Chapter 12
Protein-Protein Interactions: Structures and Druggability

David B. Ascher, Harry C. Jubb, Douglas E.V. Pires, Takashi Ochi, Alicia Higueruelo, and Tom L. Blundell

Abstract While protein-protein interfaces have promised a range of benefits over conventional sites in drug discovery, they present unique challenges. Here we describe recent developments that facilitate many aspects of the drug discovery process – including characterization and classification of interfaces, identifying druggable sites and strategies for inhibitor development.

12.1 Historical Background

Over the past 40 years structure-guided approaches have become increasingly central to the discovery and design of new therapeutics. Initially the focus was on modification of natural substrates or molecules known to bind and inhibit enzymes or cell surface receptors. Over the past 20 years new hits have been derived largely from screening using either whole cell assays or enzyme assays with chemical libraries that may number hundreds of thousands of drug-like compounds. Chemical libraries have been refined to make them more closely compliant with the Lipinsky Rule-of-Five that requires molecules to be less large (MW <500), less lipophilic (LogP <5), less flexible and have the requisite number of hydrogen bond donors and acceptors, features of molecules that have led to successful therapeutics [1]. This has led to many successful drugs reaching the market but at exponentially increasing costs.

One feature of drug-discovery campaigns has been the tendency to select targets that have been defined as "druggable", often leading to a focus on large enzyme superfamilies where one member has been target for a successful drug campaign; examples include aspartic proteinases [2–4], metallo proteinases [5–9], transferases [10, 11] and protein kinases, all of which have well defined concave active sites [12]. Designs have often been mechanism based, reflecting either the co-factor or

D.B. Ascher • H.C. Jubb • D.E.V. Pires • T. Ochi • A. Higueruelo • T.L. Blundell (✉)
Department of Biochemistry, Sanger Building, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, UK
e-mail: dascher@svi.edu.au; tom@cryst.bioc.cam.ac.uk

G. Scapin et al. (eds.), *Multifaceted Roles of Crystallography in Modern Drug Discovery*, NATO Science for Peace and Security Series A: Chemistry and Biology,
DOI 10.1007/978-94-017-9719-1_12

an enzyme intermediate or transition state. These have been optimized to useful leads often using structure-guided approaches. More recently with structures of membrane proteins available there has interest has revived in the GPCRs and other members of large membrane protein superfamilies that have been classically successful targets for phenotypic screening approaches.

The development of fragment-based drug discovery has allowed the more effective exploration of chemical space for ligands targeting a particular binding site by using much smaller, chemically diverse libraries composed of smaller molecules, thereby decreasing complexity. These small molecules, usually between 100 and 300 Da and consistent with the Rule-of-Three [13], have been termed 'fragments'. Detection and development of these fragments are complicated by the fact that they typically bind with much weaker affinities, in the mM range, than larger molecules. Wells and colleagues pioneered a way around this problem, tethering them by exploiting thiol-containing fragments that react through a reversible disulphide bond with a cysteine residue that has been engineered into a protein [14]. This proved particularly successful for protein-protein interactions or where sites require trapping a particular conformer. Another approach that has been successfully used to screen and validate fragment binding has relied on highly sensitive biophysical methods [15, 16]; including nuclear magnetic resonance (NMR; [17–19]), X-ray crystallography [20, 21], surface plasmon resonance (SPR; [22–24]), differential scanning fluorimetry (DSF; [25–27]) or isothermal calorimetry (ITC; [28–31]). Fragment hits identified from either approach are subsequently elaborated by crosslinking or 'growing' the fragment, while maintaining strong interactions for each group added. This method has proven very effective when used as part of a structure-guided approach [18, 32, 33].

12.2 Obtaining Selectivity with Multiprotein Systems

It has become increasingly evident over the past decade that it is difficult to obtain selectivity, especially with transition and intermediate state analogues of enzymes or those targeting co-factor binding sites. In particular the challenges with targeting protein kinases has become particularly clear as pharmaceutical companies have increased the numbers of superfamily members that can be assayed. Much of the optimism of getting very good selectivity with protein kinase inhibitors by exploiting sub-pockets around the ATP binding site has been moderated by the discovery that sub-clusters of protein kinases with similar co-factor binding sites are recognised by many of the molecules previously thought to be selective.

One of the ways of improving selectivity is to move away from targeting active sites towards regulatory multiprotein systems that are critical to cell activity [34]. A particularly fruitful line of investigation has been to target the various protein interactions that regulate different members of one superfamily. Thus, receptor tyrosyl kinases have multiple different regulatory extracellular regions that may interact with secondary receptors as well as their ligands, often leading to clustering

on the cell surface. For example fibroblast growth factor receptor (FGFR) recognizes the FGF ligand at a binding site between extracellular domains d2 and d3 but a secondary receptor the proteoglycan heparan sulphate is also critical for activity and tends to mediate clustering of receptor and ligand [35–38]. The MET receptor tyrosyl kinase differs in its extracellular region that recognizes a very different protein growth factor, HGF/SF [39–41]; in this case the heparan sulphate secondary receptor is important for some splice forms. Other receptor tyrosyl kinases such as epidermal growth factor (EGF) receptor and the insulin receptor are related evolutionarily but have different domain organizations of the region recognizing the ligands. These systems give opportunities for greater selectivity of inhibitors, either by targeting protein-protein interfaces in the assemblies directly (orthosteric inhibitors; see ref [42] for review) or indirectly through allosteric binding sites that stabilize conformers incommensurate with ligand binding at another site, as for example identified for the FGF receptor [43].

Intracellular signaling pathways are also regulated by multiprotein systems of similar complexity. The model of receptor activation that leads to a pathway of interactions, somewhat resembling the metabolic pathways familiar to biochemists, appears to be giving way to the idea that large multiprotein systems often assemble to regulate many intracellular kinases and these mediate interactions between the cell membrane and various cytoplasmic and nuclear targets [44]. In the nucleus complex regulatory systems, for example mediating DNA double-strand break repair, involve multi-component systems. Non-homologous end-joining (NHEJ) requires many factors: the Ku70 and 80 heterodimer that assembles on double-strand breaks (DSBs); the scaffolding proteins XRCC4 and XLF which interact with Ku; the key protein for recruiting NHEJ proteins at DNA ends, the ligase (DNA ligase IV) that joins the ends [45]; and the DNA-PKcs that is involved in signalling and regulating DNA repair. These proteins assemble together with other proteins such as the nuclease Artemis that also has binding sites for the ligase and DNA- PKcs [46]. Ku interacts directly with DNA-PKcs [47, 48], DNA ligase IV [49] and XLF [50] in a DNA-dependent manner, and recruits NHEJ proteins *in vivo* only when DSBs are generated.

Such complex multiprotein assemblies at the membrane, in the cytoplasm and in the nucleus often regulate cellular processes through co-location of various critical components. However, they also likely play a role in increasing signal to noise. Although binary interactions between two proteins would often occur opportunistically in the cell, especially in the cell membrane or in the limited environment of the nucleus or cytoplasm, a weak binary interaction followed by interactions of further components would give a cooperative but reversible assembly of a large multiprotein complex, allowing selective signaling regulation in the cell [44].

There are some occasions where binary systems are required in signaling and regulatory processes. These are often mediated by concerted folding and binding of one protein. This was recognized more than 30 years ago in the polypeptide hormones like glucagon which are disordered in solution but can associate with the receptor in a cooperatively formed secondary or supersecondary structure, often

first binding an anchor residue, which forms a hotspot of the interaction [51]. Such concerted-folding-and-binding is found widely in intracellular systems. A good example is BRCA2 BRC4 repeat interaction with RAD51, in which a phenylalanine, the anchor residue, recognizes a well-defined pocket on the RAD51 [52]. This then also allows a much smaller pocket, a second hotspot to be recognized, the interaction probably being driven by unhappy water, leading to high selectivity for alanine. The remaining part of the BRC4 repeat then folds onto the surface of the RAD51 through a weaker and less well defined interaction involving folding and binding of a helix onto a further hotspot. The cooperative folding and binding constitutes provides a second mechanism for obtaining selectivity and has been widely studied for intracellular systems by Wright and coworkers [53].

The two mechanisms of gaining selectivity – co-operative assembly of multi protein regulatory assemblies or co-operative folding and binding – both involve protein-protein interactions. Here we review these interactions, focusing on how they are defined experimentally, the nature of the interfaces that mediate the interactions, the effects of mutations at protein-protein interfaces and their roles in genetic diseases, and the druggability of either the isolated interfaces or the interfaces themselves.

12.3 How to Define Structures of Multiprotein Assemblies

The complex nature of many of the regulatory assemblies demands different techniques to characterize the stoichiometry of the interactions that vary in space and time. These range quite broadly in the resolution and detail that they provide: from assessments of stoichiometries and molecular radii, to overall topology of a complex provided by small angle X-ray scattering and electron microscopy, to the atomic resolution provided by X-ray crystallography and increasingly single particle cryo-electron microscopy (Fig. 12.1a). The information obtained from these diverse techniques is often complementary and help provide an overall understanding of a given complex.

Nanospray mass spectrometry (MS) can accurately determine stoichiometry of macromolecular complexes as large as a few MDa [54]. An advantage of the technique is that samples do not need to be homogenous, and it can detect different oligomeric states existing in equilibrium together. Hence, it is very useful for macromolecules that form dynamic complexes. Furthermore, nanospray MS can provide topological information of macromolecular assembles [55]. One of limitations is the buffer, which should be volatile, for example ammonium acetate. If nonvolatile chemicals are essential for the proteins or complex, alternative methods to study stoichiometry may be required. These include analytical ultracentrifugation (AUC) and size-exclusion chromatography combined with multi-angle light scattering (SEC-MALS), techniques that have been vital in establishing the oligomeric states of proteins previously thought to be monomeric, and the stoichiometry of their modulation by small molecules [56–58].

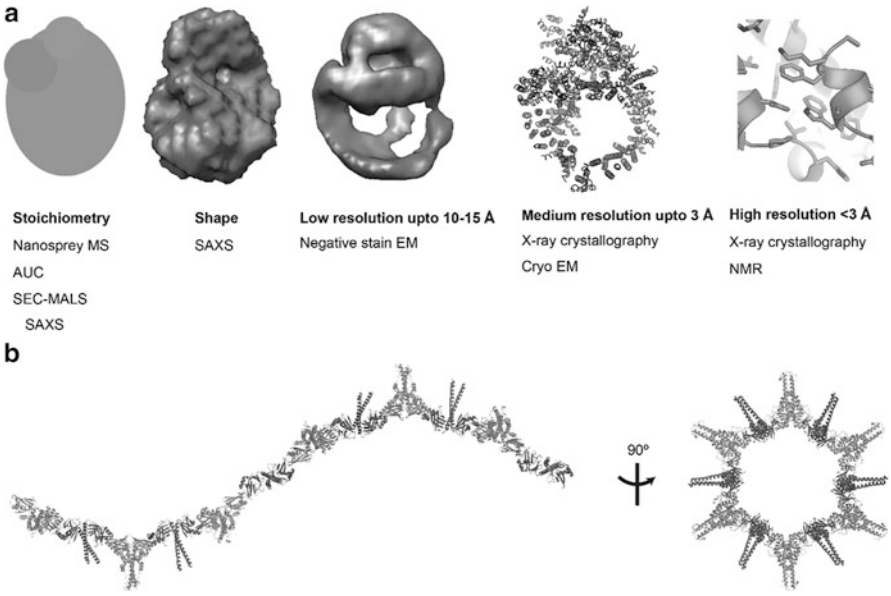

Fig. 12.1 Complementary-biophysical analyses provide different resolutions of structural information. (**a**) Comparison of biophysical techniques and resolutions. Structures of DNA-PKcs (BioIsis ID: 1DPKCY) [70], DNA-PK (EMDB ID: EMD-1210) [72], DNA-PKcs in complex with Ku80 C-terminus (PDB ID: 3KGV) [46], LigIV in complex with Artemis (PDB code ID: 3W1B) [85] were used to show examples of structural information obtained from SAXS, negative-stain EM and X-ray crystallography (low and high resolution). (**b**) Structure of the XRCC4/XLF filament. The structure of the XRCC4 (*dark grey*)/XLF (*light grey*) complex shows an alternative left-handed filament from two different views

Small angle X-ray scattering (SAXS) can also be used to determine stoichiometry of target proteins. A recent advance in SAXS allowed us to measure approximate molecular weights of only the proteins of interest from scattering data [59]. SAXS is also a good way to get a general idea of the solution structure and is particularly useful if structures of individual protein components are available [60, 61]. It can also be used to study protein complexes that have dynamic protein-protein interactions by combining with size-exclusion chromatography [62]. X-ray crystallography has proved the mainstay for defining complexes although interactions with smaller peptide ligands are often also accessible through NMR. Increasingly high-resolution cryoelectron microscopy is becoming very powerful and can even recognize small ligands binding to complex structures [63]. Isothermal calorimetry can be used to measure the thermodynamics of the interactions and surface plasmon resonance the kinetics.

For an example of the characterization of increasingly complex structures let us return to NHEJ, involved in repair of DNA-double-strand breaks and introduced above. The kinase catalytic subunit, DNA-PKcs, is a 460 kDa protein containing a long HEAT-repeat region of about 3,000 of the total 4,000 amino acid residues [64].

The C-terminal kinase region has α-helical FAT and FATC domains, in addition to the kinase domain. The high-resolution structure of the kinase region of mTOR, which is a paralog of DNA-PKcs, has a similar structure and was defined at high resolution X-ray crystallography [65]. The structure of the DNA-PK complex, which includes the Ku hetero-dimer, has been studied using EM, SAXS and X-ray crystallography. Early EM structures of DNA-PKcs identified two regions: crown/head and base/palm [66–68]. The resolution of a cryo-EM structure of DNA-PKcs was extended to 7 Å resolution many years later, revealing the secondary structure of the molecule [69]; however, the crystal structure of DNA-PKcs was required to determine unequivocally the location of the kinase [46]. The crystal structure of DNA-PKcs in complex with the C-terminal of Ku80 at 6.6 Å resolution showed that the HEAT repeats form a circular structure, consistent with the early EM models of DNA-PKcs. Recent EM and SAXS studies of DNA-PKcs showed that it undergoes a large conformational change upon autophosphorylation [70, 71], which is difficult to study using crystallography. Although the structure of whole DNA-PK complex has been investigated using EM and SAXS [70, 72], it remains unresolved where Ku70/80 binds on DNA-PKcs but both techniques consistently showed a hetero-hexameric complex of these proteins at two DNA ends. DNA-PK on DNA ends was also observed by atomic force microscopy (AFM) [73, 74]. AFM is particularly useful to see where proteins are bound on DNA because it can visualize naked DNA. Since these techniques are complementary to each other, accumulated structural studies of DNA-PK provide insights into how the complex binds DNA spatially.

In addition to DNA-PK, the structure of the ligase holoenzyme has been studied extensively and it provides a good example as to how the techniques complement each other. The DNA ligase IV, XRCC4 and XLF complex contributes to the last step of NHEJ [75]. Crystal structures of complexes of XRCC4 and XLF, structural paralogs that are both homodimers, show that they form a left-handed helical filament (Fig. 12.1b) [76–79]. A concentration-dependent formation of the filament is shown by gel filtration, nanospray mass spectrometry and SAXS [79] and a scanning force microscopy of the complex demonstrates that DNA stabilizes filament formation [76]. Interestingly, filament formation is inhibited by the strong and stable LigIV interaction with XRCC4 [80], likely due to the catalytic domains of LigIV, because LigIV/XRCC4 without the domains still forms the filament with XLF. SAXS studies of LigIV/XRCC4 indicate that the catalytic domains are flexibly tethered to a tandem repeat of BRCT domains [75, 81], which interact with XRCC4 [82, 83]. Thus, the dynamic nature of the catalytic domains prevents XLF from interacting with XRCC4. Negative-stain EM of LigIV/XRCC4 shows that the catalytic region is fixed at the N-terminal head domain of XRCC4 [84].

DNA-PK and ligase holoenzymes together with other NHEJ proteins such as Artemis are present at DNA ends. However, the dynamic nature of these protein-protein interactions makes it difficult to study structurally, which has high demand on homogeneity. Since we have atomic structures of the individual core NHEJ proteins, cryo-EM may be a reasonable technique to observe the entire complex. A key point for the success will depend upon how much we can stabilize the

complex. For this optimization, nanospray MS and other techniques will help inform distribution of stoichiometry and stability of the complex.

These challenges in defining the spatial and temporal interactions of the NHEJ system are likely to be common to many of the multicomponent systems in the cell. Nevertheless they provide a basis for understanding their roles in cell regulation and signaling, and some data that is proving useful in the design of chemical tools that can be used to selectively modulate cell activity and provide the first clues about how to proceed in discovering candidate drugs.

12.4 Organization of PPI Information: Description of Piccolo and Credo

The wealth of data publicly available in the PDB [86] allows structural comparison of interacting proteins with a complete range of partners (solvents, small molecules, small peptides, saccharides, nucleic acids and other proteins). However, in order to use this information meaningfully and efficiently for drug discovery, the data in the PDB, a flat-file-based databank, need to be better organized. Efforts to do this range from resources specialized in one type of structural interaction, like beta-sheet or alpha-helix motifs [87, 88] or domain-domain interactions [89] to resources that emphasize the mechanistic aspects of interactions, like the ASD Allosteric Database [90, 91].

However, integrating specialized resources is challenging, as we discovered with the sister databases developed in our laboratory, describing structural interactions with atomic detail for protein-protein (PICCOLO, [92]), protein-nucleic acid (BIPA, [93]) and protein-ligand complexes (CREDO, [94]). A new CREDO database [95] has now been developed with the aim of enclosing under a single resource not only all pairwise atomic interactions of inter- and intra- molecular complexes from the PDB, but also disparate data relevant to drug discovery; these include SNP databases (OMIM and COSMIC [96, 97]), mappings to sequence data from UniProt [98] and EnsEMBL [99], ChEMBL [100] binding data and the small molecules fragmented with an enhanced RECAP [101] algorithm. In addition to providing relational data structures for storing protein structure data at model, chain, residue, ligand and atom level, CREDO provides chemoinformatic routines to analyze small molecule data, such as fingerprint generation, similarity and substructure searching and chemical fragmentation. Where entities in PDB structures are involved in non-bonding interactions, such as in protein-ligand and protein-protein interactions, pairwise atomic contacts are explicitly characterized, for example as hydrogen bonding, ion pair, metal complex or specific aromatic ring interactions. These pairwise atomic contacts are stored as structural interaction fingerprint SIFTs [102], used for clustering interactions to identify common patterns or to study molecular recognition, so making CREDO a comprehensive analysis platform for drug discovery. The information on intermolecular interactions is integrated with

further chemical and biological data. The database implements useful data structures and algorithms such as cheminformatics routines to create a comprehensive analysis platform for drug discovery. The database can be accessed through a web-based interface, downloads of data sets and web services at http://structure.bioc.cam.ac.uk/credo.

The information in CREDO allows the user to move from target to target using a residue map (that links sequence to structure) to UniProt [98], or to analyze the intra-residue network interactions at the protein interfaces for correlation with hotspots.

12.5 Distribution of Protein-Protein Interactions and Pocket Size

Organizing the wealth of publicly available structural protein-protein interaction (PPI) data has made it clear that PPI interfaces, the chemical surfaces through which proteins interact with each other, come in many different shapes and sizes [92]. Multiprotein complexes can assemble from globular protomers, interacting with partners to form homo-complexes of sequence-and-structure-identical protomers; between partners of different sequence to form hetero-complexes, or between globular partners and short peptides or even longer polypeptides that are often disordered before binding [103]. There are also examples of peptide-peptide associations [104]. The range of different domains and peptides involved in interactions provides diversity in PPI binding sites [103, 105, 106]. While PPI interfaces have historically been described as being "flat and featureless" the growing number of examples of orthosteric PPI modulators [107], which compete for the binding site of one protein to another, speaks to the fact that not all PPI associations are as featureless and un-amenable to chemical modulation as was sometimes thought.

Concavity is generally accepted as a feature of many protein-ligand interactions, where binding deep into a protein's surface may maximize the interaction area between a protein and a small-molecule ligand, and where ligand binding may be entropically favorable through the ejection of water molecules from the protein's solvation shell into bulk solvent [108, 109]. Computational analyses have shown that where PPI interfaces have been successfully modulated, surface concavity at the binding site usually exists not in the single, large volume cavities found in "traditional" drug targets, but rather in multiple small, geometrically clustered concavities [110]. Examples of these kinds of concavities at interfaces have been identified as being used as "complemented pockets" by protein partners [111] involving deeply bound single residues. The concept is related to the "hotspot" hypothesis that single, buried residues or clusters of residues contribute a large proportion of interface interaction energy [112, 113].

The similar "anchor" hypothesis states that energetically important, solvent-buried residues at the interface are involved in initial, fast lock-and-key type recognition, followed by a more gradual relaxation of peripheral residues by an

induced fit mechanism to form the mature interface [114, 115]. Exploitation of these anchor sites has been used in the design of orthosteric inhibitors [116, 117]. Protein-peptide interfaces may be particularly amenable to orthosteric modulation over other associations, at least in part due to their tendency to consist of single interacting "segments" where a linear binding epitope contributes a large proportion of the binding free energy of the interface [118], and the conformation of the globular partner is typically fixed such that the surface presents concavity to a disordered binding partner which subsequently folds on binding [103, 111]. The extensive buried surface area no doubt contributes to the affinity and compensates for the loss of entropy on folding. However, even in larger, more globular interfaces, there is a tendency for a single linear epitope to contribute large proportions of the interface's interaction energy [119, 120].

Much effort both in academia and industry has focused on the modulation of pairwise PPIs through the development of competitive, orthosteric inhibitors of interface formation. Although many interactions are dimeric, the disruption of a dimer forming part of a larger complex is likely to disrupt the assembly of that complex. As discussed above many biological systems involve the coordination of multiple protomers through strong but reversible cooperative binding to achieve high signal-to-noise ratios in cellular processes. There are advantages to attempting to chemically modulate multiprotein systems by targeting pre-bound structures, to either stabilize the complex to have a therapeutic effect or to ablate the formation of a higher order complex. "Interfacial inhibitors", or stabilizers, binding juxtaposed to a PPI binding site have been explored as a potential PPI modulation strategy [121]; in part successes may stem from the existence of better defined binding sites in the periphery of an existing interface [122]. There is also increasing interest in allosteric strategies, where the binding site is distal to the interface, for disrupting multiple protein assembly.

Systematic analysis in our laboratory of over 9,000 pairwise, non-overlapping PPI interfaces, organized in our databases and filtered for structure quality, has indicated that protein-peptide interfaces make more extensive use of concavity than other kinds of interfaces, both on average and at their deepest. However, in spite of being flatter on average, a large proportion of globular-globular interfaces make use of small pockets of concavity, through deeply bound residues. The landscapes of PPI interface surfaces make more subtle use of concavity than traditional targets, therefore requiring innovative approaches for drug discovery.

12.6 Mutations & Interfaces: Their Role in Diseases

Mutations are a natural consequence of evolution, and understanding how they interact with their partners can yield insights into protein function, diseases and help guide a range of experimental efforts including protein engineering and drug development. The first efforts to understand the effects of mutations focused on their ability to alter protein folding and stability. The pioneering method SDM

[123, 124] used a statistical potential energy function, derived from environment-specific residue propensities in structural families [125], in order to predict the change in free energy of folding upon mutation for a given protein.

The leap in computational power provided by new architectures, together with the rapid increase in protein experimental and structural data generation, has created new opportunities for enhancing the existing approaches (for a review of available methods see [126]). This scenario led to the development of mCSM [127], a novel machine learning method which is proving to be significantly more accurate and scalable than previous approaches. mCSM uses the concept of graph-based signatures to represent the three-dimensional environment of a wild-type residue, which are then used to train highly accurate predictive models, capable of quantitatively assessing the effects of mutations. These signatures have been proposed previously and successfully adapted in a range of applications including: protein inter-residue analysis [128], protein automatic structural classification and function prediction [129] and receptor-based ligand prediction [130]. However all these methods present only a portion of the story necessary to understand fully the effects of mutations as they did not take into account the multitude of interactions vital for normal cellular function. In this context, tools for assessing the impact of mutations on protein-protein interfaces are necessary.

Some of the early approaches to predict the effect of mutations on the binding free energy of specific protein-protein complexes included energy-function based approaches [131–133] and more computationally intensive calculations [134–141]. These methods, however, focused on mutations to alanine, which will be discussed further below. While alanine scanning is extremely important experimentally, in order to understand the broad array of genetic variations, and mutations in diseases, a more challenging demand was to develop methods capable of predicting the effects of any mutation. To this end we developed mCSM-PPI (http://bleoberis.bioc.cam.ac.uk/mcsm/protein_protein) which employs the mCSM signatures used as evidence to train predictive models based on experimentally measured effects on protein-protein affinity from the SKEMPI database [142]. The method has shown to be effective, performing extremely well in comparison with other methods, achieving a Pearson's correlation coefficient of up to 0.8 (as shown in Table 12.1).

Other methods have been recently described including BeAtMuSiC [143], ZEMu [144] and those described by Li et al. [145] and Moal et al. [146]. Since these methods were developed concurrently, a comparative analysis of their relative performance is of general interest. As shown in Table 12.1, mCSM can outperform methods that employ a range of different techniques, some of them computationally intensive, such as Molecular Dynamics and Molecular Mechanics.

Structure-based methods like mCSM-PPI are essential tools for understanding the relation between the quantitative effects of mutations in protein-protein affinity and their roles in Mendelian diseases and in cancer, as well as to shed light on the understanding of their mechanism of action.

One example of the usefulness of such methods is the recently published work of Gossage et al. [147] of mutations on von Hippel-Lindau disease (VHL) and their relation with propensity or risk of developing renal cell carcinoma (RCC).

Table 12.1 Predicting the effects of mutations on protein-protein interactions

Method	Technique	Data set (# of mutations)	Correlation (SE)	mCSM-PPI performance on similar data set	Web server	Refs.
mCSM-PPI	Structural signatures – ML	SKEMPI (2007)	0.80 (1.25)	N/A	Yes	[127]
BeAtMuSiC	Statistical potentials	SKEMPI (2007)	0.40 (1.80)	0.58 (1.55)	Yes	[143]
FoldX	Energy function	SKEMPI (1844)	0.37 (2.14)	0.58 (1.55)	Yes	[131]
Li et al.	MD	SKEMPI (1844)	0.58 (1.55)	0.58 (1.55)	No	[145]
ZEMu	MD	SKEMPI (1254)	0.51 (1.34)	0.58 (1.55)	No	[144]
Moal et al.	Contact potentials	SKEMPI (1949)	0.73 (NR)	0.80 (1.25)	No	[146]

Relative performance of computational approaches to predict the effect of mutations on the binding free energy of protein-protein complexes
ML machine learning, *MD* molecular dynamics, *NR* not reported

An integrated computational approach was developed using structural information to understand the relation between the severity of phenotype, and the predicted effects of mutations on the stability of the pVHL protein and the change in affinity between pVHL and its protein partners (elongin B, elongin C and HIF-α peptide). The method, called Symphony (http://structure.bioc.cam.ac.uk/symphony), was able to predict the effects of mutations associated with RCC with high levels of sensitivity and specificity. A database of predictions for mutations not yet observed has also been developed.

mCSM-PPI has also been capable of giving a rationale for the affects of mutations on PPIs that are related to inherited RCC in other genes, including the P15-CDK6 complex. Figure 12.2a, b show examples of the affects of mutations on the stability of pVHL-HIF-α and P15-CDK6 complexes that were correctly identified by mCSM-PPI to dramatically reduce protein-protein affinity and potentially disrupt the complex.

Another successful application of computational predictors for understanding the mechanism of action of mutations in Mendelian diseases is the study of alkaptonuria (AKU). AKU is a rare, inherited metabolic disease caused by defective homogentisate 1,2-dioxygenase (HGD) as a result of mutations that disrupt its activity, many of them occurring in the protein-protein interfaces of its homo-hexamer (as the example shown in Fig. 12.2c). By using the predictions obtained by DUET [148] and mCSM-PPI, the mutations described in AKU were classified as belonging to one of three possible mechanism classes: protomer-destabilizing, PPI-destabilizing and active site mutations. Figure 12.2d depicts the distributions of these mutations on the structure of the human HGD.

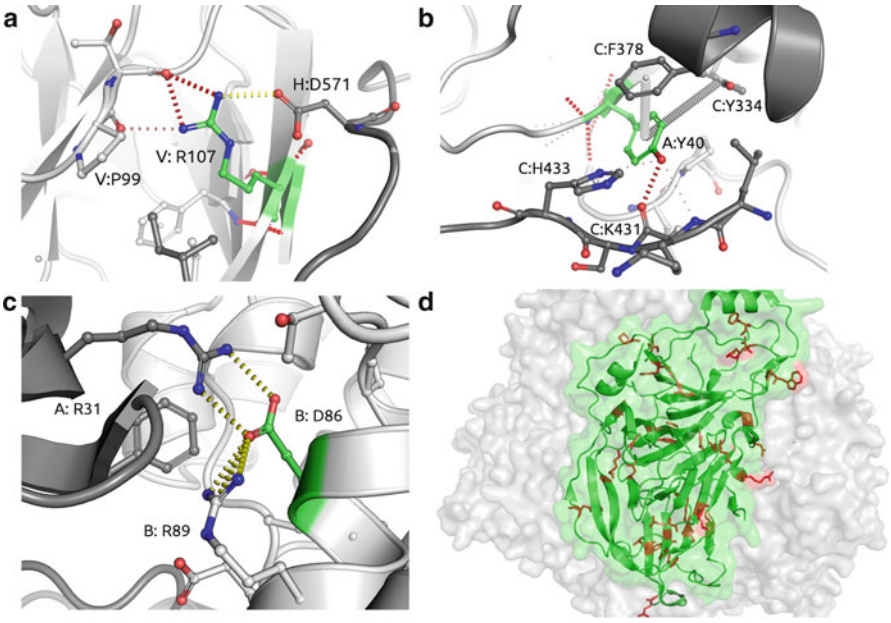

Fig. 12.2 Structural analysis of effects of mutations on PPIs and their role in cancer and Mendelian diseases. Panels (**a–c**) depict the interaction network made by important interface residues mutated in von Hippel-Lindau disease, alkaptonuria and renal cell carcinoma, respectively, whose effects were predicted to be highly destabilizing by mCSM-PPI. The chains of the binding partners are shown in dark grey. Panel (**a**) shows the interactions made by ARG107 in the interface between VHL and the HIF-2α peptide. Mutation to Proline will perturb local secondary structure and disrupt strong intra and inter-molecular hydrogen bonds and charged interactions. Panel (**b**) shows the TYR40 residue of the human HGD, which forms strong inter- and intra-molecular pi-pi interactions that are lost upon mutation to serine in AKU. Panel (**c**) shows the P15-CDK6 complex. Mutation of residue ASP86 on P15 to asparagine results in a loss of a charged interaction at the core of the interface. Panel (**d**) highlights a selection of AKU mutations within human HGD that mCSM-PPI predicts will reduce protein-protein affinity, leading to a loss of enzyme activity. These are spread across the extensive binding interface

12.7 Mutations & Interfaces: Hotspot Identification

Experimental and computational alanine scanning have been popular approaches to identifying amino acids that are critical for the formation of the complex, termed hotspot residues [112, 132, 133]. Robetta alanine scanning defines hotspot residues as those that upon mutation to alanine are predicted to decrease the binding energy by at least $\Delta\Delta G$ 1.0 kcal/mol.

Using mCSM-PPI, an alanine scan of 743 mutations within 19 different protein-protein complexes was performed to identify potential hotspot residues. The distribution of changes (Fig. 12.3) in binding free energy are consistent with the hypothesis that the loss of hotspot residues would have a significant effect

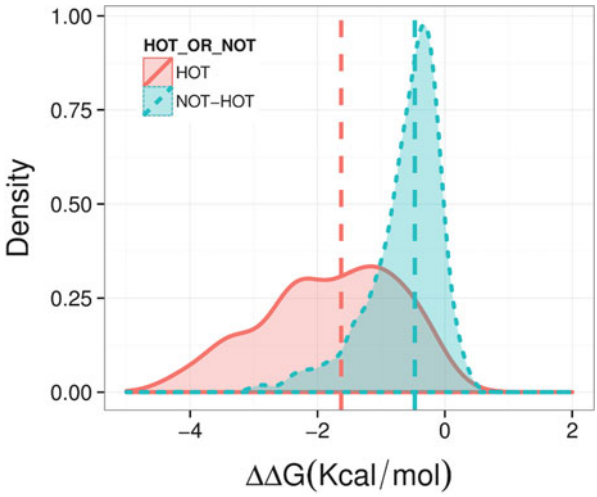

Fig. 12.3 Density distribution of protein-protein affinity change predictions by mCSM-PPI for mutations that were experimentally assigned as hotspots. While the majority of the neutral mutations (*dotted-line curve*) were predicted to have little impact, most of the hotspots (*solid-line curve*) were predicted to have a significant destabilizing effect on the protein-protein complex

upon the affinity of two binding partners. The predicted change in binding energy showed that mCSM-PPI predictions correlated strongly with the experimental data ($r > 0.7$). This indicates that mCSM-PPI could also be a powerful tool for hotspot identification.

12.8 Examples of Success Using FBDD to Target PPI's

PPI's have been successfully modulated by compounds that mimic protein interaction elements, including proteomimetics [149], foldamers [150], peptide aptamers [151], antibodies [152, 153] and affibodies [154], where unfavorable pharmacokinetic properties are modified by the use of drug carriers or chemical modifications like PEGylation [155]. However, the development of more traditional pharmaceutical small molecule modulators, which remains highly desired, is proving a viable strategy, as demonstrated by several small molecule PPI inhibitors currently used therapeutically including the anti-HIV drug Maraviroc, an inhibitor of the CCR5-gp120 interaction, and Titrobifan, a glycoprotein IIb/IIIa inhibitor used in cardiovascular disease.

Two main resources store small molecule data modulating protein interfaces: the 2P2I database [156] dedicated to structural complexes of orthosteric PPI inhibitors of PPI and the TIMBAL database [157] that holds small molecule data for PPI modulators (inhibitors and stabilizers). Comparison of the original contents of the TIMBAL database [158] with known drugs and standard screening compounds revealed that small molecules disturbing protein assemblies were bigger, more lipophilic and with less polar features than the drugs and standard synthetic molecules. Analysis of the contacts these inhibitors made when the structure was

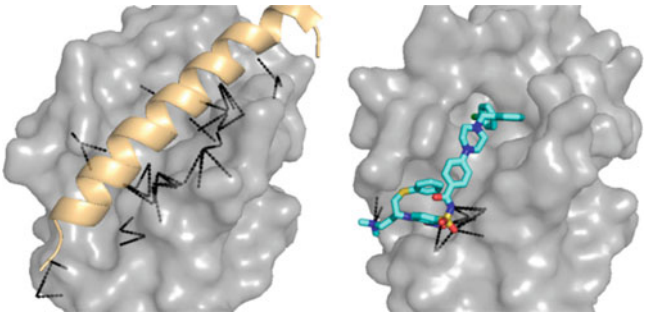

Fig. 12.4 Bcl-XL. *Left*: pdb code 2BZW, Bcl-XL (*surface representation*) bound to BAD (*cartoon representation*). *Right*: pdb code 2YXJ, Bcl-XL (*surface representation*) with the Abbott compound ABT-737 (*sticks representation*). Only polar contacts are shown

available corroborated that on average these small molecules were engaging mainly in hydrophobic contacts with the protein target.

These observations raised the question as to whether the lipophilicity is a requirement for binding to protein-protein interfaces or a reflection of common sins in drug discovery [159]. Comparison of protein complexes with PPI inhibitors, including synthetic and natural small molecules, small peptides and other proteins, highlighted the fact that protein complexes and natural molecules tend to interact with higher ratios of polar to non-polar contacts than synthetic small molecules. Contrasting the few cases where structures exist for both the protein-protein and the protein-PPI inhibitor complexes, synthetic small molecules were shown to miss available polar contact opportunities at the protein interface [160]. Figure 12.4 shows a graphical view of this concept.

When a structure is available, however, it can provide very useful insight and plays a crucial role in the development of fragment hits. The cytokine interleukin-2 (IL-2) induces T cell proliferation through binding to its heterotrimeric receptor. The structure of a small molecule inhibitor of this interaction identified by Hoffman-La Roche revealed that binding to IL-2 induced a significant conformational change to create a hydrophobic binding pocket that could accommodate the inhibitor [161]. This region overlapped with hotspot residues identified by alanine scanning mutagenesis [162, 163]. Based on this information, Wells and colleagues created a series of 11 cysteine mutants to identify small molecule inhibitors through tethering [161]. This identified a number of fragments that were shown to bind with sub-micromolar affinity. Medicinal chemistry was able to improve this affinity further to the low micromolar range. The crystal structures of the complexes, however, revealed two fragments bound to close but distinct sites. Linking these fragments together, they were able to achieve nanomolar inhibitors of the IL-2 interaction.

An example of the power of biophysical FBDD approaches to target a PPI is the development of inhibitors of the interaction between the human recombinase RAD51 and BRCA2 [52, 164]. Initial screening of a fragment library by thermal shift, followed by validation using NMR and X-ray crystallography resulted in the

structures of approximately 80 fragments bound to RAD51, which disrupted the interaction with BRAC2. With structural information in hand, fragment growing or fragment linking can be employed to identify larger compounds from one or more fragment starting points. In general, as a fragment is expanded to make additional interactions in a fragment-based drug discovery campaign, affinities tend to be increased by 3–5 orders of magnitude [165–169]. The growth of the fragments bound to RAD51 was guided by the co-crystallized structures together with the structure of RAD51 in complex with the BRC4 region of BRAC2, and was able to improve the K_D from the mM to sub mM range [52]. More recently nanomolar affinities have been achieved by the Cambridge Group.

How can we tackle these challenging interfaces using chemistry that brings more polar specific contacts into play? On the one hand, interfaces where a flexible peptide binds to a continuous epitope in a concerted folding seem to offer more opportunities for "ligandability" [170] than preformed globular protein partners assemblies [103]. On the other hand, fragment-based approaches [171, 172] give pivotal advantages for these targets as they identify the hotspots by binding and yield less hydrophobic hits [173].

12.9 Final Thoughts

PPI's play a crucial role within the cell and their perturbation can lead to a range of diseases. They also present attractive and selective sites for drug development. Significant improvements in methodology have allowed the development of some highly selective modulators. Although targeting protein-protein interfaces still presents considerable technical challenges, as our understanding of these sites continues to expand, so too will our ability to modulate them selectively.

References

1. Lipinski CA, Lombardo F, Dominy BW et al (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 46:3–26
2. Blundell T, Sibanda BL, Pearl L (1983) Three-dimensional structure, specificity and catalytic mechanism of renin. Nature 304:273–275
3. Foundling SI, Cooper J, Watson FE et al (1987) High resolution X-ray analyses of renin inhibitor-aspartic proteinase complexes. Nature 327:349–352
4. Lapatto R, Blundell T, Hemmings A et al (1989) X-ray analysis of HIV-1 proteinase at 2.7 A resolution confirms structural homology among retroviral enzymes. Nature 342:299–302
5. Dhanaraj V, Dealwis CG, Frazao C et al (1992) X-ray analyses of peptide-inhibitor complexes define the structural basis of specificity for human and mouse renins. Nature 357:466–472
6. Albiston AL, Morton CJ, Ng HL et al (2008) Identification and characterization of a new cognitive enhancer based on inhibition of insulin-regulated aminopeptidase. FASEB J 22:4209–4217

7. Ascher DB, Polekhina G, Parker MW (2012) Crystallization and preliminary X-ray diffraction analysis of human endoplasmic reticulum aminopeptidase 2. Acta Crystallogr Sect F: Struct Biol Cryst Commun 68:468–471
8. Chai SY, Yeatman HR, Parker MW et al (2008) Development of cognitive enhancers based on inhibition of insulin-regulated aminopeptidase. BMC Neurosci 9(Suppl 2):S14
9. Ye S, Chai SY, Lew RA et al (2008) Identification of modulating residues defining the catalytic cleft of insulin-regulated aminopeptidase. Biochem Cell Biol 86:251–261
10. Parker LJ, Ascher DB, Gao C et al (2012) Structural approaches to probing metal interaction with proteins. J Inorg Biochem 115:138–147
11. Parker LJ, Italiano LC, Morton CJ et al (2011) Studies of glutathione transferase P1-1 bound to a platinum(IV)-based anticancer compound reveal the molecular basis of its activation. Chemistry 17:7806–7816
12. Zhang J, Yang PL, Gray NS (2009) Targeting cancer with small molecule kinase inhibitors. Nat Rev Cancer 9:28–39
13. Congreve M, Carr R, Murray C et al (2003) A 'rule of three' for fragment-based lead discovery? Drug Discov Today 8:876–877
14. Erlanson DA, Braisted AC, Raphael DR et al (2000) Site-directed ligand discovery. Proc Natl Acad Sci U S A 97:9367–9372
15. Blundell TL, Jhoti H, Abell C (2002) High-throughput crystallography for lead discovery in drug design. Nat Rev Drug Discov 1:45–54
16. Congreve M, Murray CW, Blundell TL (2005) Structural biology and drug discovery. Drug Discov Today 10:895–907
17. Lepre CA, Moore JM, Peng JW (2004) Theory and applications of NMR-based screening in pharmaceutical research. Chem Rev 104:3641–3676
18. Murray CW, Blundell TL (2010) Structural biology in fragment-based drug design. Curr Opin Struct Biol 20:497–507
19. Cala O, Guilliere F, Krimm I (2014) NMR-based analysis of protein-ligand interactions. Anal Bioanal Chem 406:943–956
20. Hartshorn MJ, Murray CW, Cleasby A et al (2005) Fragment-based lead discovery using X-ray crystallography. J Med Chem 48:403–413
21. Caliandro R, Belviso DB, Aresta BM et al (2013) Protein crystallography and fragment-based drug design. Future Med Chem 5:1121–1140
22. Navratilova I, Hopkins AL (2010) Fragment screening by surface plasmon resonance. ACS Med Chem Lett 1:44–48
23. Navratilova I, Hopkins AL (2011) Emerging role of surface plasmon resonance in fragment-based drug discovery. Future Med Chem 3:1809–1820
24. Shepherd CA, Hopkins AL, Navratilova I (2014) Fragment screening by SPR and advanced application to GPCRs. Biol Prog Biophys Mol 116:113–123
25. Pantoliano MW, Petrella EC, Kwasnoski JD et al (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. J Biomol Screen 6:429–440
26. Lo MC, Aulabaugh A, Jin G et al (2004) Evaluation of fluorescence-based thermal shift assays for hit identification in drug discovery. Anal Biochem 332:153–159
27. Kranz JK, Schalk-Hihi C (2011) Protein thermal shifts to identify low molecular weight fragments. Methods Enzymol 493:277–298
28. Leavitt S, Freire E (2001) Direct measurement of protein binding energetics by isothermal titration calorimetry. Curr Opin Struct Biol 11:560–566
29. Gozalbes R, Carbajo RJ, Pineda-Lucena A (2010) Contributions of computational chemistry and biophysical techniques to fragment-based drug discovery. Curr Med Chem 17: 1769–1794
30. Ladbury JE, Klebe G, Freire E (2010) Adding calorimetric data to decision making in lead discovery: a hot tip. Nat Rev Drug Discov 9:23–27
31. Valkov E, Sharpe T, Marsh M et al (2012) Targeting protein-protein interactions and fragment-based drug discovery. Top Curr Chem 317:145–179

32. Whittle PJ, Blundell TL (1994) Protein structure–based drug design. Annu Rev Biophys Biomol Struct 23:349–375
33. Surade S, Blundell TL (2012) Structural biology and drug discovery of difficult targets: the limits of ligandability. Chem Biol 19:42–50
34. Blundell TL, Srinivasan N (1996) Symmetry, stability, and dynamics of multidomain and multicomponent protein systems. Proc Natl Acad Sci U S A 93:14243–14248
35. Pellegrini L, Burke DF, von Delft F et al (2000) Crystal structure of fibroblast growth factor receptor ectodomain bound to ligand and heparin. Nature 407:1029–1034
36. Harmer NJ, Ilag LL, Mulloy B et al (2004) Towards a resolution of the stoichiometry of the fibroblast growth factor (FGF)-FGF receptor-heparin complex. J Mol Biol 339:821–834
37. Robinson CJ, Harmer NJ, Goodger SJ et al (2005) Cooperative dimerization of fibroblast growth factor 1 (FGF1) upon a single heparin saccharide may drive the formation of 2:2:1 FGF1.FGFR2c.heparin ternary complexes. J Biol Chem 280:42274–42282
38. Brown A, Robinson CJ, Gallagher JT et al (2013) Cooperative heparin-mediated oligomerization of fibroblast growth factor-1 (FGF1) precedes recruitment of FGFR2 to ternary complexes. Biophys J 104:1720–1730
39. Chirgadze DY, Hepple JP, Zhou H et al (1999) Crystal structure of the NK1 fragment of HGF/SF suggests a novel mode for growth factor dimerization and receptor binding. Nat Struct Biol 6:72–79
40. Gherardi E, Youles ME, Miguel RN et al (2003) Functional map and domain structure of MET, the product of the c-met protooncogene and receptor for hepatocyte growth factor/scatter factor. Proc Natl Acad Sci U S A 100:12039–12044
41. Gherardi E, Sandin S, Petoukhov MV et al (2006) Structural basis of hepatocyte growth factor/scatter factor and MET signalling. Proc Natl Acad Sci U S A 103:4046–4051
42. Higueruelo AP, Jubb H, Blundell TL (2013) Protein-protein interactions as druggable targets: recent technological advances. Curr Opin Pharmacol 13:791–796
43. Herbert C, Schieborr U, Saxena K et al (2013) Molecular mechanism of SSR128129E, an extracellularly acting, small-molecule, allosteric inhibitor of FGF receptor signaling. Cancer Cell 23:489–501
44. Bolanos-Garcia VM, Wu Q, Ochi T et al (2012) Spatial and temporal organization of multiprotein assemblies: achieving sensitive control in information-rich cell-regulatory systems. Philos Transact A Math Phys Eng Sci 370:3023–3039
45. Sibanda BL, Critchlow SE, Begun J et al (2001) Crystal structure of an Xrcc4-DNA ligase IV complex. Nat Struct Biol 8:1015–1019
46. Sibanda BL, Chirgadze DY, Blundell TL (2010) Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. Nature 463:118–121
47. Singleton BK, Torres-Arzayus MI, Rottinghaus ST et al (1999) The C terminus of Ku80 activates the DNA-dependent protein kinase catalytic subunit. Mol Cell Biol 19:3267–3277
48. Gell D, Jackson SP (1999) Mapping of protein-protein interactions within the DNA-dependent protein kinase complex. Nucleic Acids Res 27:3494–3502
49. Nick McElhinny SA, Snowden CM, McCarville J et al (2000) Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol 20:2996–3003
50. Yano K, Morotomi-Yano K, Wang SY et al (2008) Ku recruits XLF to DNA double-strand breaks. EMBO Rep 9:91–96
51. Sasaki K, Dockerill S, Adamiak DA et al (1975) X-ray analysis of glucagon and its relationship to receptor binding. Nature 257:751–757
52. Pellegrini L, Yu DS, Lo T et al (2002) Insights into DNA recombination from the structure of a RAD51-BRCA2 complex. Nature 420:287–293
53. Dyson HJ, Wright PE (2002) Coupling of folding and binding for unstructured proteins. Curr Opin Struct Biol 12:54–60
54. Hernandez H, Robinson CV (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. Nat Protoc 2:715–726
55. Sharon M, Robinson CV (2007) The role of mass spectrometry in structure elucidation of dynamic protein complexes. Annu Rev Biochem 76:167–193

56. Ascher DB, Cromer BA, Morton CJ et al (2011) Regulation of insulin-regulated membrane aminopeptidase activity by its C-terminal domain. Biochemistry 50:2611–2622
57. Ascher DB, Wielens J, Nero TL et al (2014) Potent hepatitis C inhibitors bind directly to NS5A and reduce its affinity for RNA. Sci Rep 4:4765
58. Polekhina G, Ascher DB, Kok SF et al (2013) Structure of the N-terminal domain of human thioredoxin-interacting protein. Acta Crystallogr D Biol Crystallogr 69:333–344
59. Rambo RP, Tainer JA (2013) Accurate assessment of mass, models and resolution by small-angle scattering. Nature 496:477–481
60. Koch MH, Vachette P, Svergun DI (2003) Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. Q Rev Biophys 36:147–227
61. Putnam CD, Hammel M, Hura GL et al (2007) X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. Q Rev Biophys 40:191–285
62. Hammel M, Yu Y, Fang S et al (2010) XLF regulates filament architecture of the XRCC4.ligase IV complex. Structure 18:1431–1442
63. Wong W, Bai XC, Brown A et al (2014) Cryo-EM structure of the Plasmodium falciparum 80S ribosome bound to the anti-protozoan drug emetine. Elife 3:e03080
64. Davis AJ, Chen BP, Chen DJ (2014) DNA-PK: a dynamic enzyme in a versatile DSB repair pathway. DNA Repair (Amst) 17:21–29
65. Yang H, Rudge DG, Koos JD et al (2013) mTOR kinase structure, mechanism and regulation. Nature 497:217–223
66. Boskovic J, Rivera-Calzada A, Maman JD et al (2003) Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. EMBO J 22:5875–5882
67. Chiu CY, Cary RB, Chen DJ et al (1998) Cryo-EM imaging of the catalytic subunit of the DNA-dependent protein kinase. J Mol Biol 284:1075–1081
68. Leuther KK, Hammarsten O, Kornberg RD et al (1999) Structure of DNA-dependent protein kinase: implications for its regulation by DNA. EMBO J 18:1114–1123
69. Williams DR, Lee KJ, Shi J et al (2008) Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding. Structure 16:468–477
70. Hammel M, Yu Y, Mahaney BL et al (2010) Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. J Biol Chem 285:1414–1423
71. Morris EP, Rivera-Calzada A, da Fonseca PC et al (2011) Evidence for a remodelling of DNA-PK upon autophosphorylation from electron microscopy studies. Nucleic Acids Res 39:5757–5767
72. Spagnolo L, Rivera-Calzada A, Pearl LH et al (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. Mol Cell 22:511–519
73. Cary RB, Peterson SR, Wang J et al (1997) DNA looping by Ku and the DNA-dependent protein kinase. Proc Natl Acad Sci U S A 94:4267–4272
74. Yaneva M, Kowalewski T, Lieber MR (1997) Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. EMBO J 16: 5098–5112
75. Ochi T, Wu Q, Blundell TL (2014) The spatial organization of non-homologous end joining: from bridging to end joining. DNA Repair (Amst) 17:98–109
76. Andres SN, Vergnes A, Ristic D et al (2012) A human XRCC4-XLF complex bridges DNA. Nucleic Acids Res 40:1868–1878
77. Hammel M, Rey M, Yu Y et al (2011) XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. J Biol Chem 286:32638–32650
78. Ropars V, Drevet P, Legrand P et al (2011) Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. Proc Natl Acad Sci U S A 108:12663–12668

79. Wu Q, Ochi T, Matak-Vinkovic D et al (2011) Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. Biochem Soc Trans 39: 1387–1392, suppl 2 p following 1392
80. Ochi T, Wu Q, Chirgadze DY et al (2012) Structural insights into the role of domain flexibility in human DNA ligase IV. Structure 20:1212–1222
81. Williams GJ, Hammel M, Radhakrishnan SK et al (2014) Structural insights into NHEJ: building up an integrated picture of the dynamic DSB repair super complex, one component and interaction at a time. DNA Repair (Amst) 17:110–120
82. Critchlow SE, Bowater RP, Jackson SP (1997) Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. Curr Biol 7:588–598
83. Grawunder U, Zimmer D, Lieber MR (1998) DNA ligase IV binds to XRCC4 via a motif located between rather than within its BRCT domains. Curr Biol 8:873–876
84. Recuero-Checa MA, Dore AS, Arias-Palomo E et al (2009) Electron microscopy of Xrcc4 and the DNA ligase IV-Xrcc4 DNA repair complex. DNA Repair (Amst) 8:1380–1389
85. Ochi T, Gu X, Blundell TL (2013) Structure of the catalytic region of DNA ligase IV in complex with an Artemis fragment sheds light on double-strand break repair. Structure 21:672–679
86. Berman HM, Westbrook J, Feng Z et al (2000) The protein data bank. Nucleic Acids Res 28:235–242
87. Dou Y, Baisnee PF, Pollastri G et al (2004) ICBS: a database of interactions between protein chains mediated by beta-sheet formation. Bioinformatics 20:2767–2777
88. Lo A, Cheng CW, Chiu YY et al (2011) TMPad: an integrated structural database for helix-packing folds in transmembrane proteins. Nucleic Acids Res 39:D347–D355
89. Mosca R, Ceol A, Stein A et al (2014) 3did: a catalog of domain-based interactions of known three-dimensional structure. Nucleic Acids Res 42:D374–D379
90. Huang Z, Zhu L, Cao Y et al (2011) ASD: a comprehensive database of allosteric proteins and modulators. Nucleic Acids Res 39:D663–D669
91. Huang Z, Mou L, Shen Q et al (2014) ASD v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42:D510–D516
92. Bickerton GR, Higueruelo AP, Blundell TL (2011) Comprehensive, atomic-level characterization of structurally characterized protein-protein interactions: the PICCOLO database. BMC Bioinf 12:313
93. Lee S, Blundell TL (2009) BIPA: a database for protein-nucleic acid interaction in 3D structures. Bioinformatics 25:1559–1560
94. Schreyer A, Blundell T (2009) CREDO: a protein-ligand interaction database for drug discovery. Chem Biol Drug Des 73:157–167
95. Schreyer AM, Blundell TL (2013) CREDO: a structural interactomics database for drug discovery. Database (Oxford) 2013:bat049
96. Hamosh A, Scott AF, Amberger JS et al (2005) Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res 33:D514–D517
97. Forbes SA, Bhamra G, Bamford S et al (2008) The catalogue of somatic mutations in cancer (COSMIC). Curr Protoc Hum Genet, Chapter 10:Unit 10 11
98. Apweiler R, Bairoch A, Wu CH et al (2004) UniProt: the Universal Protein knowledgebase. Nucleic Acids Res 32:D115–D119
99. Hubbard T, Barker D, Birney E et al (2002) The Ensembl genome database project. Nucleic Acids Res 30:38–41
100. Gaulton A, Bellis LJ, Bento AP et al (2011) ChEMBL: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 40:D1100–D1107
101. Lewell XQ, Judd DB, Watson SP et al (1998) RECAP-Retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. J Chem Inf Comput Sci 38:511–522
102. Deng Z, Chuaqui C, Singh J (2004) Structural interaction fingerprint (SIFt): a novel method for analyzing three-dimensional protein-ligand binding interactions. J Med Chem 47: 337–344

103. Blundell TL, Sibanda BL, Montalvao RW et al (2006) Structural biology and bioinformatics in drug design: opportunities and challenges for target identification and lead discovery. Philos Trans R Soc Lond B Biol Sci 361:413–423
104. Nair SK, Burley SK (2003) X-ray structures of Myc-Max and Mad-Max recognizing DNA. Molecular bases of regulation by proto-oncogenic transcription factors. Cell 112:193–205
105. Fletcher S, Hamilton AD (2006) Targeting protein-protein interactions by rational design: mimicry of protein surfaces. J R Soc Interface 3:215–233
106. Jones S, Thornton JM (1996) Principles of protein-protein interactions. Proc Natl Acad Sci U S A 93:13–20
107. Arkin MR, Tang Y, Wells JA (2014) Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. Chem Biol 21:1102–1114
108. Cooper A (1999) Thermodynamic analysis of biomolecular interactions. Curr Opin Chem Biol 3:557–563
109. Breiten B, Lockett MR, Sherman W et al (2013) Water networks contribute to enthalpy/entropy compensation in protein-ligand binding. J Am Chem Soc 135:15579–15584
110. Fuller JC, Burgoyne NJ, Jackson RM (2009) Predicting druggable binding sites at the protein-protein interface. Drug Discov Today 14:155–161
111. Li X, Keskin O, Ma B et al (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 344:781–795
112. Clackson T, Wells JA (1995) A hot spot of binding energy in a hormone-receptor interface. Science 267:383–386
113. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280:1–9
114. Rajamani D, Thiel S, Vajda S et al (2004) Anchor residues in protein-protein interactions. Proc Natl Acad Sci U S A 101:11287–11292
115. Ben-Shimon A, Eisenstein M (2010) Computational mapping of anchoring spots on protein surfaces. J Mol Biol 402:259–277
116. Meireles LM, Domling AS, Camacho CJ (2010) ANCHOR: a web server and database for analysis of protein-protein interaction binding pockets for drug discovery. Nucleic Acids Res 38:W407–W411
117. Koes DR, Camacho CJ (2012) PocketQuery: protein-protein interaction inhibitor starting points from protein-protein interaction structure. Nucleic Acids Res 40:W387–W392
118. London N, Movshovitz-Attias D, Schueler-Furman O (2010) The structural basis of peptide-protein binding strategies. Structure 18:188–199
119. London N, Raveh B, Movshovitz-Attias D et al (2010) Can self-inhibitory peptides be derived from the interfaces of globular protein-protein interactions? Proteins 78:3140–3149
120. London N, Raveh B, Schueler-Furman O (2013) Druggable protein-protein interactions–from hot spots to hot segments. Curr Opin Chem Biol 17:952–959
121. Pommier Y, Marchand C (2012) Interfacial inhibitors: targeting macromolecular complexes. Nat Rev Drug Discov 11:25–36
122. Gao M, Skolnick J (2012) The distribution of ligand-binding pockets around protein-protein interfaces suggests a general mechanism for pocket formation. Proc Natl Acad Sci U S A 109:3784–3789
123. Topham CM, Srinivasan N, Blundell TL (1997) Prediction of the stability of protein mutants based on structural environment-dependent amino acid substitution and propensity tables. Protein Eng 10:7–21
124. Worth CL, Preissner R, Blundell TL (2011) SDM–a server for predicting effects of mutations on protein stability and malfunction. Nucleic Acids Res 39:W215–W222
125. Overington J, Donnelly D, Johnson MS et al (1992) Environment-specific amino acid substitution tables: tertiary templates and prediction of protein folds. Protein Sci 1:216–226
126. Kucukkal TG, Yang Y, Chapman SC et al (2014) Computational and experimental approaches to reveal the effects of single nucleotide polymorphisms with respect to disease diagnostics. Int J Mol Sci 15:9670–9717

127. Pires DE, Ascher DB, Blundell TL (2014) mCSM: predicting the effects of mutations in proteins using graph-based signatures. Bioinformatics 30:335–342
128. da Silveira CH, Pires DE, Minardi RC et al (2009) Protein cutoff scanning: a comparative analysis of cutoff dependent and cutoff free methods for prospecting contacts in proteins. Proteins 74:727–743
129. Pires DE, de Melo-Minardi RC, Ados Santos M et al (2011) Cutoff Scanning Matrix (CSM): structural classification and function prediction by protein inter-residue distance patterns. BMC Genomics 12(Suppl 4):S12
130. Pires DE, de Melo-Minardi RC, da Silveira CH et al (2013) aCSM: noise-free graph-based signatures to large-scale receptor-based ligand prediction. Bioinformatics 29:855–861
131. Guerois R, Nielsen JE, Serrano L (2002) Predicting changes in the stability of proteins and protein complexes: a study of more than 1000 mutations. J Mol Biol 320:369–387
132. Kortemme T, Baker D (2002) A simple physical model for binding energy hot spots in protein-protein complexes. Proc Natl Acad Sci U S A 99:14116–14121
133. Kortemme T, Kim DE, Baker D (2004) Computational alanine scanning of protein-protein interfaces. Sci STKE 2004:pl2
134. Huo S, Massova I, Kollman PA (2002) Computational alanine scanning of the 1:1 human growth hormone-receptor complex. J Comput Chem 23:15–27
135. Gouda H, Kuntz ID, Case DA et al (2003) Free energy calculations for theophylline binding to an RNA aptamer: comparison of MM-PBSA and thermodynamic integration methods. Biopolymers 68:16–34
136. Kollman PA, Massova I, Reyes C et al (2000) Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. Acc Chem Res 33:889–897
137. Kollman P (1993) Free energy calculations: applications to chemical and biochemical phenomena. Chem Rev 93:2395–2417
138. Kong X, Brooks C (1996) Lambda-dynamics: a new approach to free energy calculations. J Chem Phys 105:2414–2423
139. Moreira I, Fernandes P, Ramos M (2007) Unravelling Hot Spots: a comprehensive computational mutagenesis study. Theor Chem Accounts 117:99–113
140. Moreira IS, Fernandes PA, Ramos MJ (2005) Accuracy of the numerical solution of the Poisson–Boltzmann equation. J Mol Struct THEOCHEM 729:11–18
141. Massova I, Kollman PA (1999) Computational alanine scanning to probe protein–protein interactions: a novel approach to evaluate binding free energies. J Am Chem Soc 121: 8133–8143
142. Moal IH, Fernandez-Recio J (2012) SKEMPI: a structural kinetic and energetic database of mutant protein interactions and its use in empirical models. Bioinformatics 28:2600–2607
143. Dehouck Y, Kwasigroch JM, Rooman M et al (2013) BeAtMuSiC: prediction of changes in protein-protein binding affinity on mutations. Nucleic Acids Res 41:W333–W339
144. Dourado DF, Flores SC (2014) A multiscale approach to predicting affinity changes in protein-protein interfaces. Proteins 82:2681–2690
145. Li M, Petukh M, Alexov E et al (2014) Predicting the impact of missense mutations on protein-protein binding affinity. J Chem Theory Comput 10:1770–1780
146. Moal IH, Fernandez-Recio J (2013) Intermolecular contact potentials for protein–protein interactions extracted from binding free energy changes upon mutation. J Chem Theory Comput 9:3715–3727
147. Gossage L, Pires DE, Olivera-Nappa A et al (2014) An integrated computational approach can classify VHL missense mutations according to risk of clear cell renal carcinoma. Hum Mol Genet 23:5976–5988
148. Pires DE, Ascher DB, Blundell TL (2014) DUET: a server for predicting effects of mutations on protein stability using an integrated computational approach. Nucleic Acids Res 42: W314–W319

149. Fletcher S, Hamilton AD (2005) Protein surface recognition and proteomimetics: mimics of protein surface structure and function. Curr Opin Chem Biol 9:632–638
150. Wilson AJ (2009) Inhibition of protein-protein interactions using designed molecules. Chem Soc Rev 38:3289–3300
151. Buerger C, Groner B (2003) Bifunctional recombinant proteins in cancer therapy: cell penetrating peptide aptamers as inhibitors of growth factor signaling. J Cancer Res Clin Oncol 129:669–675
152. Seidah NG (2013) Proprotein convertase subtilisin kexin 9 (PCSK9) inhibitors in the treatment of hypercholesterolemia and other pathologies. Curr Pharm Des 19:3161–3172
153. Traczewski P, Rudnicka L (2011) Treatment of systemic lupus erythematosus with epratuzumab. Br J Clin Pharmacol 71:175–182
154. Nord K, Gunneriusson E, Ringdahl J et al (1997) Binding proteins selected from combinatorial libraries of an alpha-helical bacterial receptor domain. Nat Biotechnol 15:772–777
155. Kaminskas LM, Ascher DB, McLeod VM et al (2013) PEGylation of interferon alpha2 improves lymphatic exposure after subcutaneous and intravenous administration and improves antitumour efficacy against lymphatic breast cancer metastases. J Control Release 168:200–208
156. Basse MJ, Betzi S, Bourgeas R et al (2013) 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. Nucleic Acids Res 41:D824–D827
157. Higueruelo AP, Jubb H, Blundell TL (2013) TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. Database (Oxford) 2013:bat039
158. Higueruelo AP, Schreyer A, Bickerton GRJ et al (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. Chem Biol Drug Des 74:457–467
159. Hann MM (2011) Molecular obesity, potency and other addictions in drug discovery. MedChemComm 2:349–355
160. Higueruelo AP, Schreyer A, Bickerton GRJ et al (2012) What can we learn from the evolution of protein-ligand interactions to aid the design of new therapeutics? PLoS One 7:e51742
161. Arkin MR, Randal M, DeLano WL et al (2003) Binding of small molecules to an adaptive protein-protein interface. Proc Natl Acad Sci U S A 100:1603–1608
162. Sauve K, Nachman M, Spence C et al (1991) Localization in human interleukin 2 of the binding site to the alpha chain (p55) of the interleukin 2 receptor. Proc Natl Acad Sci U S A 88:4636–4640
163. Zurawski SM, Vega F Jr, Doyle EL et al (1993) Definition and spatial location of mouse interleukin-2 residues that interact with its heterotrimeric receptor. EMBO J 12:5113–5119
164. Scott DE, Ehebauer MT, Pukala T et al (2013) Using a fragment-based approach to target protein-protein interactions. Chembiochem 14:332–342
165. Wyatt PG, Woodhead AJ, Berdini V et al (2008) Identification of N-(4-piperidinyl)-4-(2,6-dichlorobenzoylamino)-1H-pyrazole-3-carboxamide (AT7519), a novel cyclin dependent kinase inhibitor using fragment-based X-ray crystallography and structure based drug design. J Med Chem 51:4986–4999
166. Howard S, Berdini V, Boulstridge JA et al (2009) Fragment-based discovery of the pyrazol-4-yl urea (AT9283), a multitargeted kinase inhibitor with potent aurora kinase activity. J Med Chem 52:379–388
167. Frederickson M, Callaghan O, Chessari G et al (2008) Fragment-based discovery of mexiletine derivatives as orally bioavailable inhibitors of urokinase-type plasminogen activator. J Med Chem 51:183–186
168. Antonysamy SS, Aubol B, Blaney J et al (2008) Fragment-based discovery of hepatitis C virus NS5b RNA polymerase inhibitors. Bioorg Med Chem Lett 18:2990–29995
169. Antonysamy S, Hirst G, Park F et al (2009) Fragment-based discovery of JAK-2 inhibitors. Bioorg Med Chem Lett 19:279–282

170. Edfeldt FNB, Folmer RHA, Breeze AL (2011) Fragment screening to predict druggability (ligandability) and lead discovery success. Drug Discov Today 16:284–287
171. Winter A, Higueruelo P et al (2012) Biophysical and computational fragment-based approaches to targeting protein-protein interactions: applications in structure-guided drug discovery. Q Rev Biophys 45:383–426
172. Jubb H, Higueruelo A, Winter A et al (2012) Structural biology and drug discovery for protein–protein interactions. Trends Pharmacol Sci 33:241–248
173. Keserü GM, Makara GM (2009) The influence of lead discovery strategies on the properties of drug candidates. Nat Rev Drug Discov 8:203–212