OPEN ACCESS Freely available online

PLoS COMPUTATIONAL BIOLOGY

Education

Deciphering Protein–Protein Interactions. Part I. Experimental Techniques and Databases

Benjamin A. Shoemaker, Anna R. Panchenko*

A Tutorial in PLoS Computional Biology

Proteins interact with each other in a highly specific manner, and protein interactions play a key role in many cellular processes; in particular, the distortion of protein interfaces may lead to the development of many diseases. To understand the mechanisms of protein recognition at the molecular level and to unravel the global picture of protein interactions in the cell, different experimental techniques have been developed. Some methods characterize individual protein interactions while others are advanced for screening interactions on a genome-wide scale. In this review we describe different experimental techniques of protein interaction identification together with various databases which attempt to classify the large array of experimental data. We discuss the main promises and pitfalls of different methods and present several approaches to verify and validate the diverse experimental data produced by high-throughput techniques.

Introduction

It is now becoming clear that protein interactions determine the outcome of most cellular processes [1–4]. Therefore, identifying and characterizing protein–protein interactions and their networks is essential for understanding the mechanisms of biological processes on a molecular level. Despite the fact that protein interactions are remarkably diverse, all protein interfaces share certain common properties. Protein interactions can be classified into different types depending on their strength (permanent and transient), specificity (specific or nonspecific), the location of interacting partners within one or on two polypeptide chains, and the similarity between interacting subunits (homo- and hetero-oligomers). It has been shown that interface types are significantly different in amino acid composition so that it is possible to predict the type of interaction interface from amino acid composition alone [5]. Earlier structural analysis of interfaces showed that most interfaces consist of completely buried cores surrounded by partially accessible rims [6,7] with the overall size of about 1600 ± 400 Å^2 (a "standard size" patch) [8]. It has been found that certain amino acids are preferred on protein interfaces and that the amino acid composition of the core differs considerably from the rim [6,7,9,10]. More recent models suggested that the protein binding site consists of a few independent highly packed regions, so called "hot spots," which contribute significantly to the free energy of binding [11–13]. Hot spots were found to be structurally conserved [14], and the energetics of interactions at the hot spots have been analyzed in several studies [15–18].

In many cellular processes, proteins recognize specific targets and bind them in a highly regular manner. The specificity of interactions in these cases is determined by structural and physico–chemical properties of two interacting proteins. As a result, there should be a certain degree of conservation in the interaction patterns between similar proteins and domains. Indeed, it has been found that close homologs almost always interact in the same way and protein–protein interactions place certain evolutionary constraints on protein sequence and structural divergence [19–24]. Recent studies confirm that the total number of interaction types or modes is limited and rather small [25–27]. On the other hand, remotely related proteins/domains can have different interaction modes [21,26,28]; and the conservation of such protein interfaces is similar to the average conservation of rest of the protein [29–32].

In this review and its companion review in the April issue [33], we attempt to classify and systemize the array of experimental and theoretical data on the identification and prediction of protein interactions. In this review we focus on the generic experimental techniques for identifying protein interactions and the databases storing the information

Editor: Fran Lewitter, Whitehead Institute, United States of America

Citation: Shoemaker BA, Panchenko, AR (2007) Deciphering protein–protein interactions. Part I. Experimental techniques and databases. PLoS Comput Biol 3(3): e42. doi:10.1371/journal.pcbi.0030042

Abbreviations: AD, domain that activates transcription; BD, domain that directs binding to a promoter DNA sequence; BIND, Biomolecular Interaction Network Database; CBM, Conserved Binding Mode database; DIP, database of interacting proteins; MS, mass spectroscopy; TAP, tandem affinity purification; Y2H, yeast two-hybrid

Benjamin A. Shoemaker and Anna R. Panchenko are with the Computational Biology Branch of the National Center for Biotechnology Information in Bethesda, Maryland, United States of America.

* To whom correspondence should be addressed. E-mail: panch@ncbi.nlm.nih.gov

Table 1. Different Experimental Methods Measuring Protein Interactions

Method	High-Throughput Approach	Living Cell Assay	Type of Interactions	Type of Characterization
Y2H [47,48]	+	In vivo	Physical interactions (binary)	Identification
Affinity purification–MS [61]	+	In vitro	Physical interactions (complex)	Identification
DNA microarrays/Gene coexpression [113]	+	In vitro	Functional association	Identification
Protein microarrays [114–116]	+	In vitro	Physical interaction (complex)	Identification
Synthetic lethality [85,86]	+	In vivo	Functional association	Identification
Phage display [117]	+	In vitro	Physical interaction (complex)	Identification
X-ray crystallography, NMR spectroscopy [84]	−	In vitro	Physical interactions (complex)	Structural and biological characterization
Fluorescence resonance energy transfer [89]	−	In vivo	Physical interaction (binary)	Biological characterization
Surface plasmon resonance [91]	−	In vitro	Physical interaction (complex)	Kinetic, dynamic characterization
Atomic force microscopy [93]	−	In vitro	Physical interaction (binary)	Mechanical, dynamic characterization
Electron microscopy [118]	−	In vitro	Physical interaction (complex)	Structural and biological characterization

High-throughput techniques are indicated with pluses (second column), and those which can provide information on interactions in vivo are shown in the third column. Fourth column indicates whether the method supplies data on physically interacting proteins in a complex ("complex") or only pairwise interactions ("binary"). Methods inferring interactions through functional association are shown as well. The type of protein interaction characterization is shown in the last column.

doi:10.1371/journal.ppat.0030042.t001

obtained from these experiments. In the second review, we present different methods to predict protein and domain interactions and discuss various challenges faced in this field with respect to limited prediction accuracy.

Experimental Methods for Identifying and Characterizing Protein Interactions

Protein interactions can be analyzed by different genetic, biochemical, and physical methods, which are listed in Table 1 and shown in Figure 1. Some techniques enable screening of a large number of proteins in a cell, such as yeast two-hybrid (Y2H), tandem affinity purification (TAP), mass spectroscopy (MS), DNA and protein microarrays, synthetic lethality, and phage display. Other methods focus on monitoring and characterizing specific biochemical and physico–chemical properties of a protein complex.

Yeast two-hybrid method. The development of the Y2H technique has considerably accelerated the screening of protein interactions in vivo. Y2H is based on the fact that many eukaryotic transcription activators have at least two distinct domains, one that directs binding to a promoter DNA sequence (BD) and another that activates transcription (AD) (Figure 1A). It was demonstrated that splitting BD and AD inactivates the transcription, but the transcription can be restored if a DNA-binding domain is physically (not necessarily covalently) associated with an activating domain [34]. According to the Y2H method, a protein of interest is fused to BD (bait). This chimeric protein is cloned in an expression plasmid, which is then transfected into a yeast cell. A similar procedure creates a chimeric sequence of another protein fused to AD (prey). If two proteins physically interact, the reporter gene is activated. The most broadly used Y2H systems are GAL4/LexA-based, where the GAL4 protein controls in yeast the expression of the LacZ gene encoding beta-galactosidase. Numerous variations of Y2H have been developed including systems with several reporter genes, one-hybrid and three-hybrid systems for identifying proteins interactions with DNA and RNA [35–38], systems for detecting interactions in mammalian and prokaryotic cells, and systems for screening the interactions between membrane proteins [39–43].

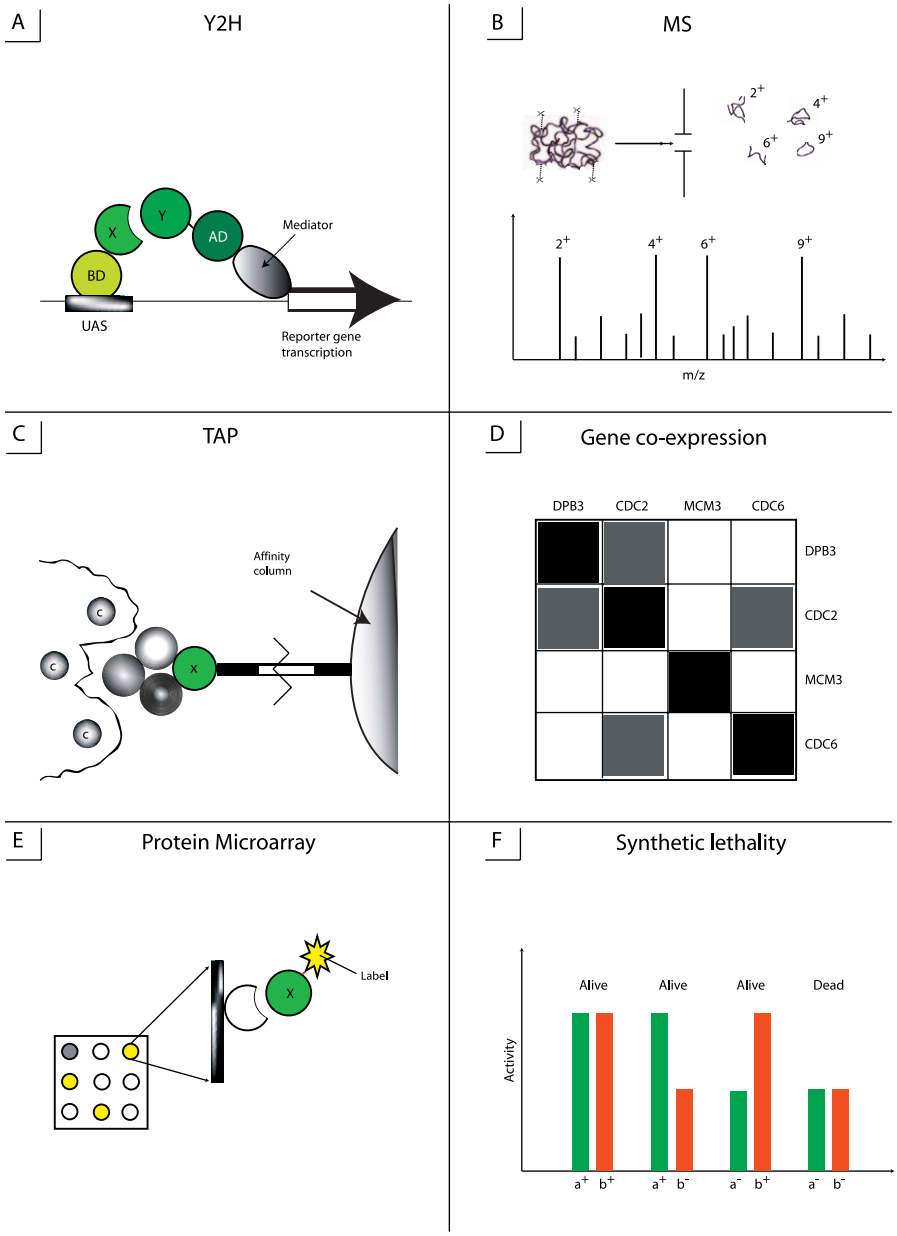
For screening entire genomes, the Y2H method has been advanced into two main approaches [44–46]: matrix-based and library-based.

In the *matrix approach,* a matrix of prey clones is created where each clone expresses a particular prey protein in one well of a plate. Then each bait strain is mated with an array of prey strains and those diploids where two chimeric proteins interact are selected based on the expression of a reporter gene and the position on a plate.

In the *library approach,* each bait is screened against an undefined prey library containing random cDNA fragments or open reading frames (ORFs). Diploid positives are selected based on their ability to grow on specific substrates; and interacting proteins are determined by DNA sequencing. The first two genome-wide analyses of the yeast "interactome" revealed 692 and 841 putative interactions, respectively [47,48]. The overlap between these two experimental studies was quite small; both methods shared only 141 interactions, about 20% of the interaction data [48]. Recently, Y2H has been used to identify interactions in worm [2], fly [1], and human [49,50].

The small overlap between Y2H experiments can be explained by different factors, among them: differences in protein interaction sampling, Y2H bias towards nonspecific interactions [51], and limitations of the Y2H method itself. For example, proteins initiating transcription by themselves cannot be targeted in Y2H experiments; and the use of sequence chimeras can impose difficulties since fusion can change the structure of a target protein. In addition, protein folding and posttranslational modifications can differ between yeast and other organisms. This makes it difficult to screen proteins from mammalian and prokaryotic cells using Y2H as well as cytoplasmic and membrane proteins. To validate the quality of Y2H protein interactions in vivo, different in vitro techniques can be used.

Mass spectroscopy. MS is a powerful method of studying macromolecular interactions in vitro. The principle of the MS method is to produce ions which can be detected based



doi:10.1371/journal.ppat.0030042.g001

Figure 1. Schematic Representations of Main Experimental Techniques Used for High-Throughput Analysis of Protein Interactions

(A) Y2H detects interactions between proteins X and Y, where X is linked to BD domain which binds to upstream activating sequence (UAS) of a promoter.

(B) MS identifies polypeptide sequence.

(C) TAP purifies protein complexes and removes the molecules of contaminants.

(D) Gene coexpression analysis produces the correlation matrix where the dark areas show high correlation between expression levels of corresponding genes.

(E) Protein microarrays (protein chips) can detect interactions between actual proteins rather than genes: target proteins immobilized on the solid support are probed with a fluorescently labeled protein.

(F) Synthetic lethality method describes the genetic interaction when two individual, nonlethal mutations result in lethality when administered together ($a^- b^-$).

on their mass-to-charge ratios, thereby allowing the identification of polypeptide sequences [36,52,53] (Figure 1B). The problem of converting protein/peptide molecules from the condensed phase into ions in the gas phase is solved by using Electrospray Ionization (ESI) [54] and Matrix Assisted Laser Desorption Ionization (MALDI) [55,56]. Different algorithms have been developed to analyze mass spectra and to identify proteins by their sequence [57–60]. Some of them find correlations between theoretical and experimental spectra while others use de novo algorithms to infer peptide sequences from theoretical interpretation of the mass spectra. Despite the usefulness of MS for the characterization

of interacting proteins, purification of protein complexes turns out to be the limiting step of their identification. To address this, TAP has been developed.

TAP method of complex purification. A TAP tag consists of two IgG binding domains of *Staphylococcus* protein A and a calmodulin binding peptide separated by the tobacco etch virus protease cleavage site [61,62] (Figure 1C). A target protein open reading frame (ORF) is fused with the DNA sequences encoding the TAP tag and is expressed in yeast where it can form native complexes with other proteins. At the first step of the TAP purification, protein A binds tightly to an IgG matrix; and after washing out the contaminants, the protease cleaves the link between protein A and IgG matrix. The eluate of this first step is then incubated with calmodulin-coated beads in the presence of calcium. After washing, the target protein complex is released. The components of each complex are screened by polyacrylamide gel electrophoresis, cleaved by proteases, and the fragments are identified by MS. Comparing Y2H and TAP–MS, it should be noted that both methods generate a lot of false positives and miss a lot of known interactions. Y2H has the advantages of being an in vivo technique and of detecting transient interactions. In contrast, TAP–MS can report on higher-order interactions beyond binary and, therefore, provides direct information on protein complexes.

Several large-scale studies of protein complexes have been performed using TAP–MS and Y2H methods [2,4,63,64]. For example, Krogan et al. showed that 7,123 protein interactions identified with high confidence in yeast can be clustered into 547 protein complexes [3].

Gene co-expression. Since the function of a protein complex depends on the functionality of all subunits, subunits should be present in stoichiometric amounts and gene expression levels of subunits in a complex should be related. Gene expression profiles can be provided, for example, from cell cycle experiments and expression levels of a gene under different conditions. Expression profile similarity can be calculated as a correlation coefficient between relative expression levels of two genes/proteins or the normalized difference between their absolute expression levels or calculated using other methods [65–69] (Figure 1D). The distribution of these quantities for target proteins then can be compared with the distributions for random noninteracting protein pairs. It was shown that the most obvious coexpression comes from permanent complexes such as ribosome and proteasome [65]. Several studies have tackled the problem of gene co-expression and demonstrated that interacting proteins in yeast are more likely to have their genes coexpressed compared with noninteracting proteins [65,70–77]. Moreover, it was shown that expression levels of physically interacting proteins coevolve, and coevolution of gene expression can be a better predictor of protein interactions than coevolution of amino acid sequences [78]. To infer the interactions between the genes, the DNA microarray methodology can be successfully used in the conjunction with the synthetic lethality method.

Synthetic lethality method. It is not very well-understood how genetic variation influences phenotype and how genes interact with each other producing different phenotypes in different strains of the same species [77,78]. These problems can be addressed by using various genetic interaction methods, the most common of which is the synthetic lethality method (Figure 1F). The synthetic lethality method produces mutations or deletions of two separate genes which are viable alone but cause lethality when combined together in a cell under certain conditions [78–83]. Since these mutations are lethal, they cannot be isolated directly and should be synthetically constructed. Synthetic interaction can point to the possible physical interaction between two gene products, their participation in a single pathway, or a similar function. For example, synthetic lethality experiments enabled the prediction of the unknown function of the YLL049W gene as belonging to the dynein–dynactin pathway, and the bridging together of the two pathways of the parallel mitotic exit network and the Cdc14 early anaphase release pathway [83].

Monitoring specific protein interactions. The most detailed information about protein interaction interfaces at the atomic level can be provided by X-ray crystallography and NMR spectroscopy, but the number of solved protein complexes remains low [84]. At the same time, the real-time characterization of interacting proteins in vivo can be achieved with various spectroscopic techniques requiring the attachment of a spectroscopic label to a target protein [87,88] (Table 1). A powerful technique in this respect is fluorescence resonance energy transfer (FRET), which can occur only if two fluorophores are located close to each other [89]. Another effective method, surface plasmon resonance (SPR), does not require spectroscopic labeling and can detect interactions between soluble ligands and immobilized receptors [90,91]; while the isothermal titration calorimetry (ITC) technique allows for direct measurement of the enthalpy of binding [92]. Recently, new methods have been developed to analyze protein interactions at the single-molecule level. For example, atomic force microscopy can fairly accurately measure interaction forces ([93]) while fluorescence techniques can characterize conformational changes in proteins upon binding [94].

Protein interaction networks derived from experiments. The fast development of experimental techniques for protein interactions has enabled the construction and systematic analysis of interaction networks [1,2,95]. Interaction maps obtained for one species can be used to predict interaction networks in other species, to identify functions of unknown proteins, and to get insight into the evolution of protein interaction patterns. The interaction map analyses and comparisons are based on the observation that many interactions are conserved among species ("interologs") [46]. Sequence-based searches for "interologs" were able to identify 16%–31% of true "interologs" (tested using Y2H system) even between remotely related species such as yeast and worm [96]. Analysis of conservation in the networks produced by gene co-expression data revealed that interologs correspond to the functionally related genes responsible for core biological processes [77]. Moreover, a multiple-species network has been constructed by identifying pairs of genes with correlated expression in different organisms. A multiple-species network has shown to perform better than a single-species network in linking together functionally related genes.

Verification of protein interactions. Validation of protein interaction data is difficult; except for small datasets on protein interactions provided by the Protein Data Bank (PDB) [84] and the Munich Information Center for Protein Sequences (MIPS) [97], there is no comprehensive gold

Table 2. Databases Available for Searching and/or Downloading Data Related to Protein Interactions

Database	Proteins/Domains	Type	Number of Interactions
DIP[a], LiveDIP	P	E,S	55,733
BIND[a]	P	E,C,S	83,517
MPact/MIPS[a]	P	E,C,F	15,488 (4,300)[b]
STRING	P	E,P,F	730,000 (proteins)
MINT[a]	P	E,C	71,854
IntAct[a]	P	E,C	68,165
BioGRID[a]	P	E,C	116,000 (30,000)[b]
HPRD	P	E,C	33,710
ProtCom	P,D	S,H	1,770
3did, Interprets	D	S,H	3,304
Pibase, ModBase	D	S,H	2,387
CBM	D	S	2,784
SCOPPI	D	S	3,358
iPfam	D	S	3,019
InterDom	D	P	30,037
DIMA	D	F,S	—
Prolinks	P	F	—

Listed are: the name of the database; the unit of interaction, protein (P) or domain (D); type of data (high-throughput experimental data (E), structural data (S), manual curation (C), functional predictions (F), and interface homology modeling (H)); and the number of interactions.

[a]Databases are members of the International Molecular Exchange Consortium (IMEx) (http://imex.sourceforge.net).

[b]Number of interactions listed in parentheses is for curated set.

doi:10.1371/journal.ppat.0030042.t002

Table 3. URLs and Primary Citations for Protein Interaction–Related Databases

Database	URL/FTP
DIP [102], LiveDIP[103]	http://dip.doe-mbi.ucla.edu
BIND [105]	http://bind.ca
MPact/MIPS [97]	http://mips.gsf.de/services/ppi
STRING [119]	http://string.embl.de
MINT [120]	http://mint.bio.uniroma2.it/mint
IntAct [121]	http://www.ebi.ac.uk/intact
BioGRID [122]	http://www.thebiogrid.org
HPRD [123]	http://www.hprd.org
ProtCom [124]	http://www.ces.clemson.edu/compbio/ProtCom
3did [108], Interprets[125]	http://gatealoy.pcb.ub.es/3did/
Pibase [107], ModBase [126]	http://alto.compbio.ucsf.edu/pibase
CBM [26]	ftp://ftp.ncbi.nlm.nih.gov/pub/cbm
SCOPPI [111]	http://www.scoppi.org/
iPfam [127]	http://www.sanger.ac.uk/Software/Pfam/iPfam
InterDom [128]	http://interdom.lit.org.sg
DIMA [129]	http://mips.gsf.de/genre/proj/dima/index.html
Prolinks [104]	http://prolinks.doe-mbi.ucla.edu/cgi-bin/functionator/pronav/

doi:10.1371/journal.ppat.0030042.t003

standard interaction set. Several methods have been proposed for verification of protein interaction data [66,67,76,98,99], and some of them are described here.

Expression profile reliability method (EPR) [66] is based on the observation that interacting proteins are coexpressed. Two distributions of expression distances are defined for noninteracting and reliably interacting proteins. The distribution of expression distances for a protein set of interest is assumed to be a linear combination of two predefined distributions with the linear coefficient that characterizes the accuracy of a given dataset.

Paralogous verification method (PVM) [66] is based on the observation that if two proteins interact, their paralogs most likely interact. It gives more reliability to the interaction of two families that contain a greater number of interactions between paralogous proteins. This method identified ~40% true interactions at a 1% error rate.

Protein localization method (PLM) [98] defines true positives as interacting proteins that are localized in the same cellular compartment and/or interacting proteins that are annotated to have a common cellular role. PLM showed that the accuracy of experimental data strongly depends on the method with up to 50% true positives detected in Y2H experiments and up to 100% true positives detected in immunoprecipitation experiments [100].

Protein and domain interaction databases. A large variety of databases exists to study binary protein interactions and the higher order interactions in protein complexes. A summary of some available databases is given in Tables 2 and 3. Different databases contain interactions obtained by direct submission from experimentalists and by mining literature and other data sources; in some cases the data is verified using automated algorithms or manual curation. In addition to direct detection of physical protein interactions, indirect methods can be used to predict the functional association between proteins or to predict the location of the interaction interface itself. There is indeed a wide range of detail characterizing the interactions available from different databases. For example, Y2H data gives the identity of interacting proteins, electron microscopy provides relative positional information of interacting proteins, and crystallography provides full atomic detail of interaction surfaces. In addition, interacting proteins can be studied either as complete units or by domains used as the units of interaction. Consequently, in this review we group all databases into protein and domain-related databases.

In spite of the interaction data diversity, there exist considerable overlaps in the datasets contained in the databases, making it difficult to recommend a single resource for a particular type of information. In one effort to deal with this redundancy, the International Molecular Exchange Consortium (IMEx) has been formed in which databases agree to share their data in a consistent and timely fashion (Table 2). In addition, a standard data model has been proposed for the representation and exchange of protein interaction data [101]. A few example databases from Table 2 will now be highlighted to illustrate different types of interaction data available.

Protein Interaction Databases

Database of Interacting Proteins. The Database of Interacting Proteins (DIP) contains experimentally determined protein interactions and includes a core subset of interactions that have passed a quality assessment [102]. Interaction data are obtained from the literature; PDB; and high-throughput methods such as Y2H, DNA and protein microarrays; and TAP–MS analysis of protein complexes. Several methods are employed to assess the quality of interaction data and are offered as a service for query interactions. DIP has links to a couple of related databases

including LiveDIP, which records information about the state of a biological interaction, such as covalently modified, conformational, or cellular location states [103]. Another database related to DIP is Prolinks, which brings together four methods of linking proteins: phylogenetic profiles, Rosetta Stone, gene neighbors, and gene clusters[104]. The database includes a Proteome Navigator tool to browse the linkages and view accompanying data.

Biomolecular Interaction Network Database. The Biomolecular Interaction Network Database (BIND) includes high-throughput experimental datasets and protein complexes from PDB [105,106]. It contains a variety of curated experimental data. A generalized data specification handles not only various types of protein interaction data, but also protein–small molecule interactions and protein–nucleic acid interactions. An interaction viewer is provided to browse the interaction space. BIND also can distinguish different functional types of interactions.

Munich MPact/MIPS database. MPact is a resource to access MIPS, which contains a manually curated yeast protein interaction dataset [97] collected by curators from the literature. The resource also includes high-throughput results for yeast, but keeps this data separate. MIPS is often used as a standard of truth database for evaluating the quality of data and the accuracy of interaction prediction methods.

Domain Interaction Databases

PIBASE database. PIBASE is a database of domain interactions from the protein structure data [107]. It uses SCOP and CATH domain definitions to find putative domain interactions. Several methods are employed to remove redundancy in structural data; for example, structural comparisons of interfaces are made between domains within one structure. The database combines physicochemical properties of protein binding sites and has a link to MODBASE [108], containing models of three-dimensional structures that allow use of PIBASE for modeling of putative domain interfaces.

3did database. 3did allows one to explore the details of domain interactions from protein structure data (yeast interactions are also included) [109]. For each domain, an overview is given of all its interactions with other domains, showing different interaction types. In some cases, dot plots of structural comparisons between interaction interfaces show the variance of the interactions between pairs of domain families. Database entries are also supplied with the GO-based functional annotations. InterPreTS is a Web-based service associated with 3did that predicts domain interactions based on sequence homology of query proteins to a database of interacting domains (DBID) [21].

Conserved Binding Mode database. The Conserved Binding Mode (CBM) database is a collection of domain interactions from the structure data where domains are defined by the Conserved Domain Database [110]. Unlike other structure-based databases, domain interactions are grouped by geometry into conserved interaction modes for each pair of domain families across all PDB structures [26]. Structural superpositions are used to infer CBMs from different members of interacting domain families docking in the same way. Such domain interactions with recurring structural themes have greater significance to be biologically relevant, unlike spurious crystal packing interactions. CBMs can also assist in analyzing protein interaction network topology by emphasizing connections made in a biological context. Finally, the CBM database can be used to categorize the specific interaction surfaces that have evolved from conserved domains and thereby allows for the homology modeling of protein interaction interfaces. A similar approach for grouping interaction patterns for SCOP domains was recently undertaken with the SCOPPI database [111].

Domain Interaction Map database. Domain Interaction Map (DIMA) database is a domain interaction map derived from phylogenetic profiling Pfam domains [97]. Instead of looking at entire protein sequences, the algorithm compares the occurrences of domains across genomes and associates similar patterns of occurrences with functional associations. The method works well for domains with moderate information content that have distinct phylogenetic profiles.

In this paper we have reviewed a wide spectrum of experimental techniques for identifying and characterizing protein interactions; each technique can provide a piece in the puzzle of mechanisms of protein recognition [112]. Despite enormous efforts in this field, the overall picture is still incomplete, which is not surprising given the enormous complexity of a cell. Indeed, proteins can behave differently in different parts of the cell, and many proteins form transient complexes that are difficult to identify. Moreover, evolutionarily conserved proteins have much better coverage in experiments than the proteins restricted to a certain organism. The low coverage together with the small overlap between different experimental methods calls for the development of theoretical approaches for interaction data verification and prediction, the topic we address in our companion review [33]. ■

Acknowledgments

The authors thank Lewis Geer for helpful discussions and Robert Yates for graphic design of the figures. This work was supported by the Intramural Research Program of the National Library of Medicine at the National Institutes of Health of the US Department of Health and Human Services.

Author contributions. BAS and ARP analyzed the data and wrote the paper.

Funding. The authors received no specific funding for this article.

Competing interests. The authors have declared that no competing interests exist.

References

1. Giot L, Bader JS, Brouwer C, Chaudhuri A, Kuang B, et al. (2003) A protein interaction map of *Drosophila melanogaster*. Science 302: 1727–1736.
2. Li S, Armstrong CM, Bertin N, Ge H, Milstein S, et al. (2004) A map of the interactome network of the metazoan *C. elegans*. Science 303: 540–543.
3. Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, et al. (2006) Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. Nature 440: 637–643.
4. Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, et al. (2006) Proteome survey reveals modularity of the yeast cell machinery. Nature 440: 631–636.
5. Ofran Y, Rost B (2003) Analysing six types of protein–protein interfaces. J Mol Biol 325: 377–387.
6. Bogan AA, Thorn KS (1998) Anatomy of hot spots in protein interfaces. J Mol Biol 280: 1–9.
7. Chakrabarti P, Janin J (2002) Dissecting protein–protein recognition sites. Proteins 47: 334–343.
8. Lo Conte L, Chothia C, Janin J (1999) The atomic structure of protein–protein recognition sites. J Mol Biol 285: 2177–2198.
9. Jones S, Thornton JM (1997) Analysis of protein–protein interaction sites using surface patches. J Mol Biol 272: 121–132.
10. Guharoy M, Chakrabarti P (2005) Conservation and relative importance

of residues across protein–protein interfaces. Proc Natl Acad Sci U S A 102: 15447–15452.
11. DeLano WL (2002) Unraveling hot spots in binding interfaces: Progress and challenges. Curr Opin Struct Biol 12: 14–20.
12. Keskin O, Ma B, Rogale K, Gunasekaran K, Nussinov R (2005) Protein–protein interactions: Organization, cooperativity and mapping in a bottom-up Systems Biology approach. Phys Biol 2: S24–S35.
13. Res I, Lichtarge O (2005) Character and evolution of protein–protein interfaces. Phys Biol 2: S36–S43.
14. 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 U S A 100: 5772–5777.
15. Sheinerman FB, Honig B (2002) On the role of electrostatic interactions in the design of protein–protein interfaces. J Mol Biol 318: 161–177.
16. 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.
17. Fernandez A, Scheraga HA (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. Proc Natl Acad Sci U S A 100: 113–118.
18. Kundrotas PJ, Alexov E (2006) Electrostatic properties of protein–protein complexes. Biophys J 91: 1724–1736.
19. Valdar WS, Thornton JM (2001) Protein–protein interfaces: Analysis of amino acid conservation in homodimers. Proteins 42: 108–124.
20. Nooren IM, Thornton JM (2003) Structural characterisation and functional significance of transient protein–protein interactions. J Mol Biol 325: 991–1018.
21. Aloy P, Ceulemans H, Stark A, Russell RB (2003) The relationship between sequence and interaction divergence in proteins. J Mol Biol 332: 989–998.
22. Littler SJ, Hubbard SJ (2005) Conservation of orientation and sequence in protein domain–domain interactions. J Mol Biol 345: 1265–1279.
23. Teichmann SA (2002) The constraints protein–protein interactions place on sequence divergence. J Mol Biol 324: 399–407.
24. Panchenko AR, Wolf YI, Panchenko LA, Madej T (2005) Evolutionary plasticity of protein families: Coupling between sequence and structure variation. Proteins 61: 535–544.
25. Aloy P, Russell RB (2004) Ten thousand interactions for the molecular biologist. Nat Biotechnol 22: 1317–1321.
26. Shoemaker BA, Panchenko AR, Bryant SH (2006) Finding biologically relevant protein domain interactions: Conserved binding mode analysis. Protein Sci 15: 352–361.
27. Kim WK, Henschel A, Winter C, Schroeder M (2006) The many faces of protein–protein interactions: A compendium of interface geometry. PLoS Comput Biol 2: e124.
28. Keskin O, Tsai CJ, Wolfson H, Nussinov R (2004) A new, structurally nonredundant, diverse data set of protein–protein interfaces and its implications. Protein Sci 13: 1043–1055.
29. Grishin NV, Phillips MA (1994) The subunit interfaces of oligomeric enzymes are conserved to a similar extent to the overall protein sequences. Protein Sci 3: 2455–2458.
30. Caffrey DR, Somaroo S, Hughes JD, Mintseris J, Huang ES (2004) Are protein–protein interfaces more conserved in sequence than the rest of the protein surface? Protein Sci 13: 190–202.
31. Korkin D, Davis FP, Sali A (2005) Localization of protein-binding sites within families of proteins. Protein Sci 14: 2350–2360.
32. Panchenko AR, Kondrashov F, Bryant S (2004) Prediction of functional sites by analysis of sequence and structure conservation. Protein Sci 13: 884–892.
33. Shoemaker BA, Panchenko AR (2007) Deciphering protein–protein interactions. Part II. Computational methods to predict protein and domain interaction partners. PLoS Comp Biol 3: e43.
34. Fields S, Song O (1989) A novel genetic system to detect protein–protein interactions. Nature 340: 245–246.
35. Fashena SJ, Serebriiskii I, Golemis EA (2000) The continued evolution of two-hybrid screening approaches in yeast: How to outwit different preys with different baits. Gene 250: 1–14.
36. Causier B (2004) Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom Rev 23: 350–367.
37. Auerbach D, Thaminy S, Hottiger MO, Stagljar I (2002) The post-genomic era of interactive proteomics: Facts and perspectives. Proteomics 2: 611–623.
38. Van Criekinge W, Beyaert R (1999) Yeast two-hybrid: State of the art. Biol Proced Online 2: 1V38.
39. Toby GG, Golemis EA (2001) Using the yeast interaction trap and other two-hybrid–based approaches to study protein–protein interactions. Methods 24: 201–217.
40. Lee JW, Lee SK (2004) Mammalian two-hybrid assay for detecting protein–protein interactions in vivo. Methods Mol Biol 261: 327–336.
41. Walhout AJ, Vidal M (2001) High-throughput yeast two-hybrid assays for large-scale protein interaction mapping. Methods 24: 297–306.
42. Aronheim A, Zandi E, Hennemann H, Elledge SJ, Karin M (1997) Isolation of an AP-1 repressor by a novel method for detecting protein–protein interactions. Mol Cell Biol 17: 3094–3102.
43. Mohler WA, Blau HM (1996) Gene expression and cell fusion analyzed by lacZ complementation in mammalian cells. Proc Natl Acad Sci U S A 93: 12423–12427.
44. Bartel PL, Roecklein JA, SenGupta D, Fields S (1996) A protein linkage map of *Escherichia coli* bacteriophage T7. Nat Genet 12: 72–77.
45. Finley RL Jr, Brent R (1994) Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. Proc Natl Acad Sci U S A 91: 12980–12984.
46. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, et al. (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. Science 287: 116–122.
47. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, et al. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. Nature 403: 623–627.
48. Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, et al. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. Proc Natl Acad Sci U S A 98: 4569–4574.
49. Ghavidel A, Cagney G, Emili A (2005) A skeleton of the human protein interactome. Cell 122: 830–832.
50. Rual JF, Venkatesan K, Hao T, Hirozane-Kishikawa T, Dricot A, et al. (2005) Towards a proteome-scale map of the human protein–protein interaction network. Nature 437: 1173–1178.
51. Deeds EJ, Ashenberg O, Shakhnovich EI (2006) A simple physical model for scaling in protein–protein interaction networks. Proc Natl Acad Sci U S A 103: 311–316.
52. Di Tullio A, Reale S, De Angelis F (2005) Molecular recognition by mass spectrometry. J Mass Spectrom 40: 845–865.
53. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 198–207.
54. Whitehouse CM, Dreyer RN, Yamashita M, Fenn JB (1985) Electrospray interface for liquid chromatographs and mass spectrometers. Anal Chem 57: 675–679.
55. Pieles U, Zurcher W, Schar M, Moser HE (1993) Matrix-assisted laser desorption ionization time-of-flight mass spectrometry: A powerful tool for the mass and sequence analysis of natural and modified oligonucleotides. Nucleic Acids Res 21: 3191–3196.
56. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10,000 daltons. Anal Chem 60: 2299–2301.
57. Yates JR III, Eng JK, McCormack AL, Schieltz D (1995) Method to correlate tandem mass spectra of modified peptides to amino acid sequences in the protein database. Anal Chem 67: 1426–1436.
58. Taylor JA, Johnson RS (1997) Sequence database searches via de novo peptide sequencing by tandem mass spectrometry. Rapid Commun Mass Spectrom 11: 1067–1075.
59. Pevzner PA, Dancik V, Tang CL (2000) Mutation-tolerant protein identification by mass spectrometry. J Comput Biol 7: 777–787.
60. Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, et al. (2004) Open mass spectrometry search algorithm. J Proteome Res 3: 958–964.
61. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, et al. (1999) A generic protein purification method for protein complex characterization and proteome exploration. Nat Biotechnol 17: 1030–1032.
62. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, et al. (2001) The tandem affinity purification (TAP) method: A general procedure of protein complex purification. Methods 24: 218–229.
63. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 415: 141–147.
64. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. Nature 415: 180–183.
65. Jansen R, Greenbaum D, Gerstein M (2002) Relating whole-genome expression data with protein–protein interactions. Genome Res 12: 37–46.
66. Deane CM, Salwinski L, Xenarios I, Eisenberg D (2002) Protein interactions: Two methods for assessment of the reliability of high throughput observations. Mol Cell Proteomics 1: 349–356.
67. Kemmeren P, van Berkum NL, Vilo J, Bijma T, Donders R, et al. (2002) Protein interaction verification and functional annotation by integrated analysis of genome-scale data. Mol Cell 9: 1133–1143.
68. Kim SK, Lund J, Kiraly M, Duke K, Jiang M, et al. (2001) A gene expression map for *Caenorhabditis elegans*. Science 293: 2087–2092.
69. Troyanskaya OG, Garber ME, Brown PO, Botstein D, Altman RB (2002) Nonparametric methods for identifying differentially expressed genes in microarray data. Bioinformatics 18: 1454–1461.
70. Troyanskaya OG (2005) Putting microarrays in a context: Integrated analysis of diverse biological data. Brief Bioinform 6: 34–43.
71. Bhardwaj N, Lu H (2005) Correlation between gene expression profiles and protein–protein interactions within and across genomes. Bioinformatics 21: 2730–2738.
72. Tornow S, Mewes HW (2003) Functional modules by relating protein interaction networks and gene expression. Nucleic Acids Res 31: 6283–6289.
73. Teichmann SA, Babu MM (2002) Conservation of gene co-regulation in prokaryotes and eukaryotes. Trends Biotechnol 20: 407–410. Discussion on p. 410.
74. Ge H, Liu Z, Church GM, Vidal M (2001) Correlation between

transcriptome and interactome mapping data from *Saccharomyces cerevisiae*. Nat Genet 29: 482–486.
75. Grigoriev A (2001) A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res 29: 3513–3519.
76. Mrowka R, Patzak A, Herzel H (2001) Is there a bias in proteome research? Genome Res 11: 1971–1973.
77. Stuart JM, Segal E, Koller D, Kim SK (2003) A gene-coexpression network for global discovery of conserved genetic modules. Science 302: 249–255.
78. Fraser HB, Hirsh AE, Wall DP, Eisen MB (2004) Coevolution of gene expression among interacting proteins. Proc Natl Acad Sci U S A 101: 9033–9038.
79. Rutherford SL (2000) From genotype to phenotype: Buffering mechanisms and the storage of genetic information. Bioessays 22: 1095–1105.
80. Hartman J, Garvik B, Hartwell L (2001) Principles for the buffering of genetic variation. Science 291: 1001–1004.
81. Bender A, Pringle JR (1991) Use of a screen for synthetic lethal and multicopy suppressee mutants to identify two new genes involved in morphogenesis in *Saccharomyces cerevisiae*. Mol Cell Biol 11: 1295–1305.
82. Ooi SL, Pan X, Peyser BD, Ye P, Meluh PB, et al. (2006) Global synthetic-lethality analysis and yeast functional profiling. Trends Genet 22: 56–63.
83. Brown JA, Sherlock G, Myers CL, Burrows NM, Deng C, et al. (2006) Global analysis of gene function in yeast by quantitative phenotypic profiling. Mol Syst Biol 2: 0001.
84. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, et al. (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 294: 2364–2368.
85. Ye P, Peyser BD, Pan X, Boeke JD, Spencer FA, et al. (2005) Gene function prediction from congruent synthetic lethal interactions in yeast. Mol Syst Biol 1: 0026.
86. Berman HM, Bhat TN, Bourne PE, Feng Z, Gilliland G, et al. (2000) The Protein Data Bank and the challenge of structural genomics. Nat Struct Biol 7: 957–959.
87. Lippincott-Schwartz J, Patterson GH (2003) Development and use of fluorescent protein markers in living cells. Science 300: 87–91.
88. Piehler J (2005) New methodologies for measuring protein interactions in vivo and in vitro. Curr Opin Struct Biol 15: 4–14.
89. Yan Y, Marriott G (2003) Analysis of protein interactions using fluorescence technologies. Curr Opin Chem Biol 7: 635–640.
90. Karlsson R (2004) SPR for molecular interaction analysis: A review of emerging application areas. J Mol Recognit 17: 151–161.
91. Cooper MA (2003) Label-free screening of bio-molecular interactions. Anal Bioanal Chem 377: 834–842.
92. Velazquez Campoy A, Freire E (2005) ITC in the post-genomic era...? Priceless. Biophys Chem 115: 115–124.
93. Yang Y, Wang H, Erie DA (2003) Quantitative characterization of biomolecular assemblies and interactions using atomic force microscopy. Methods 29: 175–187.
94. Margittai M, Widengren J, Schweinberger E, Schroder GF, Felekyan S, et al. (2003) Single-molecule fluorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. Proc Natl Acad Sci U S A 100: 15516–15521.
95. Butland G, Peregrin-Alvarez JM, Li J, Yang W, Yang X, et al. (2005) Interaction network containing conserved and essential protein complexes in *Escherichia coli*. Nature 433: 531–537.
96. Matthews LR, Vaglio P, Reboul J, Ge H, Davis BP, et al. (2001) Identification of potential interaction networks using sequence-based searches for conserved protein–protein interactions or "interologs." Genome Res 11: 2120–2126.
97. Guldener U, Munsterkotter M, Oesterheld M, Pagel P, Ruepp A, et al. (2006) MPact: The MIPS protein interaction resource on yeast. Nucleic Acids Res 34: D436–D441.
98. Sprinzak E, Sattath S, Margalit H (2003) How reliable are experimental protein–protein interaction data? J Mol Biol 327: 919–923.
99. Bader JS, Chaudhuri A, Rothberg JM, Chant J (2004) Gaining confidence in high-throughput protein interaction networks. Nat Biotechnol 22: 78–85.
100. Masters SC (2004) Co-immunoprecipitation from transfected cells. Methods Mol Biol 261: 337–350.
101. Hermjakob H, Montecchi-Palazzi L, Bader G, Wojcik J, Salwinski L, et al. (2004) The HUPO PSI's molecular interaction format—A community standard for the representation of protein interaction data. Nat Biotechnol 22: 177–183.
102. Salwinski L, Miller CS, Smith AJ, Pettit FK, Bowie JU, et al. (2004) The Database of Interacting Proteins: 2004 update. Nucleic Acids Res 32: D449–D451.
103. Duan XJ, Xenarios I, Eisenberg D (2002) Describing biological protein interactions in terms of protein states and state transitions: The LiveDIP database. Mol Cell Proteomics 1: 104–116.
104. Bowers PM, Pellegrini M, Thompson MJ, Fierro J, Yeates TO, et al. (2004) Prolinks: A database of protein functional linkages derived from coevolution. Genome Biol 5: R35.
105. Alfarano C, Andrade CE, Anthony K, Bahroos N, Bajec M, et al. (2005) The Biomolecular Interaction Network Database and related tools 2005 update. Nucleic Acids Res 33: D418–D424.
106. Bader GD, Hogue CW (2000) BIND—A data specification for storing and describing biomolecular interactions, molecular complexes and pathways. Bioinformatics 16: 465–477.
107. Davis FP, Sali A (2005) PIBASE: A comprehensive database of structurally defined protein interfaces. Bioinformatics 21: 1901–1907.
108. Pieper U, Eswar N, Davis FP, Braberg H, Madhusudhan MS, et al. (2006) MODBASE: A database of annotated comparative protein structure models and associated resources. Nucleic Acids Res 34: D291–D295.
109. Stein A, Russell RB, Aloy P (2005) 3did: Interacting protein domains of known three-dimensional structure. Nucleic Acids Res 33: D413–D417.
110. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, et al. (2002) CDD: A database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 30: 281–283.
111. Winter C, Henschel A, Kim WK, Schroeder M (2006) SCOPPI: A structural classification of protein–protein interfaces. Nucleic Acids Res 34: D310–D314.
112. Panchenko AR, Shoemaker BA (2006) ISMB tutorial 2006: Protein–protein interactions: Structure and systems approaches to analyze diverse genomic data. Available: http://www.ncbi.nlm.nih.gov/CBBresearch/Panchenko/ismb__tutorial2006.ppt. Accessed 16 February 2007.
113. Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 95: 14863–14868.
114. MacBeath G, Schreiber SL (2000) Printing proteins as microarrays for high-throughput function determination. Science 289: 1760–1763.
115. Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, et al. (2001) Global analysis of protein activities using proteome chips. Science 293: 2101–2105.
116. Jones RB, Gordus A, Krall JA, MacBeath G (2006) A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature 439: 168–174.
117. Smith GP (1985) Filamentous fusion phage: Novel expression vectors that display cloned antigens on the virion surface. Science 228: 1315–1317.
118. Baumeister W, Grimm R, Walz J (1999) Electron tomography of molecules and cells. Trends Cell Biol 9: 81–85.
119. von Mering C, Jensen LJ, Snel B, Hooper SD, Krupp M, et al. (2005) STRING: Known and predicted protein–protein associations, integrated and transferred across organisms. Nucleic Acids Res 33: D433–D437.
120. Zanzoni A, Montecchi-Palazzi L, Quondam M, Ausiello G, Helmer-Citterich M, et al. (2002) MINT: A molecular INTeraction database. FEBS Lett 513: 135–140.
121. Hermjakob H, Montecchi-Palazzi L, Lewington C, Mudali S, Kerrien S, et al. (2004) IntAct: An open source molecular interaction database. Nucleic Acids Res 32: D452–D455.
122. Stark C, Breitkreutz BJ, Reguly T, Boucher L, Breitkreutz A, et al. (2006) BioGRID: A general repository for interaction datasets. Nucleic Acids Res 34: D535–D539.
123. Peri S, Navarro JD, Amanchy R, Kristiansen TZ, Jonnalagadda CK, et al. (2003) Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 13: 2363–2371.
124. Kundrotas PJ, Alexov E (2006) PROTCOM: Searchable database of protein complexes enhanced with domain–domain structures. Nucleic Acids Res 34: D247–D251.
125. Aloy P, Russell RB (2003) InterPreTS: Protein interaction prediction through tertiary structure. Bioinformatics 19: 161–162.
126. Pieper U, Eswar N, Braberg H, Madhusudhan MS, Davis FP, et al. (2004) MODBASE, a database of annotated comparative protein structure models, and associated resources. Nucleic Acids Res 32: D217–D222.
127. Finn RD, Marshall M, Bateman A (2005) iPfam: Visualization of protein–protein interactions in PDB at domain and amino acid resolutions. Bioinformatics 21: 410–412.
128. Ng SK, Zhang Z, Tan SH, Lin K (2003) InterDom: A database of putative interacting protein domains for validating predicted protein interactions and complexes. Nucleic Acids Res 31: 251–254.
129. Pagel P, Oesterheld M, Stumpflen V, Frishman D (2006) The DIMA web resource—Exploring the protein domain network. Bioinformatics 22: 997–998.