (2003) Structural basis for molecular recognition. Adv. Prot. Chem. 61, 9–73.
15 Bahadur, R. P., Chakrabarti, P., Rodier, F. and Janin, J. (2003) Dissecting subunit interfaces in homodimeric proteins. Proteins 53, 708–719.

16 Bahadur, R. P., Chakrabarti, P., Rodier, F. and Janin, J. (2004) A dissection of specific and non-specific protein-protein interfaces. J. Mol. Biol. 336, 943–955.
17 Bahadur, R. P., Rodier, F. and Janin, J. (2007) A dissection of the protein–protein interfaces in icosahedral virus capsids J. Mol. Biol. 367, 574–590.
18 Janin, J., Rodier, F., Chakrabarti, P. and Bahadur, R. P. (2007). Macromolecular recognition in the Protein Data Bank. Acta Crystallogr. D, 63, 1–8.
19 Janin, J. and Rodier, F. (1995) Protein-protein interaction at crystal contacts. Proteins 23, 580–587.
20 Nooren, I. M. A. and Thornton, J. M. (2003) Diversity of protein-protein interactions. EMBO J. 22, 3486–3482.
21 Janin, J., Miller, S. and Chothia, C. (1988) Surface, subunit interfaces and interior of oligomeric proteins. J. Mol. Biol. 204, 155–164.
22 Miller, S., Janin, J., Lesk, A. M. and Chothia, C. (1987) Interior and surface of monomeric proteins. J. Mol. Biol. 196, 641–656.
23 Larsen, T. A., Olson, A. J. and Goodsell, D. S. (1998) Morphology of protein-protein interfaces. Structure 6, 421–427.
24 Goodsell, D. S. and Olson, A. J. (2000) Structural symmetry and protein function. Annu. Rev. Biophys. Biomol. Struct. 29, 105–153.
25 Nooren, I. M. A. and Thornton, J. M. (2003) Structural characterisation and functional significance of transient protein-protein interactions. J. Mol. Biol. 325, 991–1018.
26 Ofran, Y. and Rost, B. (2003) Analysing six types of protein-protein interfaces. J. Mol. Biol. 325, 377–387.
27 Miller, S., Lesk, A. M., Janin, J. and Chothia, C. (1987) The accessible surface area and stability of oligomeric proteins. Nature 328, 834–836.
28 Miller, S. (1989) The structure of interfaces between subunits of dimeric and tetrameric proteins. Protein Eng. 3, 77–83.
29 Brinda, K. V., Kannan, N. and Vishveshwara, S. (2002) Analysis of homodimeric protein interfaces by graph-spectral methods. Protein Eng. 15, 265–277.
30 Brinda, K. V. and Vishveshwara, S. (2005) Oligomeric protein structure networks: Insights into protein-protein interactions. BMC Bioinformatics 6, 296–310.
31 Tsuchiya, Y., Kinoshita, K., and Nakamura, H. (2006) Analyses of homo-oligomer interfaces of proteins from the complementarity of molecular surface, electrostatic potential and hydrophobicity. Protein Eng. Des. Sel. 9, 421–429.
32 Janin, J. (1997) Specific *vs* non-specific contacts in protein crystals. Nat. Struct. Biol. 4, 973–974.
33 Carugo, O. and Argos, P. (1997) Protein-protein crystal-packing contacts. Protein Sci. 6, 2261–2263.
34 Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. and Bell, J. A. (1997) Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. Proteins 28, 494–514.
35 Ponstingl, H., Henrick, K. and Thornton, J. M. (2000) Discriminating between homodimeric and monomeric proteins in the crystalline state. Proteins 41, 47–57.
36 Valdar, S. J. and Thornton, J. M. (2001) Conservation helps to identify biologically relevant crystal contacts. J. Mol. Biol. 313, 399–416.
37 Saha, R. P., Bahadur, R. P. and Chakrabarti, P. (2005) Inter-residue contacts in proteins and protein-protein interfaces and their use in characterizing the homodimeric interface. J. Proteome Res. 4, 1600–1609.
38 Liu, S., Li, Q., and Lai, L. (2006) A combinatorial score to distinguish biological and nonbiological protein-protein interfaces. Proteins 64, 68–78.
39 Zhu, H., Domingues, F. S., Sommer, I., and Lengauer, T. (2006) NOXclass: prediction of protein-protein interaction types. BMC Bioinformatics 7, 364–374.
40 Pal, A., Chakrabarti, P., Bahadur, R. P., Rodier, F. and Janin, J. (2007). Peptide segments in protein–protein interfaces. J. Biosci. 32, 101–111.
41 Lee, B. and Richards, F. M. (1971) The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55, 379–400.
42 Connolly, M. L. (1983) Solvent-accessible surfaces of proteins and nucleic acids. Science 221, 709–713.
43 Laskowski, R. A. (1995) SURFNET: A program for visualizing molecular surfaces, cavities, and intermolecular interactions. J. Mol. Graph. 13, 323–330.
44 Gose, E., Johnsonbaugh, R. and Jost, S. (2002) Pattern recognition and image processing, Prentice Hall, New Delhi, pp. 199–226.
45 Sheinerman, F. B., Norel, R. and Honig, B. (2000) Electrostatic aspects of protein-protein interactions. Curr. Opin. Struct. Biol. 10, 153–159.
46 Schreiber, G. and Fersht, A. R. (1996) Rapid, electrostatically assisted association of proteins. Nat. Struct. Biol. 3, 427–431.
47 Bhat, T. N., Bentley, G. A., Boulot, G., Greene, M. I., Tello, D., Dall'Acqua, W., Souchon, H., Schwarz, F. P. and Mariuzza, R. A. (1994) Bound water molecules and conformational stabilization help mediate an antigen-antibody association. Proc. Natl. Acad. Sci. USA 91, 1089–1093.
48 Huang, K., Lu, W., Anderson, S., Laskowski, M. J. and James, M. N. G. (1995) Water molecules participate in proteinase-inhibitor interactions: Crystal structure of Leu18, Ala18, and Gly18 variants of turkey ovomucoid inhibitor third domain complexed with *Streptomyces griseus* proteinase B. Protein Sci. 4, 1985–1997.
49 Janin, J. (1999) Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. Structure 7, R277–279.
50 Papoian, G. A., Ulander, J. and Wolynes, P. G. (2003) Role of water-mediated interactions in protein-protein recognition landscapes. J. Am. Chem. Soc. 125, 9170–9178.
51 Mustata, G. and Briggs, J. M. (2004) Cluster analysis of water molecules in alanine racemase and their putative structural role. Protein Eng. 17, 223–234.
52 Rodier, R., Bahadur, R. P., Chakrabarti, P. and Janin, J. (2005) Hydration of protein-protein interfaces. Proteins 60, 36–45.
53 Levy, Y. and Onuchic, J. N. (2006) Water mediation in protein folding and molecular recognition. Annu. Rev. Biophys. Biomol. Struct. 35, 389–415.
54 Li, Z. and Lazaridis, T. (2007) Water at biomolecular binding interfaces. Phys. Chem. Chem. Phys. 9, 573–581.
55 Raschke, T. M. (2006) Water structure and interactions with protein surfaces. Curr. Opin. Struct. Biol. 16, 152–159.
56 Mihalek, I., Res, I. and Lichtarge, O. (2007) On itinerant water molecules and detectability of protein–protein interfaces through comparative analysis of homologues. J. Mol. Biol. 369, 584–595.
57 Teyra, J. and Pisabarro, M. T. (2007) Characterization of interfacial solvent in protein complexes and contribution of wet spots to the interface description. Proteins 67, 1087–1095.
58 Smith, G. R., Sternber, M. J. E. and Bates, P. A. (2005) The relationship between the flexibility of proteins and their conformational states on forming protein–protein complexes with an application to protein–protein docking. J. Mol. Biol. 347, 1077–1101.
59 Crosio, M. P., Janin, J. and Jullien, M. (1992) Crystal packing in six crystal forms of pancreatic ribonuclease. J. Mol. Biol. 228, 243–251.
60 Wodak, S. J. and Janin, J. (1978) Computer analysis of protein-protein interaction. J. Mol. Biol. 124, 323–342.
61 Cherfils, J., Duquerroy, S. and Janin, J. (1991) Protein-protein recognition analyzed by docking simulation. Proteins 11, 271–280.
62 Zacharias, M. (2003) Protein–protein docking with a reduced protein model accounting for side-chain flexibility. Protein Sci. 12, 1271–1282.
63 Connolly M. L. (1985) Molecular surface triangulation. J. Appl. Crystallogr. 18, 499–505.

64 Connolly, M. L. (1986) Shape complementarity at the hemoglobin alpha 1 beta 1 subunit interface. Biopolymers 25, 1229–1247.
65 Jiang, F. and Kim, S. H. (1991) "Soft docking": Matching of molecular surface cubes. J. Mol. Biol. 219, 79–102.
66 Shoichet, B. K. and Kuntz, I. D. (1991) Protein docking and complementarity. J. Mol. Biol. 221, 327–246.
67 Wang, H. (1991) Grid-search molecular accessible surface algorithm for solving the protein docking problem. J. Comp. Chem. 12, 746–750.
68 Shoichet, B. K., Bodian, D. L. and Kuntz, I. D. (1992) Molecular docking using shape descriptors. J. Comput. Chem. 13, 380–397.
69 Lawrence, M. C. and Colman, P. M. (1993) Shape complementarity at protein-protein interfaces. J. Mol. Biol. 234, 946–950.
70 Fischer, D., Norel, R., Wolfson, H. J. and Nussinov, R. (1993) Surface motifs by a computer vision technique: Searches, detection, and implications for protein-ligand recognition. Proteins 16, 278–292.
71 Norel, R., Fischer, D., Wolfson, H. J. and Nussinov, R. (1994) Molecular surface recognition by a computer vision-based technique. Protein Eng. 7, 39–46.
72 Katchalski-Katzir, E., Shariv, I., Eisenstein, M., Friesem, A. A., Aflalo, C. and Vakser, I. A. (1992) Molecular surface recognition: Determination of geometric fit between proteins and their ligands by correlation techniques. Proc. Natl. Acad. Sci. USA 89, 2195–2199.
73 Gabb, H. A., Jackson, R. M. and Sternberg, M. J. E. (1997) Modelling protein docking using shape complementarity, electrostatic and biochemical information. J. Mol. Biol. 272, 106–120.
74 Palma, P. N., Krippahl, L., Wampler, J. E., and Moura, J. J. G. (2000) Bigger: A new (soft) docking algorithm for predicting protein interactions. Proteins 39, 372–384.
75 Janin, J., Henrick, K., Moult, J., Eyck, L. T., Sternberg, M. J. E., Vajda, S., Vakser, I. and Wodak, S. J. (2003) CAPRI: A Critical Assessment of PRedicted Interactions. Proteins 52, 2–9.
76 Mendez, R., Leplae, R., De Maria, L. and Wodak, S. J. (2003) Assessment of blind predictions of protein-protein interactions: current status of docking methods. Proteins 52, 51–67.
77 Halperin, I., Ma, B., Wolfson, H. J. and Nussinov, R. (2002) Principles of docking: an overview of search algorithms and a guide to scoring functions. Proteins 47, 409–443.
78 Bonvin, A. M. (2006) Flexible protein-protein docking. Curr. Opin. Struct. Biol. 16, 194–200.
79 Smith, G. R. and Sternberg, M. J. E. (2002) Prediction of protein-protein interactions by docking methods. Curr. Opin. Struct. Biol. 12, 28–35.
80 Fernandez-Racio, J., Totrov, M. and Abagyan, R. (2003) ICM-DISCO docking by global energy optimization with fully flexible side-chains. Proteins 52, 113–117.
81 Gray, J. J., Moughon, S. E., Kortemme, T., Schueler-Furman, O., Misura, K. M. S., Morozov, A. V. and Baker, D. (2003) Protein-protein docking predictions or the CAPRI experiments. Proteins 52, 118–122.
82 Bastard, K., Prevost, C. and Zacharias, M. (2006) Accounting for loop flexibility during protein–protein docking. Proteins 62, 956–969.
83 May, A. and Zacharias, M. (2005) Accounting for global protein deformability during protein–protein and protein–ligand docking. Biochim. Biophys. Acta 1754, 225–231.
84 May, A. and Zacharias, M. (2007) Energy minimization in low-frequency normal modes to efficiently allow for global flexibility during systematic protein-protein docking. Proteins (in press).
85 May, A. and Zacharias, M. (2007) Protein-protein docking in CAPRI using ATTRACT to account for global and local flexibility. Proteins 69, 774–780.
86 Miyazawa, S. and Jernigan, R. L. (1999) An empirical energy potential with a reference state for protein fold and sequence recognition. Proteins 36, 357–369.
87 Chen, R. and Weng, Z. (2002) Docking unbound proteins using shape complementarity, desolvation, and electrostatics. Proteins 47, 281–294.
88 Moont, G., Gabb, H. A. and Sternberg, M. J. E. (1999) Use of pair potentials across protein interfaces in screening predicted docked complexes. Proteins 35, 364–373.
89 Moreira, I. S., Fernandes, P. A. and Ramos, M. J. (2007) Hot spots – A review of the protein-protein interface determinant amino-acid residues. Proteins 68, 803–812.
90 Bogan, A. A. and Thorn, K. S. (1998) Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280, 1–9.
91 Rajamani, D., Thiel, S., Vajda, S. and Camacho, C. J. (2004) Anchor residues in protein-protein interactions. Proc. Natl. Acad. Sci. USA 101, 11287–11292.
92 Glaser, F., Steinberg, D.M., Vakser, I. A. and Ben-Tal, N. (2001) Residue frequencies and pairing preferences at protein-protein interfaces. Proteins 43, 89–102.
93 Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., Fischer, S., Gavin, A. C., Bork, P., Superti–Furga, G., Serrano, L. and Russell, R. B. (2004) Structure-based assembly of protein complexes in yeast. Science 303, 2026–2029.
94 Chothia, C. (1992) One thousand families for the molecular biologist. Nature 357, 543–544.
95 Nakatsu, T., Kato, H. and Oda, J. (1998) Crystal structure of asparagine synthetase reveals a close evolutionary relationship to class II aminoacyl-tRNA synthetase. Nat. Struct. Biol. 5, 15–19.
96 Safo, M. K. and Abraham, D. J. (2001) The X-ray structure determination of bovine carbonmonoxy hemoglobin at 2.1 Å resolution and its relationship to the quaternary structures of other hemoglobin crystal forms. Protein Sci. 6, 1091–1099.
97 Bhat, T. N., Bentley, G. A., Fischmann, T. O., Boulot, G. and Poljak, R. J. (1990) Small rearrangements in structures of Fv and Fab fragments of antibody D1.3 on antigen binding. Nature 347, 483–485.
98 Strop, P., Sedlacek, J., Stys, J., Kaderabkova, Z., Blaha, I., Pavlickova, L., Pohl, J., Fabry, M., Kostka, V., Newman, M., Frazao, C., Shearer, A., Tickle, I. J. and Blundell, T. L. (1990) Engineering enzyme subsite specificity: Preparation, kinetic characterization, and X-ray analysis at 2.0-Å resolution of Val111Phe site-mutated calf chymosin. Biochemistry 29, 9863–9871.
99 Weston, S. A., Camble, R., Colls, J., Rosenbrock, G., Taylor, I., Egerton, M., Tucker, A. D., Tunnicliffe, A., Mistry, A., Mancia, F., de la Fortelle, E., Irwin, J., Bricogne, G. and Pauptit, R. A. (1998) Crystal structure of the anti-fungal target *N*-myristoyl transferase. Nat. Struct. Biol. 5, 213–221.
100 Rafferty, J. B., Somers, W. S., Saint-Girons, I. and Phillips, S. E. (1989) Three-dimensional crystal structures of *Escherichia coli* met repressor with and without corepressor. Nature 341, 705–710.
101 Ahmad, F. K., Engel, C. K. and Gilbert, G. P. (1998) Crystal structure of the BTB domain from PLZF. Proc. Natl. Acad. Sci. USA 95, 12123–12128.

To access this journal online:
http://www.birkhauser.ch/CMLS