Introduction

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.

A Tutorial in PLoS Computational Biology

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

Benjamin A. Shoemaker, Anna R. Panchenko

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.

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.

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.

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.

A Tutorial in PLoS Computational Biology

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

Benjamin A. Shoemaker, Anna R. Panchenko

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.
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 necessarily covalently associated with an activating domain [34]. According to the Y2H method, a protein of interest is necessarily covalently associated with an activating domain [34]. According to the Y2H method, a protein of interest is necessarily covalently associated with an activating domain [34]. According to the Y2H method, a protein of interest is necessarily covalently associated with an activating domain [34]. According to the Y2H method, a protein of interest is necessarily covalently associated with an activating domain [34].

**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

---

**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
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 7,125 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 YL049W gene as belonging to the dynemin–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
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

---

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

| Database | Proteins/Domains | Type | Number of Interactions |
|----------|------------------|------|------------------------|
| DIP, LiveDIP | P | ES | 55,733 |
| BIND | P | E.C.S | 83,517 |
| MPAct/MIPS* | P | E.C.F | 15,488 (4,300)* |
| STRING | P | E.P.F | 730,000 (proteins) |
| MINT* | P | E.C | 71,854 |
| IntAct* | P | E.C | 68,165 |
| BioGRID* | P | E.C | 116,000 (30,000)* |
| HPRD | P | E.C | 33,710 |
| ProtCom | P,D | S.H | 1,770 |
| 3did, Interacts | D | S.H | 3,304 |
| Pibase, ModBase | D | S.H | 2,387 |
| CBM | D | S | 2,784 |
| SCOPOPI | 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.

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

*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/ppl |
| 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/comprob/ProteCom |
| 3did [108], Interacts [125] | http://gatealoy.pcb.ub.es/3did/ |
| Pibase [107], ModBase [126] | http://altom.compbio.ucsf.edu/pibase |
| CBM [26] | ftp://ftp.ncbi.nlm.nih.gov/pub/cbm |
| SCOPOPI [111] | http://www.scoppi.org/ |
| iPfam [127] | http://www.sanger.ac.uk/Software/Pfam/iPfam |
| InterDom [128] | http://interdom.it.org.sg |
| DIMA [129] | http://mips.gsf.de/genec/proj/dima/index.html |
| Prolinks [104] | http://prolinks.doe-mbi.ucla.edu/cgi-bin/ functionator/pronav/|

doi:10.1371/journal.ppat.0030042.t003

*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
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 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, Boche S, et al. (2006) Proteome survey reveals modular architecture 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: 354–353.
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: 15417–15422.
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, Rogake 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 interactions. Phys Biol 2: S36–S43.
14. Ma B, Elakam T, Woloson 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: 14096–14101.
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 DM, 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 interface 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 VI, Panchenko LA, Madjar 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. Kortemme T, Henschel A, Winter C, Schroeder M (2006) The many faces of protein–protein interaction: A compendium of interface geometry. PLoS Comput Biol 2: e124.
28. Keskin O, Tsai CJ, Woloson H, Nussinov R (2004) A new, structurally nonredundant, diverse data set of protein–protein interfaces and its implications. Protein Sci. 13: 1045–1055.
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, Sonnhammer ES, Hughes JD, Mintsiris 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) Conservation of orientation and sequence in protein–domain interactions. J Mol Biol 345: 1265–1279.
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. Fodor SP, Szeverinski 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. Causer B (2004) Studying the interactome with the yeast two-hybrid system and mass spectrometry. Mass Spectrom Rev 23: 350–367.
37. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. Nature 422: 180–183.
38. Puig O, Caspary F, Rigaut G, Rutz B, Bouveret E, et al. (2001) The tandem mass spectrometry search algorithm. J Proteome Res 2: 1V38.
transcriptome and interactome mapping data from Saccharomyces cerevisiae. Nat Genet 29: 482–486.
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 functional linkages derived from coevolution. Genome Biol 5: R53.
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 (2003) BIND—A data specification for storing and describing biomolecular interactions, molecular complexes and pathways. Bioinformatics 16: 465–477.
107. Davis FP, Salit A (2005) PIRBASE: 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. Steiman A, Russell RB, Aloy P (2005) 3dIF: Interacting protein domains of known three-dimensional structure. Nucleic Acids Res 35: D413–D417.
110. Marchler-Bauer A, Panchenko AR, Shoemaker BA, Thiessen PA, Geer LY, et al. (2005) CDD: A database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res 33: 281–283.
111. Winter C, Henschel A, Kim WK, Schroeder M (2006) SCOOPi: A structural classification of protein–protein interfaces. Nucleic Acids Res 34: D510–D514.
112. Panchenko AR, Shoemaker BA (2006) DMB tutorial 2006: Protein–protein interactions: Structure and systems approaches to analyze diverse genomic data. Available: http://www.ncbi.nlm.nih.gov/CBBResearch/Panchenko/DMB_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 290: 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 SR, Gorard 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. (2005) Development of human protein reference database as an initial platform for approaching systems biology in humans. Genome Res 15: 2363–2371.
124. Kundrotas PJ, Alexev F (2006) PROTCOM: Searchable database of protein complexes enhanced with domain–domain structures. Nucleic Acids Res 34: D247–D251.
125. Aloy P, Russell RB (2003) InterPreT: 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) iPam: Visualization of protein–protein interactions in PDB at domain and amino acid resolution. Bioinformatics 21: 410–412.
128. Ng SK, Zhang Z, Tan SH, Lin K (2005) InterDom: A database of putative interacting protein domains for validating predicted protein interactions and complexes. Nucleic Acids Res 33: 251–254.
129. Pagel P, Osterheld M, Stumpflen V, Frishman D (2006) The DIMA web resource—Exploring the protein domain network. Bioinformatics 22: 997–998.