Journal of Structural Biology 179 (2012) 347–358

Contents lists available at SciVerse ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Review

Protein interactions in 3D: From interface evolution to drug discovery

Christof Winter [a], Andreas Henschel [b], Anne Tuukkanen [a], Michael Schroeder [a,*]

[a] Biotechnology Center, Technische Universität Dresden, Tatzberg 47–51, 01307 Dresden, Germany
[b] Masdar Institute of Science and Technology, P.O. Box 54224, Abu Dhabi, United Arab Emirates

ARTICLE INFO

Article history:
Available online 1 May 2012

Keywords:
Structural protein interaction
Databases
Evolution
Drug repositioning
Protein interaction prediction

ABSTRACT

Over the past 10 years, much research has been dedicated to the understanding of protein interactions. Large-scale experiments to elucidate the global structure of protein interaction networks have been complemented by detailed studies of protein interaction interfaces. Understanding the evolution of interfaces allows one to identify convergently evolved interfaces which are evolutionary unrelated but share a few key residues and hence have common binding partners. Understanding interaction interfaces and their evolution is an important basis for pharmaceutical applications in drug discovery.

Here, we review the algorithms and databases on 3D protein interactions and discuss in detail applications in interface evolution, drug discovery, and interface prediction.

1. Introduction

Interactions between proteins are at the heart of virtually every cellular process. Structures of protein complexes from the Protein Data Bank (Berman et al., 2000) provide detailed insight into the protein interaction interfaces. They are a valuable resource to understand how and why proteins interact. Since the seminal work by Jones and Thornton (1996) on principles of protein–protein interactions, many data collections to study structural principles of protein–protein interactions have been presented. First and foremost, the question arises what is an interface? Different approaches exist to define an interface based on distances of residue pairs or buried surface area. More refined measures consider e.g. the role of water in the interface. Drilling down, it is an open problem how to define and identify key residues that act as hot spot. These are important, since they form the basis to understand how robust an interface is to mutations. On the one hand, interfaces can be highly specific with a single mutation interrupting an interaction. An example are small GTPases interacting with their specific effectors. On the other hand, there are interfaces which share nothing but a few key residues in the same orientation and still they bind the same interaction partners. An example are chymotrypsin and subtilisin, whose overall sequence and structure bear no resemblance – yet they share three residues in exactly the same orientation, which allows inhibition by numerous common inhibitors. This wide range from fragile to robust interfaces, poses the more general question on how unique faces (note, an interface has two faces) are binding their partners. Is it a common phenomenon that one face can bind evolutionary unrelated faces from other proteins? This is important since it also forms the basis for applications in drug discovery.

This review is organised as follows: First, we will briefly summarise the databases and approaches to define and characterise structural interfaces. Next, we elaborate how these resources shed light on the evolution of interfaces, on applications in drug discovery, and prediction of interactions. We conclude by discussing the promises and limitations of these approaches and by providing a summary of the main 3D interaction resources.

1.1. Interface definitions for structural protein–protein interactions

What is an interaction interface? In principle, the interface consists of those amino acids that participate in the binding and hence physically contribute to the adhesion of two proteins. Various methods have been proposed to define the interface. These methods can be divided into three groups based on distance criteria between residue pairs, on buried surface area and on refined measures considering for example water in the interface. One should bear in mind, however, that this is often not straightforward to measure: dynamics of side chains and backbones result in varying distances, and the absence of hydrogens in many structures makes hydrogen bridges difficult to identify. Therefore, all these measures can only be an approximation.

Considering distance as a measure, two residues from different proteins are defined to be interacting if

- the distance between two C_α or between two C_β atoms from each residues is less than 9.0 Å
- the distance between any of their atoms, one from each, is less than 5.0 Å

* Corresponding author. Fax: +49 351 463 400 61.
E-mail address: ms@biotec.tu-dresden.de (M. Schroeder).

1047-8477/$ - see front matter © 2012 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jsb.2012.04.009

- the distance between any of their atoms, one from each, is less than the sum of their corresponding van der Waals radii plus 0.5 Å.
  Although different, Tsai et al. (1996) have stated that the above definitions yield consistent results.

Alternatively, the interface can be defined as

- the buried surface, which is the difference in accessible surface area (ASA) between the bound and unbound forms of the two-

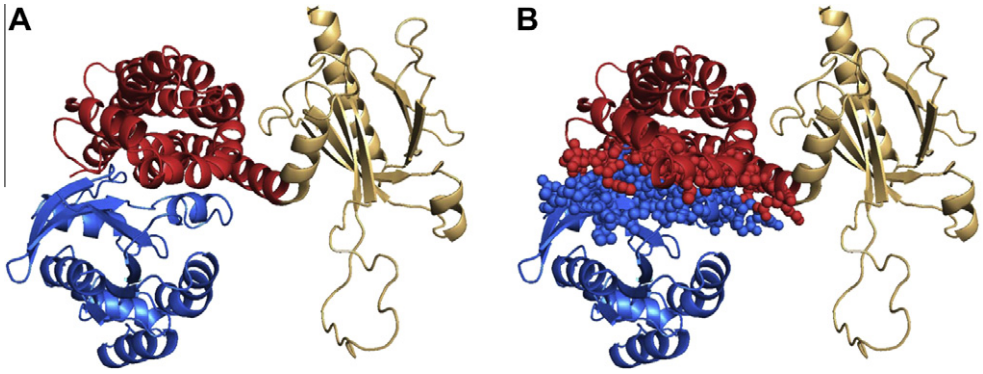

Fig.1. A structural protein–protein interaction. (A) Structure of the protein complex of a small GTPase (RAC1, blue) with a guanine nucleotide exchange factor (TIAM1, red/khaki). (B) The same complex with interface atoms between the two proteins as defined by the 5 Å rule (see text) highlighted as spheres. Note that only the DH domain (red) of TIAM1 contributes to the interface, whereas the membrane-anchoring PH domain (khaki) does not.

Table 1
Online resources for structural protein–protein interactions. The following databases were excluded from this overview as they were not online for a period of two months as of May 2011: PSIMAP (Park et al., 2001), BID (Fischer et al., 2003), iPfam (Finn et al., 2005), InterPare (Gong et al., 2005a), and SNAPPI (Jefferson et al., 2007). Interface definitions: 5–5, at least 5 residues within 5 Å distance; dASA, difference in accessible surface area; vdW, van der Waals.

Database	Domain definition	Number of interactions domain	Total	Interface definition	Interface clustering	Query options	Interface details	Visualisation	Download	Last update	References
3did	Pfam	6260	173,616	5–5	Yes	PDB ID, domain name, keyword, accession number, sequence, GO	Interaction topology, interface HMMs, experimental information	Network, Jmol	Flat file, MySQL	weekly	Stein et al. (2005)
PIBASE	SCOP, CATH, Chain	2387	264,015[a]	1–6 and dASA > 300 Å^2	Yes	PDB ID, keyword	Interacting residues (list), ASA	Rasmol	MySQL	9/2010	Davis and Sali (2005)
SCOPPI	SCOP	4630	105,547	5–5	Yes	PDB ID, domain name, keyword, accession number, GO	Conservation of interface residues, interface size, dASA, permanent/transient nature of interaction, interface classification, GO annotation	Interface images	Flat file upon request	6/2010	Winter et al. (2006)
SCOWLP	SCOP	5571	235,444	5–9, exclusion of crystal contacts[b]	Yes	PDB ID, domain ID or name, keyword	Dry and water mediated contacts, secondary structure	Interaction pattern, Jmol	MySQL	5/2011	Teyra et al. (2006)
PRISM	Chain	8205	49,512	sum of vdW radii plus 0.5 (Å)–6	Yes	PDB ID	dASA	Interface images, Chime		2008	Ogmen et al. (2005)
PSIBASE	SCOP	2387		5–5	No	PDB ID, keyword	–	Chime, Jmol	Flat file	9/2005	Gong et al. (2005b)
DOMINE	Pfam	6634		iPfam and 3did	No	Keyword, domain ID or name, GO term	–	–	MySQL	9/2010	Raghavachari et al. (2008)
ProtCID	Pfam		67,382	1–5 plus 10–12	Yes	PDB ID, domain ID, sequence	dASA	–	Flat file	4/2011	Xu and Dunbrack (2011)

[a] SCOP definition.
[b] Crystal contacts are excluded using NOXCLASS (Zhu et al., 2006).

protein complex. ASA was first described by Lee and Richards (1971).

Recently, two more refined methods were proposed.

- Fischer et al. (2006) have shown that a Voronoi polyhedra-based approach (Voronoi, 1908), which was first applied to proteins by Richards (1974), is more accurate than accessible surface or distance-based cutoff methods.
- Teyra and Pisabarro (2007) take water molecules at the interface into account. Starting with a 9 Å distance cutoff, they also consider an interaction when two residues are bridged through a water molecule. They have introduced the concept of the wet spots (Teyra and Pisabarro, 2007; Samsonov et al., 2008), which consist of residues that only interact through water.

Fig. 1 illustrates the interface of a protein complex as defined by the 5 Å rule (i.e. all residues of one protein which have at least one atom in less than 5.0 Å distance to an atom of another protein are considered part of the interface).

According to the current view of protein interaction, only a small subset of all interface residues is actually essential for recognition or binding. These essential residues are defined as residues that abolish protein–protein interactions if mutated, and are commonly referred to as hot spots. They are structurally conserved areas on the protein surface, often rich in tryptophan, tyrosine and arginine residues. Such hot spots have been found to contribute significantly in the binding free energy (Bogan and Thorn, 1998; Kortemme and Baker, 2002; Res and Lichtarge, 2005). It has been postulated that only very few of the residues in protein–protein interfaces are absolutely essential for the interaction and less than 5% of interface residues contribute more than 2 kcal/mol to binding free energy (Bogan and Thorn, 1998). The core hot spot residues are surrounded by a ring of energetically unimportant residues which occludes bulk solvent.

1.2. Databases for structural protein–protein interactions

The above definitions of interfaces led to various databases providing access to all structural protein interactions in the PDB such as 3did, iPfam, PIBASE, SCOPPI,[1] SCOWLP, PRISM, PSIBASE, DOMINE and PSIMAP (Stein et al., 2005; Finn et al., 2005; Davis and Sali, 2005; Winter et al., 2006; Teyra et al., 2006; Ogmen et al., 2005; Gong et al., 2005b; Raghavachari et al., 2008; Park et al., 2001). These databases are usually based on domain–domain interactions. Domain definitions are taken from SCOP, the structural classification of proteins (Murzin et al., 1995), from CATH (Orengo et al., 1997), from Pfam (Bateman et al., 2004), or from the conserved domain database, CDD (Marchler-Bauer et al., 2005). Most of the databases use distance-based cutoff methods to define the interface residues of the domain–domain interface. Query options include PDB ID, domain name, keywords, Gene Ontology terms, and protein sequence. Table 1 gives an overview about currently online databases and their content.

2. Application examples

What did we learn so far from the data collected in these structural interaction databases? Besides insight into general principles on the geometry and physico-chemical properties of protein interaction interfaces, application in three areas of research has led to interesting discoveries. First, a better understanding of how interfaces evolve. Second, providing a framework for drug repositioning, i.e. finding new targets for known drugs. Third, development of tools and algorithms to predict protein interactions based on structural knowledge. The next three sections summarise key results in these three areas.

2.1. Understanding the evolution of interfaces

2.1.1. Minimal binding consensus versus interface specificity

Interaction interfaces evolve. How much can an interface change until there is no more interaction? One would expect that small changes in interface residues have small impact on the interaction, and big changes have big impact (Fig. 2). There are, however, notable exceptions:

On the one hand, small changes in the interface can have large impact on the interaction behaviour. Small GTPases and their effectors (such as GTPase activating and guanine nucleotide exchanging proteins) are good examples for forming specific interactions (Hall, 2000). Here, one or two changes at key positions of the interaction surface can prevent binding, for example of Ras, Rap, or Raf to an effector Ras binding domain (Wohlgemuth et al., 2005; Kiel et al., 2005). Another example is that of cyclin-dependent kinases (CDK) 2 and 7. CDK2 was demonstrated to bind to Cyclin-dependent kinase inhibitor (CDKN) 3 in a co-crystallization study (Song et al., 2001). CDK7, although structurally almost identical to CDK2 (RMSD of the structural alignment of 1.1 Å), was found to not interact with CDKN3 (Lolli et al., 2004).

On the other hand, there are interfaces that require just a few key residues at the right position. For example, chymotrypsin and subtilisin have both a catalytic triad of the three residues histidine, aspartate, and serine in exactly the same conformation in common and can both bind common substrates (Pattabiraman and Lawson, 1972). Another prominent example for such a minimal binding consensus is the PxxP motif (two prolines with any two amino acids in between), which is bound by SH3 domains (Ren et al., 1993). More generally, Davey et al. (2011) summarise viral mimicry of native protein interaction interfaces. All of these examples lead to the notion of a minimal binding consensus. It

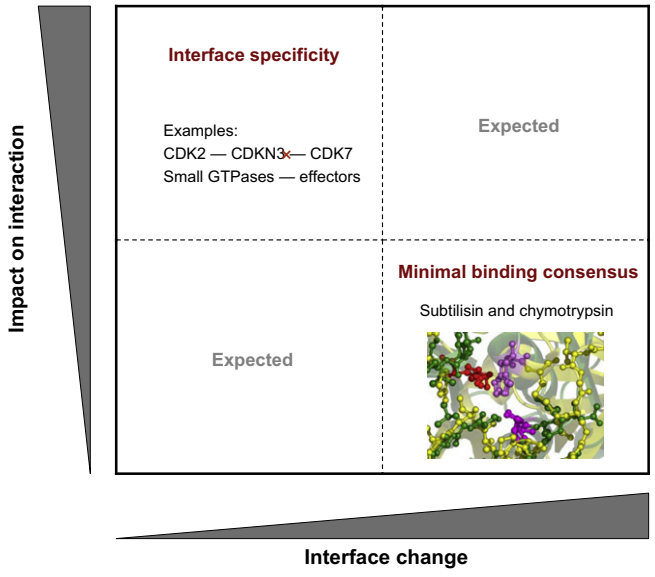

Fig.2. Impact of changes in the binding interface on the interaction. In general, the larger the interface change, the bigger the expected impact on the interaction. There are, however, interesting cases where small interface changes account for considerable difference in interaction behaviour (upper left quadrant), and cases where big interface changes that retain only a minimal consensus set of binding residues still allow binding of the same interaction partner. Examples are discussed in the text.

[1] Please note that SCOPPI is developed by the authors.

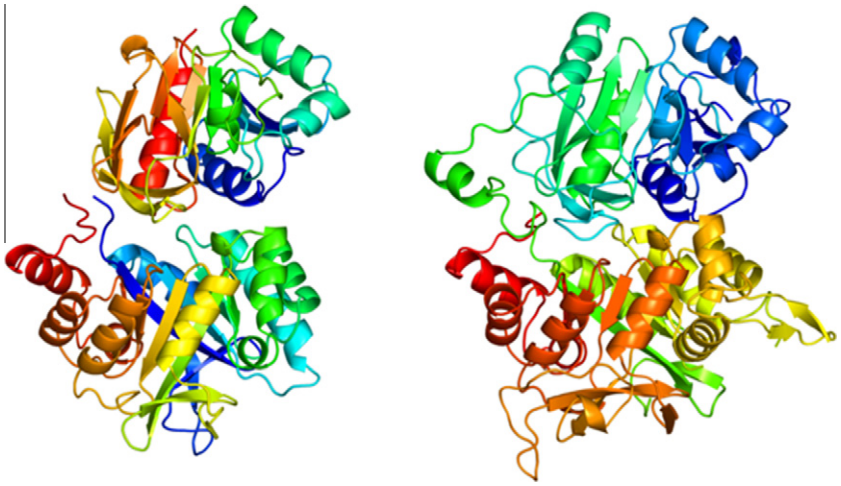

Fig.3. Binding site conservation after gene fusion. Imidazole glycerolphosphate synthase catalyses the formation of the imidazole ring in histidine biosynthesis. The structure of this enzyme in *Thermotoga maritima*, a hyperthermophile bacterium (left), and in *Saccharomyces cerevisiae* (right) is shown. The functional enzyme consists of a glutamine amidotransferase domain (top), and a cyclase domain (bottom). In *T. maritima*, these domains are located on two separate polypeptide chains, forming a heterodimeric protein complex. In *S. cerevisiae*, the two domains are fused together to one polypeptide chain, while preserving the geometry of the association.

should be noted that minimal binding consensus examples are not the result of changing interfaces starting from homologous interactions, but instead are often the result of convergent evolution, which will be discussed below.

Both extremes, high binding specificity and the minimal binding consensus pose difficult computational problems. Specific binding implies that prediction of protein interactions is difficult since it cannot always be derived from global sequence or structure properties. The minimal binding consensus implies that computational methods have to identify key residues to implement binding site comparison for remotely similar interfaces. Next, we summarise general results on the evolution of interfaces.

2.1.2. Evolutionary insights

Structural data on protein interaction can shed light on the evolution of protein interfaces. Using the SCOPPI database as a starting point, Kim et al. (2006b) classified structural interfaces according to their geometry. Their analysis revealed that 40% of protein interactions between homologues associate in multiple orientations. This allowed studying gene fusion events detected by conventional sequence homology: in one-third of these cases the fused and non-fused proteins associate in alternative binding orientations. Fig. 3 shows an example of gene fusion where the binding orientation of the fused and non-fused protein is conserved. The interface classification of Kim et al. (2006b) further revealed that any evolutionary analysis such as interface conservation is potentially skewed by multiple binding orientations and interaction partners. Finally, assuming that an interface is ancient if it is common to all three kingdoms (archea, bacteria, and eukaryotes), the taxonomic distribution of different interface classes suggests that ancient interactions are symmetric homodimers, and that asymmetric homodimers as well as heterodimers occurred later during evolution. Indeed, heterodimers are most frequently observed in eukaryotes, suggesting that their generation happened after kingdom separation.

Kim et al. (2006a) combined structural information with protein interaction networks, resulting in a structural interaction network, SIN. Relating network topology to three-dimensional structure and the number of distinct binding interfaces, they were able to subdivide hubs in the network into "singlish-interface" and "multi-interface" hubs. The former use the one binding site for several mutually exclusive interactions, whereas the latter have several binding sites for simultaneously possible interactions. These subclasses show different evolutionary properties, such as greater conservation of multi-interface hubs.

2.1.3. Convergent evolution

The analyses by Kim et al. (2006a,b) show that sometimes binding sites may be re-used by different binding partners. The faces of these different partners can hence be said to have converged in evolution to bind the common partner. In general, convergent evolution of interfaces can be observed when two unrelated proteins that do not share a common ancestor exhibit a remarkably similar interface structure. As noted above, a classic example of convergent evolution is that of the serine proteases chymotrypsin, a pancreatic digestive enzyme found in vertebrates, and subtilisin, found in the prokaryote *Bacillus subtilis*. Although the tertiary structures of the two enzymes bear no resemblance, the active site residues are in almost exactly the same position (Wright et al., 1969). Both proteins have independently evolved to carry out the same enzymatic function.

Convergent evolution can be found in particular in viruses mimicking native interfaces (Davey et al., 2011). With comprehensive structural protein interaction databases, it became possible to systematically search for examples of convergent evolution of protein–protein interfaces. Henschel et al. (2006) screened the Protein Data Bank for pairs of protein–protein interactions A–B and A–C where B and C are unrelated (different SCOP superfamilies) and bind to equivalent sites of A. This was named the ABAC[2] method. One ABAC example from Henschel et al. (2006) is shown in Fig. 4. The crystal structures of the complexes of caspase with the p35 protein from baculoviruses (Xu et al., 2001) and of caspase with XIAP, an endogenous inhibitor of the IAP family (Riedl et al., 2001) were solved by two independent groups. Both p35 and XIAP act as apoptotic suppressors. Aligning the two complexes using the common partner caspase reveals that both p35 and XIAP bind to the same surface patch of the caspase, which involves the active site of the caspase. Fig. 4 shows the remarkable structural similarity of the interacting motifs of p35 (red) and XIAP (blue), while the rest of the structures adopts a completely different fold. Interestingly, the interacting motifs are also different on the sequence level, with only one aspartate residue being identical. This aspartate residue (marked with an arrowhead in Fig. 4) is contacting a well-conserved potential hot spot on the caspase surface consisting of two tryptophane and

[2] Please note that ABAC is developed by the authors.

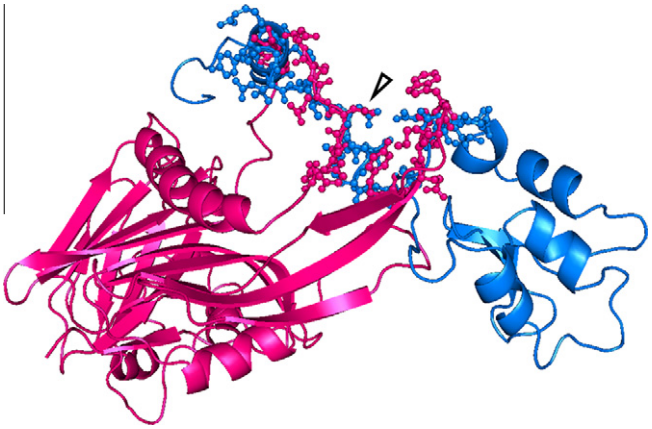

Fig.4. Convergent evolution of interfaces. Baculovirus protein p53 (red) mimics binding motif of the human inhibitor of apoptosis (blue). Both structures are aligned using their common partner, the catalytic domain of caspase (not shown). Interface residues are shown in balls and sticks. The arrowhead marks a prominent and potentially critical aspartate residue in almost identical position which makes contact to a hot spot region of the caspase interface.

one arginine residues. In total, 914 such examples were identified with the ABAC method (Henschel et al., 2006). Most of them concern enzymes.

2.1.4. How complete is the structural interface space?

The above analyses rely on the availability of structural data. As of May 2011, the Protein Data Bank (PDB) contains 73,000 structures. Since 2005 there is a linear trend in growth of PDB, with 7500 new structures each year. If the growth continues at this pace (ignoring the fact that many remaining proteins such as membrane proteins will be hard to tackle by X-ray crystallography) it will take at least 20 years until the human proteome is covered structurally (5000 human protein structures known today, a total of 17,000 proteins, and a current annual growth of 600 structures, counting only non-redundant proteins at maximum 50% sequence identity). To obtain structures of all complexes of the human interactome would take much longer. Even with a conservative estimate of the human interactome of 150,000 (Hart et al., 2006), the current coverage of 2800 protein complexes with the current annual growth of 300 complexes (again non-redundant at 50% sequence identity) would account for almost 500 years until the human interactome is complete in terms of structure.

However, we might be much closer to a complete picture of the structural interface space than this. Aloy and Russell (2004) suggested that there are only 10,000 interaction types. Recently, Gao and Skolnick (2010) claimed that the interface space is saturated and even close to complete. This assumption is based on the observation that even without structural similarity between individual monomers that form dimeric complexes, 90% of the native dimer interfaces have a close structural neighbor with similar geometry and interfacial contact pattern. The relatively flat interacting surfaces account for highly degeneration in geometry of the interfaces. According to Gao and Skolnick (2010), only about 1000 distinct interface types exist, and the library of protein interfaces is close to complete. Yet, they also agree that an experimentally determined database of all representative quaternary structures is not likely in the near future.

2.2. Drug discovery

Convergent evolution and the minimal binding consensus are important beyond our understanding of evolution: The principle can be exploited in drug discovery.

2.2.1. Protein–protein interfaces as drug targets

Targeting the interfaces between proteins with small molecule drugs has enormous therapeutic potential, given that many aberrant signaling protein interactions exist for example in cancer (Wells and McClendon, 2007). Yet, it is challenging. Protein–protein interfaces are often flat, have their binding energy delocalized over the whole surface, and often lack the grooves and pockets present at the surfaces of small-molecule binding sites (Hopkins and Groom, 2002). Protein interactions, at least non-enzymatic ones, were hence long thought not to be druggable. Two findings, however, challenge this view. First, the existence of hot spots in binding interfaces means that it is sufficient to target a small subset of the interface in order to considerably diminish the binding energy. Second, although an interface may appear flat in a snapshot view obtained by X-ray crystallography, molecular dynamics simulations reveal that small pockets open up for short time, which are druggable (Eyrisch and Helms, 2007, 2009).

High-throughput screening involving small-molecule libraries, often done in company research environments, could identify several compounds, some of which are already in clinical trials, that successfully target protein–protein interfaces (Wells and McClendon, 2007). One example is shown in Fig. 5A. Tumour-necrosis factor (TNF) is important for inflammatory response, and drugs targeting it are used for treating arthritis. Using fragment screening (He et al., 2005), a class of small molecules was recently discovered that disrupt TNF by binding and displacing one of the three monomers that constitute the TNF trimer (Fig. 5A). Although these small molecules are not seriously considered as drug candidates because of their moderate binding affinities, this finding clearly demonstrates that interfaces in oligomeric proteins are able to bind small molecules. Also, they provide a starting point for developing similar molecules with higher affinities.

Another approach is the combination of high throughput virtual and experimental screening, as demonstrated by Betzi et al. (2007). This method identified several compounds that inhibit the interaction between the Nef protein from human immunodeficiency virus type 1 (HIV-1) and host SH3 domain protein (Fig. 5B). The interface between Nef and the SH3 domain is formed by a proline-rich region and a so-called RT loop binding region (dashed line in Fig. 5B, left). A library of 1420 compounds was docked to this RT loop binding region, and ten candidates with promising chemical and geometrical properties were selected for experimental evaluation in a two-hybrid assay. One compound showed effect on the Nef-SH3 interaction, which was also validated in pull-down experiments. The core structure of this first inhibitor was then used as a template to screen a 435,000 compound database for similar compounds. Subjecting the 70 most similar compounds to two-hybrid assays followed by pull-down experiments found a second potent inhibitor of the Nef-SH3 interaction. NMR experiments confirmed binding of this second inhibitor to Nef (Fig. 5B, right). Notably, persons infected with HIV-1 strains that have deletions of the Nef gene develop AIDS symptoms much more slowly than those infected with standard HIV strains, making Nef a valuable target for pharmaceutical intervention.

2.3. Structure-based protein interaction prediction

As discussed above, structural interactions shed light on interface evolution and applications in drug development. These applications rely heavily on the availability of data. While Gao and Skolnick (2010) argue that the interface space is saturated, it remains an open problem how to predict protein interactions. Since protein complex structures provide a molecular view on how proteins interact, knowledge inferred from structural protein interactions is helpful to predict novel protein interactions.

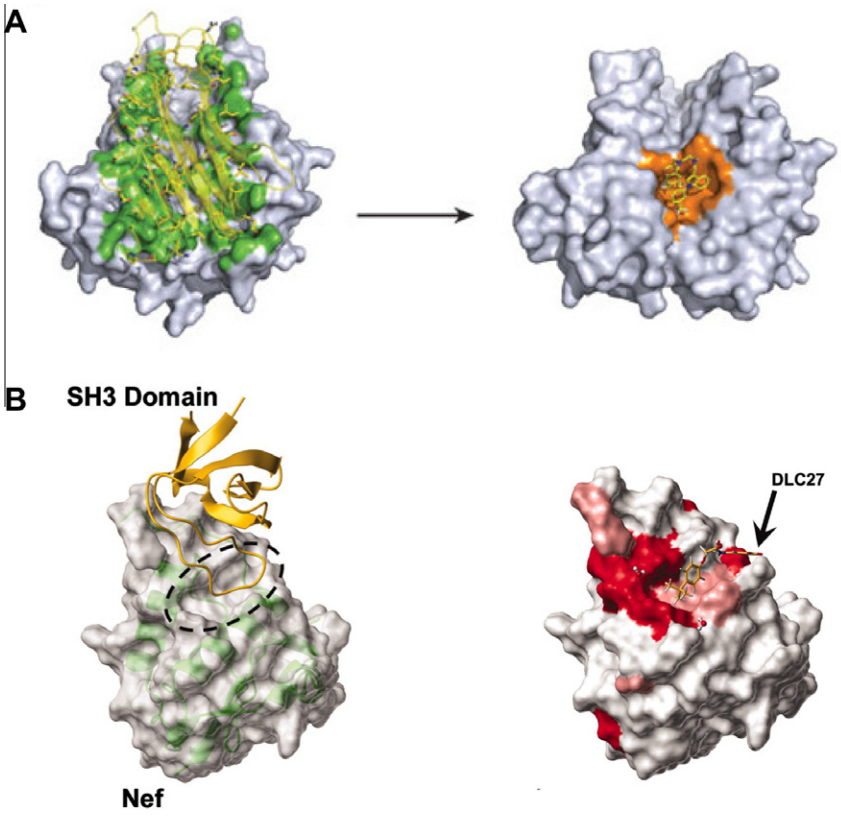

Fig.5. Drugs designed to target protein–protein interactions. (A) Left: The structure of tumour-necrosis factor TNF, which is composed of three monomers, is shown on the left. Two of the TNF monomers are shown in surface representation (grey), and the third monomer is shown in ribbon representation (yellow). The contact surface on the TNF dimer is shown in green. Right: The structure of the TNF dimer in complex with the small molecule SP304, which disrupts the TNF trimer. SP304 is shown in stick representation, the contact surface on the TNF dimer is colored in orange. Image from Wells and McClendon (2007). (B) Left: Structure of the HIV protein Nef binding to host SH3 domains. The interface exhibits a RT loop binding region with a hydrophobic groove (dashed line). This region was selected for virtual and experimental screening for binding compounds. Right: The small molecule compound DLC27 identified binds to the RT loop binding region and is a potent inhibitor of the Nef-SH3 interaction. Image from Betzi et al. (2007).

2.3.1. Homology-based approaches

Homology based transfer of interaction information through interologs has been attempted previously on sequence level (Walhout et al., 2000; Yu et al., 2004). The idea is that two proteins are assumed to be interacting if they have homologs in other species that are known to be interacting. Structure-based prediction methods often start by employing a search for protein complex structures that are homologous to the query sequences. These known complex structures are then used as templates to structurally model the interaction between query sequences. This method has the advantage of not only inferring protein interactions but also providing insights into how proteins interact on the atomic level of the interaction interface.

Aloy and Russell were the first to describe a method that uses complexes of known 3D structure to test for putative interactions between the homologues of the proteins contained in that complex (Aloy and Russell, 2002, 2003). Given a 3D complex and alignments of homologues of the interacting proteins, their method uses an empirical potential to assess the fit of any possible interacting pair on the complex. Using this approach combined with electron microscopy density maps, models were built for 54 protein complexes in yeast (Aloy et al., 2004). A method termed *multimeric threading* uses threading of both partners on a known complex was proposed Lu et al. (2002). Its application to the yeast proteome (Lu et al., 2003) was the first genome–scale interaction modelling study.

Several methods have been proposed to predict interfaces for a given protein. ProMate (Neuvirth et al., 2004), an interface prediction program, uses surface properties derived from protein–protein interactions for which the structures of both the unbound and bound states were known. Support vector machines were successfully employed by Koike and Takagi (2004), Bordner and Abagyan (2005), and Bradford and Westhead (2005), using patch hydrophobicity, conservation, electrostatic potential, solvation energy and residue accessible surface area as features. Aytuna et al. (2005) describe a method that combines structure and sequence conservation in protein interfaces. Its predictions are part of the PRISM database (Ogmen et al., 2005). Based on the PIBASE database (Davis and Sali, 2005), MODTIE (Davis et al., 2006, 2007) and HOMOLOBIND (Davis, 2011) predict binary protein interactions and higher-order protein complexes binding sites from a given set of protein sequences. Henschel et al. (2007) built Hidden–Markov-Model descriptors of interfaces derived from the SCOPPI database that were used to predict protein and ligand binding sites on protein sequences. Dawelbait et al. (2007) used structural interaction templates found by threading followed by homology modeling and energy minimization.

In the past years, also several web servers for structure-based interaction prediction have emerged, such as 3D-partner (Chen et al., 2007), HOMCOS (Fukuhara and Kawabata, 2008), Protinfo PPC (Kittichotirat et al., 2009), and Struct2Net (Singh et al., 2010).

2.3.2. Docking-based approaches

Structure-based protein interaction prediction is closely related to protein–protein docking, which attempts to use geometric and steric considerations to fit two proteins of known structure into a

bound complex. Docking competitions such as CAPRI, the critical assessment of predicted interactions (Janin, 2010), show that enzyme–inhibitor interactions which are characterised by large and well conserved interfaces can be predicted with good success, but that the general problem of docking is far from being solved (Vajda and Camacho, 2004). In particular, docking has not been widely used to predict interaction partners of a protein. The main reason for this is that the list of models produced by current docking algorithms contains a large number of false positives (Janin, 2010). This makes it difficult to correctly rank them, and most often the correct complex is not the one with the highest docking score. It has therefore been regarded beyond the scope of current protein docking algorithms to detect interaction partners (Russell et al., 2004; Aloy and Russell, 2006). Recently, however, Wass et al. (2011) challenged this view. In a proof of principle high-throughput docking experiment, they demonstrated that a rigid-body docking algorithm can distinguish between interacting and non-interacting proteins. To this end, they generated models for known interacting proteins (the widely used Mintseris et al. (2005) benchmark set representing 56 true interactions) and compared the score distribution of these models with those of models generated from a large background set of different structures that are unlikely to interact with either of the partners (51,632 false interactions). The majority of the benchmark interactions had better scores than 85% of the background set interactions, and three benchmark interactions (among them one enzyme–inhibitor example) had a better score than all of the background set. Some benchmark interactions, however, were indistinguishable from the background. In particular, eight examples had a score distribution that was worse than 50% of the score distributions from the background set. Most proteases (particularly subtilisin) were among these examples, potentially due to the broad substrate specificity of such enzymes. Overall, a receiver operating characteristic analysis to assess the global ability of the method in separating the 56 true positive from the 51,632 true negative pairs showed an area under curve accuracy of 80%.

As it is known from small-molecule docking that consensus scoring based on multiple algorithms improves prediction results (Yang et al., 2005), a similar strategy could be applied to high-throughput protein–protein docking. One limitation is that this might increase the computing time beyond feasibility – Wass et al. (2011) had to use a supercomputing facility and generated over one billion complex models. The recent advent of cloud computing in the life sciences has the potential to help overcome this limitation. In a pilot study, the pharmaceutical company Pfizer outsourced docking experiments of antibodies to epitopes with RosettaDock (Gray et al., 2003) to a commercial cloud service provider. This reduced the time spent on computing of one antibody model from two to three months on their in-house computing cluster to merely one day in the cloud environment (Kraut, 2009).

2.3.3. Accuracy of structure-based interaction prediction

How accurate are such structure-based interaction prediction approaches? As described above, a problem with homology-based prediction is the fact that changes in key residues that are critical for interaction specificity can render an interaction impossible. A solution to this problem is to take into account the calculated interaction energy of the homology modelled complex. Prediction of protein interactions that use homology modelling and energy calculations have shown promising results: Kiel et al. (2007) have determined a genome-wide Ras-effector interaction network based on homology models, with a high accuracy of predicting binding and non-binding domains. For binding site predictions on the sequence level, ProMate achieves a 70% success rate in predicting interface location. HOMOLOBIND predicts residues in protein binding sites with an estimated true positive rate of 88% at a false positive rate of 1% (Davis, 2011).

3. Structural protein–protein interactions

The above insights into evolution, drug development, and interaction prediction are based on the knowledge of the molecular details of structural protein–protein interactions. 3D interaction databases provide such knowledge and have made such studies possible. The next two sections summarise these databases and the type of data they provide.

3.1. Online databases of structural protein–protein interactions

3.1.1. PSIMAP/PSIBASE

The first protein interaction family map was created by Park et al. (2001). Individual protein domain interactions where classified by domain assignments into family–family interactions using SCOP. Sequential interaction data was taken from MIPS, whereas structural information was used from the PDB. The scale-free shape of the family interaction graph could be demonstrated: the interaction map exposes few versatile families, whereas most families interact only with one, two or three other families. This family–family classification was extended in the PSIMAP and PSIBASE databases (Bolser et al., 2003; Dafas et al., 2004; Gong et al., 2005b).

3.1.2. iPfam

The iPfam database (Finn et al., 2005) was amongst the first interaction databases that aimed for regional information of the interface, not derived from computational prediction nor experimental studies but from multi-domain and multi-chain complexes in PDB. The database iPfam is, as the name suggests, built using the Pfam database of protein domains and families derived from multiple sequence alignments. The Pfam domain boundary definitions were mapped on the structures using the UniProt to PDB mapping (Golovin et al., 2004). iPfam is the most highly cited database on structural protein interactions.

3.1.3. 3did

Beside the collection of domain-domain interactions in proteins from high-resolution three-dimensional structures, 3did (Stein et al., 2005) offers an overview of how similar in structure are interactions between different members of the same protein family. It is based on domains defined by Pfam. Additionally, the database contains yeast interactions and Gene Ontology annotations. Recent add-ons include domain–peptide interactions based on linear motifs, and an interaction classification based on interface residues.

3.1.4. SCOPPI

SCOPPI (Winter et al., 2006) is a derivative of the approach taken in PSIMAP. It is also based on a rigorous family–family interaction classification, but additionally distinguishes interface types. It has been used for protein–protein interaction prediction method using the structural interolog method. SCOPPI was the first to cluster binding sites and to characterise these clusters.

3.1.5. SCOWLP

As described in the introduction, the SCOWLP database (Teyra et al., 2006) introduced the concept of wet spots – amino acids that are bridged by water molecules in the interface via hydrogen bonds. Besides protein–protein interfaces, SCOWLP also provides interfaces to DNA, RNA, peptides, and sugars. Protein complexes within a family are clustered at different binding similarity cutoffs

to provide an interface classification. The NOXCLASS SVM classifier (Zhu et al., 2006) is used to identify crystal contacts in order to exclude them.

3.1.6. PIBASE

In contrast to other structural databases, domain definitions in PIBASE (Davis and Sali, 2005) are taken from both SCOP and CATH. It provides a collection of structural and functional protein properties imported from various sources including MODBASE, SwissProt, GO, OMIM, and InterPro. PIBASE clusters interacting domains by binding site topology.

3.1.7. PRISM

The database of Protein Interactions by Structural Matching (PRISM, Ogmen et al. (2005)) is unique in the structural classification databases in that it does not employ structural domain definitions. Rather it is derived from a non-redundant dataset of interface structures. An interface is derived from a multi chain PDB structure where interacting residues are identified as those with a pair of atoms (each from opposite chains) with a distance smaller than the sum of their van der Waals radii plus a threshold of 0.5 Å. Additionally, scaffold residues, i.e. those with their C_α within 6 Å of an interface residue of the opposite chain are included. The 50,000 interfaces are clustered using geometric hashing. PRISM's interface types are taken from chain–chain interactions. It has been pointed out, however, that for the investigation of interaction types, domain–domain interactions are more appropriate: the protein repertoire uses a combination of domains to account for the large diversity of functions (Chothia et al., 2003). A domain as a structural and evolutionary unit provides a simpler view on binding sites than chains that are multi-domain proteins.

3.1.8. Conserved binding mode database (CBM)

Likewise, CBM (Shoemaker et al., 2006) is a domain interaction database, but additionally has information of the geometric association of the domains. It can furthermore assist in constructing topological networks from connections derived from a biological context. Finally, CBM categorises protein interaction surfaces and thereby allows homology modelling of interfaces. The domain definition is taken from NCBI conserved domain database (Marchler-Bauer et al., 2005).

3.1.9. DOMINE

DOMINE (Raghavachari et al., 2008) is a collection of known and predicted protein domain interactions, with domains defined by Pfam. Known interactions are collected from the iPfam and 3did databases. Predicted protein domain interactions are obtained by a variety of 13 different computational approaches.

3.1.10. SNAPPI-DB

SNAPPI-DB (Jefferson et al., 2007) is a feature rich domain–domain interaction database with domain definitions from SCOP, CATH and Pfam. It provides a programming API for convenient retrieval of domain–domain interactions. The clustering of interfaces has been done using the iRSMD method of Aloy et al. (2003). The iRMSD focuses on the orientation of the centers of mass of the interacting domains. The disadvantage of the iRMSD approach is that it is insensitive to the location of the actually interacting atoms. The interfaces are taken from the biological units from PQS rather than PDB's asymmetric units. SNAPPI-DB links to InterPro, Gene Ontology and SwissProt.

3.2. Properties of protein interaction interfaces

Since the seminal work by Chothia and Janin (1975), a number of studies have been carried out on the physico-chemical characterisation of protein interfaces. The discovery of the characteristics of each interface forms a basis to understand the rules of molecular recognition. General principles of protein–protein interactions have been proposed (Hubbard and Argos, 1994; Jones and Thornton, 1996; Valdar and Thornton, 2001; Gao et al., 2004; Bahadur et al., 2004). Many authors defined measures for interface size, shape complementarity, protrusion, segmentation and secondary structure. It is important to notice that in the vast diversity of protein interactions these characteristics differ strongly. Given the structure of two interaction candidate proteins, these measures both help to determine the binding ability as well as to identify the actual binding site. This knowledge driven approach has inspired many machine learning approaches. Further important aspects for protein interactions have been investigated, as will be discussed in the following.

3.2.1. Duration of interaction

A fundamental distinction between interactions is by their duration, that is whether they occur permanently or transiently. Nooren and Thornton (2003) further distinguish between weak and strong transient interactions, where the former follow a principle informally termed *kiss and run*, while the latter are in contact and their connection is disrupted through some trigger. A slightly deviating definition, regarding the duration as well as the functional aspect, divides protein–protein interactions into obligate and non-obligate. An obligate interaction is a permanent interaction between proteins which do not occur individually, but only as a complex. A prominent example are the alpha and beta chains of hemoglobin, which are only fully functional upon complex formation.

3.2.2. Evolutionary conservation

Structurally and functionally important residues are often well conserved. It is reasonable to assume that key residues of a vital protein–protein interaction are spared from mutations, as the loss of an interaction either has fatal consequences for the organism, or displays an evolutionary disadvantage. Consequently, functional interfaces are commonly associated with a lower mutability than other, non-functional parts of the protein surface. The identification of binding sites based on evolutionary conservation, however, remains controversial. While some authors claim a stronger conservation of interfaces than the rest of the surface (Nooren and Thornton, 2003; Bordner and Abagyan, 2005), others question the statistical significance (Caffrey et al., 2004). A remaining problem is how to clearly identify non-functional parts of the protein surface, as we are probably still far from complete knowledge about all interactions. Surface parts considered non-functional could actually be important for yet unknown interactions. It is hence not sufficient to base binding site detection solely on conservation scores. However, selective pressure often leads to the conservation of important protein features, such as binding behaviour. Therefore, it is generally beneficial to include evolutionary conservation scores, as for example employed by Li et al. (2004) as well as Bordner and Abagyan (2005).

3.2.3. Hydrophobicity

Surface patches with high hydrophobicity are energetically unfavourable in a watery solution, but favourable when in contact with other hydrophobic surfaces. Hence, their occurrence can be linked to binding sites. Gallet et al. (2000) proposed an interface detection method that predicts binding sites by analysing the hydrophobicity distribution in sequences. Especially permanent interactions between globular proteins were found to involve more hydrophobic residues. The largest hydrophobic surface patches were often found to participate in protein binding, at least to some extent. This approach has been outperformed by several

techniques that employed further features. According to Nooren and Thornton (2003), hydrophobicity has less discriminative power for transient interactions, as transient interactions are often established by hydrogen bonds from polar side chains.

3.2.4. Hot spots and hot regions

Hot spots are interface residues that dominantly contribute to the binding free energy. Their identification requires experimental approaches such as alanine scanning (Bogan and Thorn, 1998). Ma et al. (2003) found that structurally conserved residues distinguish between binding sites and exposed protein surfaces. Halperin et al. (2004) showed that hot spots are often observed to couple across two-chain interfaces. Darnell et al. (2007) describe knowledge-based models that allow predicting hot spots. Predictions are available online (Darnell et al., 2008).

3.2.5. Sequential patterns

Conserved sequence motifs in and around the interacting residues are indicators for interaction sites.

3.2.6. Secondary structure

Interface residues can be classified according to the secondary structure they are part of, that is alpha-helices, beta-strands, turns, and loops. Less common are 3_{10}-helices and π-helices. Secondary structure elements are commonly determined in protein structures using the DSSP algorithm (Kabsch and Sander, 1983), which is based on an electrostatic definition to identify hydrogen bonds. Further, less regular structural motifs exist (Milner-White et al., 2004). Preissner et al. (1998) investigated protein interfaces from 351 selected protein complex structures. They classify binding sites according to the participating secondary structure elements.

3.2.7. Electrostatics

Calculation of the electrostatic potential of protein–protein complexes revealed that protein–protein interfaces display electrostatic complementarity (McCoy et al., 1997). Electrostatic potentials of proteins surfaces can be calculated by solving the Poisson–Boltzmann equation (Fogolari et al., 2002). Frequently, binding sites are identified by opposite charges in opposite interfacing patches.

3.2.8. Hydrogen bonds

The bonds formed by hydrogen atoms between a hydrogen-donor and a hydrogen-acceptor are the strongest contributors to binding energies.

3.2.9. Water bridges

Bridging water molecules are those forming two hydrogen bonds, one to each interface side. A comprehensive analysis of so-called wet-spots can be found in the SCOWLP database by Teyra et al. (2006).

3.2.10. Buried surface

A probe sphere of radius 1.4 Å (the approximate size of water) is rolled around the van der Waals surface of a protein. The area is measured for the interacting proteins before and after complex formation. The difference of the two areas yields the size of the buried surface, and thus the size of an interface.

3.2.11. Shape complementarity

Interacting proteins often have large surface patches that are in direct contact with each other. Especially enzymes that select their substrates specifically exhibit a high degree of shape complementarity with respect to their substrates. This phenomenon was observed by Emil Fischer, who proposed the "lock and key" model in 1894 (Fischer, 1894). To account for conformational flexibility of protein–protein interactions, Koshland (1958) proposed the "induced fit theory" as a modification to the lock and key model. According to this theory, substrates do not rigidly dock to their enzymes, but constantly perform small rearrangements of their side chains. Shape complementarity is a purely structural feature and hence does not translate directly into sequence representations. Therefore, shape complementarity is not directly usable in predictions, where only sequences are given. However, shape complementarity can be the result of a long evolutionary process involving mutations on one side and compensatory mutations on the opposite side of the interaction. These correlated mutations are detectable on sequence level.

3.2.12. Intra/inter–chain interactions

Domain–domain interactions can be observed between different proteins or within the same protein.

3.2.13. Homo and hetero dimers

Protein dimers can be built from equal or different domains. Many proteins occur as homo dimers, with two identical domains binding to each other. The distinction between homo dimers and hetero dimers can be made on various levels, such as sequence level, family level, or superfamily level. These factors influence the formation of protein–protein complexes. Jones and Thornton (1996) suggest a rough classification into four different types of protein–protein complexes: homo-dimeric proteins, hetero-di-

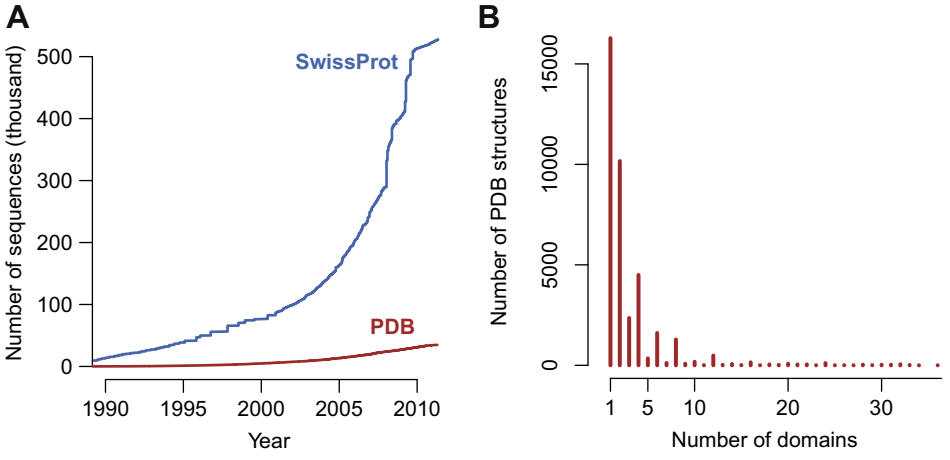

Fig.6. Limited availability of structural data. (A) Regarding the growth of the SwissProt database of known protein sequences, and the Protein Data Bank (PDB) of known protein structures, there is an increasing sequence–structure gap. (B) Many structures in PDB contain only one domain, with no information on interfaces.

meric proteins, enzyme–inhibitor complexes, and antibody–protein complexes. The comparison between the complexes highlights differences that reflect their biological roles. It will be of great interest to determine the most relevant characteristics for the particular purpose of binding site prediction. Elcock and McCammon (2001) found that interface conservation helps to correctly predict the oligomerization state of a protein based on crystal structures.

4. Limitations

While the overall achievements of the structural interaction approach are impressive and cover a wide area, there are two inherent limits to it:

- Data bias: Despite the growth of the Protein Data Bank (PDB), there is a huge gap between the amount of structural and sequence data (Fig. 6A). This gap is even wider if one considers that most of PDB consists of monomers (Fig. 6B), and that only some 27% of PDB are dimers and only some 7% are heterodimers. Additionally, the PDB is biased towards smaller complexes and disease relevant targets. Nonetheless, the argument can be made that the interface space is saturated (Gao and Skolnick, 2010).
- Specificity and minimal binding consensus: The presence or absence of a few key residues can determine an interaction or not. In highly specific interactions a few missing key residues interrupt the interaction and in convergently evolved interfaces only a few key residues are needed for the interaction. This implies that any algorithm to compare interfaces must be able to detect these key residues. Overall, there are some concrete successes to this end, but the general problem is challenging and remains open.

5. Conclusions and outlook

Over the past ten years, structural protein interactions have been studied in detail. Algorithms and large-scale databases were developed and so far proved useful to broaden our understanding of interface evolution and in drug discovery. A key challenge for the future is the use of protein interactions to model large complexes. Several problems arise: First, how to expand the limited data? One approach will integrate structural interactions with experimental high-throughput interaction data and literature curated protein interactions. Second, how to deal with uncertainty and error in the models of individual structures when compiling large complexes? This is more challenging than protein docking, where exact structural information of the unbound proteins is available, as one has to deal with structural uncertainty. We are in need of new methods to judge the quality of such models not only based on interface complementarity and strength, but also on the confidence in the predicted or inferred interface structure. Third, how to assemble individual interactions into larger complexes. First successes by Sali and co-workers in modelling the nuclear pore complex prove feasibility, but to date there are no systematic approaches how to assemble models of large complexes. Forth, how to use microscopy data to support the above modelling challenges. This will range from the use of electron microscopy maps as low resolution templates for the large complexes to the use of atomic force microcopy to study intra- and inter-molecular forces.

Acknowledgements

Funding by the EU (Ponte) and BMWi (GoOn, GeneCloud) is kindly acknowledged.

References

Aloy, P., Bottcher, B., Ceulemans, H., Leutwein, C., Mellwig, C., et al., 2004. Structure-based assembly of protein complexes in yeast. Science 303 (5666), 2026–2029.
Aloy, P., Ceulemans, H., Stark, A., Russell, R.B., 2003. The relationship between sequence and interaction divergence in proteins. J. Mol. Biol. 332 (5), 989–998.
Aloy, P., Russell, R.B., 2002. Interrogating protein interaction networks through structural biology. Proc. Natl. Acad. Sci. USA 99 (9), 5896–5901.
Aloy, P., Russell, R.B., 2003. InterPreTS: protein interaction prediction through tertiary structure. Bioinformatics 19 (1), 161–162.
Aloy, P., Russell, R.B., 2004. Ten thousand interactions for the molecular biologist. Nat. Biotechnol. 22 (10), 1317–1321.
Aloy, P., Russell, R.B., 2006. Structural systems biology: modelling protein interactions. Nat. Rev. Mol. Cell Biol. 7 (3), 188–197.
Aytuna, A.S., Gursoy, A., Keskin, O., 2005. Prediction of protein–protein interactions by combining structure and sequence conservation in protein interfaces. Bioinformatics 21 (12), 2850–2855.
Bahadur, R.P., Chakrabarti, P., Rodier, F., Janin, J., 2004. A dissection of specific and non-specific protein–protein interfaces. J. Mol. Biol. 336 (4), 943–955.
Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., et al., 2004. The Pfam protein families database. Nucleic Acids Res. 32 (database issue), D138–D141.
Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., et al., 2000. The Protein Data Bank. Nucleic Acids Res. 28 (1), 235–242.
Betzi, S., Restouin, A., Opi, S., Arold, S.T., Parrot, I., et al., 2007. Protein–protein interaction inhibition (2P2I) combining high throughput and virtual screening: application to the HIV-1 Nef protein. Proc. Natl. Acad. Sci. USA 104 (49), 19256–19261.
Bogan, A.A., Thorn, K.S., 1998. Anatomy of hot spots in protein interfaces. J. Mol. Biol. 280 (1), 1–9.
Bolser, D., Dafas, P., Harrington, R., Park, J., Schroeder, M., 2003. Visualisation and graph-theoretic analysis of a large-scale protein structural interactome network. BMC Bioinform. 4 (45).
Bordner, A.J., Abagyan, R., 2005. Statistical analysis and prediction of protein–protein interfaces. Proteins 60 (3), 353–366.
Bradford, J.R., Westhead, D.R., 2005. Improved prediction of protein–protein binding sites using a support vector machines approach. Bioinformatics 21 (8), 1487–1494.
Caffrey, D.R., Somaroo, S., Hughes, J.D., Mintseris, J., Huang, E.S., 2004. Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci. 13 (1), 190–202.
Chen, Y.-C., Lo, Y.-S., Hsu, W.-C., Yang, J.-M., 2007. 3D-partner: a web server to infer interacting partners and binding models. Nucleic Acids Res. 35 (web server issue), W561–W567.
Chothia, C., Gough, J., Vogel, C., Teichmann, S.A., 2003. Evolution of the protein repertoire. Science 300 (5626), 1701–1703.
Chothia, C., Janin, J., 1975. Principles of protein–protein recognition. Nature 256 (5520), 705–708.
Dafas, P., Bolser, D., Gomoluch, J., Park, J., Schroeder, M., 2004. Using convex hulls to extract interaction interfaces from known structures. Bioinformatics 20 (10), 1486–1490.
Darnell, S.J., LeGault, L., Mitchell, J.C., 2008. KFC Server: interactive forecasting of protein interaction hot spots. Nucleic Acids Res. 36 (web server issue), W265–269.
Darnell, S.J., Page, D., Mitchell, J.C., 2007. An automated decision-tree approach to predicting protein interaction hot spots. Proteins 68 (4), 813–823.
Davey, N.E., Trave, G., Gibson, T.J., 2011. How viruses hijack cell regulation. Trends Biochem. Sci. 36 (3), 159–169.
Davis, F., Sali, A., 2005. PIBASE: a comprehensive database of structurally defined protein interfaces. Bioinformatics 21 (9), 1901–1907.
Davis, F.P., 2011. Proteome-wide prediction of overlapping small molecule and protein binding sites using structure. Mol. Biosyst. 7 (2), 545–557.
Davis, F.P., Barkan, D.T., Eswar, N., McKerrow, J.H., Sali, A., 2007. Host pathogen protein interactions predicted by comparative modeling. Protein Sci. 16 (12), 2585–2596.
Davis, F.P., Braberg, H., Shen, M.-Y., Pieper, U., Sali, A., et al., 2006. Protein complex compositions predicted by structural similarity. Nucleic Acids Res. 34 (10), 2943–2952.
Dawelbait, G., Winter, C., Zhang, Y., Pilarsky, C., Grützmann, R., et al., 2007. Structural templates predict novel protein interactions and targets from pancreas tumour gene expression data. Bioinformatics 23 (13).
Elcock, A.H., McCammon, J.A., 2001. Identification of protein oligomerization states by analysis of interface conservation. Proc. Natl. Acad. Sci. USA 98 (6), 2990–2994.
Eyrisch, S., Helms, V., 2007. Transient pockets on protein surfaces involved in protein–protein interaction. J. Med. Chem. 50 (15), 3457–3464.
Eyrisch, S., Helms, V., 2009. What induces pocket openings on protein surface patches involved in protein–protein interactions? J. Comput. Aided Mol. Des. 23 (2), 73–86.
Finn, R.D., Marshall, M., Bateman, A., 2005. iPfam: visualization of protein–protein interactions in PDB at domain and amino acid resolutions. Bioinformatics 21 (3), 410–412.
Fischer, E., 1894. Einfluss der Configuration auf die Wirkung der Enzyme. Ber. Dtsch. Chem. Ges. 27, 2984–2993.

Fischer, T.B., Arunachalam, K.V., Bailey, D., Mangual, V., Bakhru, S., et al., 2003. The binding interface database (BID): a compilation of amino acid hot spots in protein interfaces. Bioinformatics 19 (11), 1453–1454.
Fischer, T.B., Holmes, J.B., Miller, I.R., Parsons, J.R., Tung, L., et al., 2006. Assessing methods for identifying pair-wise atomic contacts across binding interfaces. J. Struct. Biol. 153 (2), 103–112.
Fogolari, F., Brigo, A., Molinari, H., 2002. The Poisson–Boltzmann equation for biomolecular electrostatics: a tool for structural biology. J. Mol. Recognit. 15 (6), 377–392.
Fukuhara, N., Kawabata, T., 2008. HOMCOS: a server to predict interacting protein pairs and interacting sites by homology modeling of complex structures. Nucleic Acids Res. 36 (web server issue), W185–W189.
Gallet, X., Charloteaux, B., Thomas, A., Brasseur, R., 2000. A fast method to predict protein interaction sites from sequences. J. Mol. Biol. 302 (4), 917–926.
Gao, M., Skolnick, J., 2010. Structural space of protein–protein interfaces is degenerate, close to complete, and highly connected. Proc. Natl. Acad. Sci. USA 107 (52), 22517–22522.
Gao, Y., Wang, R., Lai, L., 2004. Structure-based method for analyzing protein–protein interfaces. J. Mol. Model 10 (1), 44–54.
Golovin, A., Oldfield, T.J., Tate, J.G., Velankar, S., Barton, G.J., et al., 2004. E-MSD: an integrated data resource for bioinformatics. Nucleic Acids Res. 32 (database issue), D211–D216.
Gong, S., Park, C., Choi, H., Ko, J., Jang, I., et al., 2005a. A protein domain interaction interface database: InterPare. BMC Bioinform. 6, 207.
Gong, S., Yoon, G., Jang, I., Bolser, D., Dafas, P., et al., 2005b. PSIbase: a database of protein structural interactome map (PSIMAP). Bioinformatics 21 (10), 2541–2543.
Gray, J.J., Moughon, S., Wang, C., Schueler-Furman, O., Kuhlman, B., et al., 2003. Protein–protein docking with simultaneous optimization of rigid-body displacement and side-chain conformations. J. Mol. Biol. 331 (1), 281–299.
Hall, A. (Ed.), 2000. GTPases. Oxford University Press.
Halperin, I., Wolfson, H., Nussinov, R., 2004. Protein–protein interactions; coupling of structurally conserved residues and of hot spots across interfaces. Implications for docking. Structure 12 (6), 1027–1038.
Hart, G.T., Ramani, A.K., Marcotte, E.M., 2006. How complete are current yeast and human protein-interaction networks? Genome Biol. 7 (11), 120.
He, M.M., Smith, A.S., Oslob, J.D., Flanagan, W.M., Braisted, A.C., Whitty, A., Cancilla, M.T., Wang, J., Lugovskoy, A.A., Yoburn, J.C., Fung, A.D., Farrington, G., Eldredge, J.K., Day, E.S., Cruz, L.A., Cachero, T.G., Miller, S.K., Friedman, J.E., Choong, I.C., Cunningham, B.C., 2005. Small-molecule inhibition of TNF-alpha. Science 310 (5750), 1022–1025.
Henschel, A., Kim, W.K., Schroeder, M., 2006. Equivalent binding sites reveal convergently evolved interaction motifs. Bioinformatics 22 (5), 550–555.
Henschel, A., Winter, C., Kim, W.K., Schroeder, M., 2007. Using structural motif descriptors for sequence-based binding site prediction. BMC Bioinform., S5.
Hopkins, A.L., Groom, C.R., 2002. The druggable genome. Nat. Rev. Drug Discov. 1 (9), 727–730.
Hubbard, S.J., Argos, P., 1994. Cavities and packing at protein interfaces. Protein Sci. 3 (12), 2194–2206.
Janin, J., 2010. Protein–protein docking tested in blind predictions: the CAPRI experiment. Mol. Biosyst. 6 (12), 2351–2362.
Jefferson, E.R., Walsh, T.P., Roberts, T.J., Barton, G.J., 2007. SNAPPI-DB: a database and API of structures, iNterfaces and alignments for protein–protein interactions. Nucleic Acids Res. 35 (database issue), D580–D589.
Jones, S., Thornton, J.M., 1996. Principles of protein–protein interactions. Proc. Natl. Acad. Sci. USA 93 (1), 13–20.
Kabsch, W., Sander, C., 1983. Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22 (12), 2577–2637.
Kiel, C., Foglierini, M., Kuemmerer, N., Beltrao, P., Serrano, L., 2007. A genome-wide Ras-effector interaction network. J. Mol. Biol. 370 (5), 1020–1032.
Kiel, C., Wohlgemuth, S., Rousseau, F., Schymkowitz, J., Ferkinghoff-Borg, J., et al., 2005. Recognizing and defining true Ras binding domains II: in silico prediction based on homology modelling and energy calculations. J. Mol. Biol. 348 (3), 759–775.
Kim, P.M., Lu, L.J., Xia, Y., Gerstein, M.B., 2006a. Relating three-dimensional structures to protein networks provides evolutionary insights. Science 314 (5807), 1938–1941.
Kim, W.K., Henschel, A., Winter, C., Schroeder, M., 2006b. The many faces of protein–protein interactions: a compendium of interface geometry. PLoS Comput. Biol. 2 (9), e124.
Kittichotirat, W., Guerquin, M., Bumgarner, R.E., Samudrala, R., 2009. Protinfo PPC: a web server for atomic level prediction of protein complexes. Nucleic Acids Res. 37 (web server issue), W519–W525.
Koike, A., Takagi, T., 2004. Prediction of protein–protein interaction sites using support vector machines. Protein Eng. Des. Sel. 17 (2), 165–173.
Kortemme, T., Baker, D., 2002. A simple physical model for binding energy hot spots in protein–protein complexes. Proc. Natl. Acad. Sci. USA 99 (22), 14116–14121.
Koshland, D.E., 1958. Application of a theory of enzyme specificity to protein synthesis. Proc. Natl. Acad. Sci. USA 44 (2), 98–104.
Kraut, A., 2009. Antibody docking on the amazon cloud. Bio-IT World (May–June).
Lee, B., Richards, F.M., 1971. The interpretation of protein structures: estimation of static accessibility. J. Mol. Biol. 55 (3), 379–400.
Li, X., Keskin, O., Ma, B., Nussinov, R., Liang, J., 2004. Protein–protein interactions: hot spots and structurally conserved residues often locate in complemented pockets that pre-organized in the unbound states: implications for docking. J. Mol. Biol. 344 (3), 781–795.
Lolli, G., Lowe, E.D., Brown, N.R., Johnson, L.N., 2004. The crystal structure of human CDK7 and its protein recognition properties. Structure 12 (11), 2067–2079.
Lu, L., Arakaki, A.K., Lu, H., Skolnick, J., 2003. Multimeric threading-based prediction of protein–protein interactions on a genomic scale: application to the *Saccharomyces cerevisiae* proteome. Genome Res. 13 (6A), 1146–1154.
Lu, L., Lu, H., Skolnick, J., 2002. MULTIPROSPECTOR: an algorithm for the prediction of protein–protein interactions by multimeric threading. Proteins 49 (3), 350–364.
Ma, B., Elkayam, T., Wolfson, H., Nussinov, R., 2003. Protein–protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. Proc. Natl. Acad. Sci. USA 100 (10), 5772–5777.
Marchler-Bauer, A., Anderson, J.B., Cherukuri, P.F., DeWeese-Scott, C., Geer, L.Y., et al., 2005. CDD: a conserved domain database for protein classification. Nucleic Acids Res. 33 (database issue), D192–196.
McCoy, A.J., Epa, V.C.E., Colman, P.M., 1997. Electrostatic complementarity at protein/protein interfaces. J. Mol. Biol. 268 (2), 570–584.
Milner-White, E.J., Nissink, J.W., Allen, F.H., Duddy, W.J., 2004. Recurring main-chain anion-binding motifs in short polypeptides: nests. Acta Crystallogr. D Biol. Crystallogr. 60 (Pt. 11), 1935–1942.
Mintseris, J., Wiehe, K., Pierce, B., Anderson, R., Chen, R., et al., 2005. Protein–Protein docking Benchmark 2.0: an update. Proteins 60 (2), 214–216.
Murzin, A.G., Brenner, S.E., Hubbard, T., Chothia, C., 1995. SCOP: a structural classification of proteins database for the investigation of sequences and structures. J. Mol. Biol. 247 (4), 536–540.
Neuvirth, H., Raz, R., Schreiber, G., 2004. ProMate: a structure based prediction program to identify the location of protein–protein binding sites. J. Mol. Biol. 338 (1), 181–199.
Nooren, I.M.A., Thornton, J.M., 2003. Diversity of protein–protein interactions. EMBO J. 22 (14), 3486–3492.
Ogmen, U., Keskin, O., Aytuna, A.S., Nussinov, R., Gursoy, A., 2005. PRISM: protein interactions by structural matching. Nucleic Acids Res. 33 (web server issue), W331–W336.
Orengo, C.A., Michie, A.D., Jones, S., Jones, D.T., Swindells, M.B., et al., 1997. CATH – a hierarchic classification of protein domain structures. Structure 5 (8), 1093–1108.
Park, J., Lappe, M., Teichmann, S.A., 2001. Mapping protein family interactions: intramolecular and intermolecular protein family interaction repertoires in the PDB and yeast. J. Mol. Biol. 307 (3), 929–938.
Pattabiraman, T.N., Lawson, W.B., 1972. Comparative studies of the specificities of -chymotrypsin and subtilisin BPN'. Studies with flexible substrates. Biochem. J. 126 (3), 645–657.
Preissner, R., Goede, A., Frommel, C., 1998. Dictionary of interfaces in proteins (DIP). Data bank of complementary molecular surface patches. J. Mol. Biol. 280 (3), 535–550.
Raghavachari, B., Tasneem, A., Przytycka, T.M., Jothi, R., 2008. DOMINE: a database of protein domain interactions. Nucleic Acids Res. 36 (database issue), D656–661.
Ren, R., Mayer, B.J., Cicchetti, P., Baltimore, D., 1993. Identification of a ten-amino acid proline-rich SH3 binding site. Science 259 (5098), 1157–1161.
Res, I., Lichtarge, O., 2005. Character and evolution of protein–protein interfaces. Phys. Biol. 2 (2), S36–S43.
Richards, F.M., 1974. The interpretation of protein structures: total volume, group volume distributions and packing density. J. Mol. Biol. 82 (1), 1–14.
Riedl, S.J., Renatus, M., Schwarzenbacher, R., Zhou, Q., Sun, C., et al., 2001. Structural basis for the inhibition of caspase-3 by XIAP. Cell 104 (5), 791–800.
Russell, R.B., Alber, F., Aloy, P., Davis, F.P., Korkin, D., et al., 2004. A structural perspective on protein–protein interactions. Curr. Opin. Struct. Biol. 14 (3), 313–324.
Samsonov, S., Teyra, J., Pisabarro, M.T., 2008. A molecular dynamics approach to study the importance of solvent in protein interactions. Proteins 73 (2), 515–525.
Shoemaker, B.A., Panchenko, A.R., Bryant, S.H., 2006. Finding biologically relevant protein domain interactions: conserved binding mode analysis. Protein Sci. 15 (2), 352–361.
Singh, R., Park, D., Xu, J., Hosur, R., Berger, B., 2010. Struct2Net: a web service to predict protein–protein interactions using a structure-based approach. Nucleic Acids Res., W508–W515.
Song, H., Hanlon, N., Brown, N.R., Noble, M.E., Johnson, L.N., et al., 2001. Phosphoprotein–protein interactions revealed by the crystal structure of kinase-associated phosphatase in complex with phosphoCDK2. Mol. Cell 7 (3), 615–626.
Stein, A., Russell, R.B., Aloy, P., 2005. 3did: interacting protein domains of known three-dimensional structure. Nucleic Acids Res. 33 (database issue), 413–417.
Teyra, J., Doms, A., Schroeder, M., Pisabarro, M.T., 2006. SCOWLP: a web-based database for detailed characterization and visualization of protein interfaces. BMC Bioinform. 7, 104.
Teyra, J., Pisabarro, M.T., 2007. Characterization of interfacial solvent in protein complexes and contribution of wet spots to the interface description. Proteins 67 (4), 1087–1095.
Tsai, C.J., Lin, S.L., Wolfson, H.J., Nussinov, R., 1996. A dataset of protein–protein interfaces generated with a sequence-order-independent comparison technique. J. Mol. Biol. 260 (4), 604–620.
Vajda, S., Camacho, C.J., 2004. Protein–protein docking: is the glass half-full or half-empty? Trends Biotechnol. 22 (3), 110–116.

Valdar, W.S., Thornton, J.M., 2001. Protein–protein interfaces: analysis of amino acid conservation in homodimers. Proteins 42 (1), 108–124.
Voronoi, G., 1908. Nouvelles applications des parametres continus a la theorie des formes quadratiques. J. Reine Angew. Math. 1908 (133), 97–102.
Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., Vidal, M., 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 287 (5450), 116–122.
Wass, M.N., Fuentes, G., Pons, C., Pazos, F., Valencia, A., 2011. Towards the prediction of protein interaction partners using physical docking. Mol. Syst. Biol., 469.
Wells, J.A., McClendon, C.L., 2007. Reaching for high-hanging fruit in drug discovery at protein–protein interfaces. Nature 450 (7172), 1001–1009.
Winter, C., Henschel, A., Kim, W.K., Schroeder, M., 2006. SCOPPI: a structural classification of protein–protein interfaces. Nucleic Acids Res. 34 (database issue), 310–314.
Wohlgemuth, S., Kiel, C., Kramer, A., Serrano, L., Wittinghofer, F., et al., 2005. Recognizing and defining true Ras binding domains I: biochemical analysis. J. Mol. Biol. 348 (3), 741–758.
Wright, C.S., Alden, R.A., Kraut, J., 1969. Structure of subtilisin BPN' at 2.5 Å resolution. Nature 221 (5177), 235–242.
Xu, G., Cirilli, M., Huang, Y., Rich, R.L., Myszka, D.G., et al., 2001. Covalent inhibition revealed by the crystal structure of the caspase-8/p35 complex. Nature 410 (6827), 494–497.
Xu, Q., Dunbrack, R.L.J., 2011. The protein common interface database (ProtCID) – a comprehensive database of interactions of homologous proteins in multiple crystal forms. Nucleic Acids Res. 39 (database issue), D761–D770.
Yang, J.-M., Chen, Y.-F., Shen, T.-W., Kristal, B.S., Hsu, D.F., 2005. Consensus scoring criteria for improving enrichment in virtual screening. J. Chem. Inf. Model 45 (4), 1134–1146.
Yu, H., Luscombe, N.M., Lu, H.X., Zhu, X., Xia, Y., et al., 2004. Annotation transfer between genomes: protein–protein interologs and protein-DNA regulogs. Genome Res. 14 (6), 1107–1118.
Zhu, H., Domingues, F.S., Sommer, I., Lengauer, T., 2006. NOXclass: prediction of protein–protein interaction types. BMC Bioinform., 27.