

Proteomic approaches for generating comprehensive protein interaction maps

Daniel Auerbach, Michael Fetchko and Igor Stagljar

The availability of complete genome sequences of numerous model organisms has initiated the development of new approaches in biological research to complement conventional biochemistry and genetics. In this context, high-throughput methods for detecting protein interactions, such as mass spectrometry and yeast two-hybrid assays, have produced vast amounts of data that can be exploited to infer protein function and regulation. In this review, we explore different genome-wide protein interaction studies and comment on their extrapolation towards understanding protein functions. It is likely that improvements of these approaches, together with more sophisticated databases and the invention of novel technologies, will help to decipher the complex interactions among proteins and to integrate interacting proteins into existing and novel cellular pathways.

Daniel Auerbach

Dualsystems Biotech

Winterthurerstrasse 190

CH-8057 Zurich, Switzerland

Michael Fetchko

Igor Stagljar*

Institute of Veterinary Biochemistry
and Molecular Biology

University of Zurich-Irchel

Winterthurerstrasse 190

CH-8057 Zurich, Switzerland

*e-mail: stagljar@vetbio.unizh.ch

▼ The explosion in genome sequencing over the past few years has brought many exciting opportunities to biological researchers. In the past decade, genome sequencing has advanced from the first complete genome sequence of a free-living organism, *Haemophilus influenzae*, in 1995 [1] to a working draft for the human genome in 2001 [2,3]. The complete genome sequences are now available for ~60 prokaryotic organisms (<http://www.tigr.org/tdb/mdb/mdbcomplete.html>) and several eukaryotic model organisms, including the baker's yeast *Saccharomyces cerevisiae* [4], the worm *Caenorhabditis elegans* [5] the fly *Drosophila melanogaster* [6], the plant *Arabidopsis thaliana* [7], the parasite *Plasmodium falciparum* [8–10] and the mouse *Mus musculus* [11] (Table 1).

Knowledge about the genome sequence of an organism, however, does not yet provide an insight into its intricate functions because the gene products, the proteins, can assume a multitude of functions. Proteomics, an emerging biomolecular discipline, comprises techniques

used to investigate the entire complement of protein varieties or the 'proteome' of an organism. The goal of proteomics is to determine when and where the proteome of an organism is expressed, and to identify post-translational modifications that regulate protein functions. Moreover, an important task of proteomics is to elucidate the interactions between individual members of the proteome in the normal and diseased state. Because disturbances of protein–protein interactions within a cell can lead to many diseases, it is of great biomedical interest to identify proteins that bind to a certain target protein and consequently help to modulate its function.

Recently, several large-scale studies have been completed that systematically identify large sets of protein–protein interactions with the aim of generating comprehensive protein interaction maps [12–20]. In this article, existing technologies will be highlighted along with the results that are derived from different genome-wide protein interaction studies. A list of emerging technologies that can potentially be used to improve these studies will also be provided. Finally, the use of proteomics data in understanding global protein networks will be discussed.

Approaches for generating protein interaction maps

Methods that are suitable for the detection of protein interactions on a large scale have to meet several criteria, the most important being automatization and acceptably low numbers of false-positive and false-negative interactions (Box 1). The ~6000 proteins expressed in the unicellular organism *S. cerevisiae* are estimated to generate ~30,000 interactions [21], and so the number of interactions of the

Table 1. Genome sequences of some organisms

| Model organism | Year sequence published | Predicted ORFs |
|--|-------------------------|----------------|
| Bacterium <i>Haemophilus influenzae</i> | 1995 | 1720 |
| Yeast <i>Saccharomyces cerevisiae</i> | 1996 | 6234 |
| Worm <i>Caenorhabditis elegans</i> | 1998 | 19,099 |
| Bacterium <i>Pseudomonas aeruginosa</i> | 2000 | 5570 |
| Fruit fly <i>Drosophila melanogaster</i> | 2000 | 13061 |
| Plant <i>Arabidopsis thaliana</i> | 2000 | ~25,000 |
| Human <i>Homo sapiens</i> | 2001 | ~35,000 |
| Parasite <i>Plasmodium falciparum</i> | 2002 | 5279 |
| Mouse <i>Mus musculus</i> | 2002 | ~30,000 |

Abbreviation: ORFs, open reading frames.

estimated 35,000 proteins derived from the human genome would be predicted to be much higher. Clearly, only a limited number of methods are able to cope with such large numbers of assays (Table 2). To date, proteome-wide interaction maps have been generated using three experimental methods: the yeast two-hybrid system [14,16,17]; affinity purification coupled to mass spectrometry [18,19]; and, on a smaller scale, proteome chips [20,22]. As discussed below, each of these methods has advantages and disadvantages and, ultimately, only a combination of several approaches, together with the use of advanced bioinformatics methods, will create reliable and informative interaction networks.

The yeast two-hybrid system

The yeast two-hybrid system originally created by Fields and Song [23] is a genetic system wherein an interaction

Box 1. Interaction terminology

The two most common problems associated with large-scale interaction networks are the detection of unspecific interactions and the failure to detect known interactions, referred to as false-positives and false-negatives, respectively.

False-positives

Every method discussed in the main text generates false-positive interactions. In the yeast two-hybrid system these can arise from the artificial localization of the interacting proteins in the nucleus or from independent activation of the transcriptional readout, whereas the affinity purification methods generate unspecific interactions through overexpression of bait proteins (i.e. proteins under investigation) and artificial lysis and purification conditions. False-positive interactions can be reduced by carrying out interaction experiments several times and scoring only reproducible interactions or by comparing datasets derived from several methods.

False-negatives

Owing to the technical limitations of each method, many interactions are not detected in large-scale interaction studies. For example, interactions between integral membrane proteins are under-represented in all studies to date and comparison of the different large-scale interaction networks suggests that their coverage is far from complete. By combining several datasets to increase their accuracy, coverage is automatically reduced because many biologically significant interactions that are detected by one method, but are absent in the dataset created by another method, are ignored.

between a known bait protein and its partner expressed from a library is detected via the reconstitution of a transcription factor and concurrent activation of a set of reporter

Table 2. High-throughput methods used to detect protein interactions

| Method | Advantages | Drawbacks |
|--|---|---|
| Yeast two-hybrid | Interactions are detected <i>in vivo</i> No artificial lysis or washing steps Has the potential to detect weak and transient interactions | Mostly detects binary interactions Forced localization of interacting proteins to the nucleus can create both false-positives and false-negatives Artificial fusion might inhibit certain interactions Bias towards certain classes of proteins with regard to cellular compartments |
| Affinity purification coupled to mass spectrometry | Detects interactions <i>in vivo</i> and in their normal cellular setting Capable of detecting interactions that depend on higher order complexes Less chance of inhibition of protein-protein interactions by artificial fusions because only one complex partner is tagged | Low-abundance proteins might be missed Some interactions might be lost during purification steps Overexpression of bait proteins might lead to the detection of false-positive interactions (HMS-PCI) Artificial fusion might inhibit certain interactions Bias towards certain classes of proteins |
| Proteome chips | Chips can be screened for a wide variety of interactions, such as protein-protein, lipid-protein or drug-protein interactions | Not all protein classes can be purified efficiently Interactions are detected <i>in vitro</i> Artificial fusion might inhibit certain interactions |

Abbreviation: HMS-PCI, high-throughput mass spectrometric protein complex identification.

genes (Fig. 1a). As a genetic system, the yeast two-hybrid system is well suited to high-throughput applications that require automatization. Large-scale two-hybrid approaches have used two complementary approaches, the so-called 'matrix approach' and the 'library screening approach' for screening large sets of proteins. Comprehensive yeast two-hybrid screens using both methods have been carried out for organisms as diverse as viruses [12,13] and higher eukaryotes [15,24] (the reader is referred to a recent review for an in-depth discussion on available interaction maps [25]). Here, we focus on the two large-scale screens that have been accomplished using *S. cerevisiae* [16,17].

In the matrix approach (Fig. 1b) a yeast strain expressing the bait protein under investigation is mated with an array of yeast strains that express many different prey proteins, and interactions are detected by growth on selective medium. The position of growing colonies in the array identifies the prey that interacts with the bait of interest. Uetz *et al.* [16] performed a matrix screen by mating 192 baits with a set of 5345 yeast open reading frames (ORFs). In this way, they identified 281 interactions.

The library screening approach (Fig. 1c) does not separate the different prey strains on an array but instead screens a set of baits against a library of either full-length ORFs or ORF fragments. Colonies expressing interacting proteins are selected and the library plasmids are isolated and sequenced to determine the identity of the interacting prey protein. This has been demonstrated in the prokaryote *Helicobacter pylori* where a large-scale library screen using 285 ORFs yielded a network encompassing 1280 protein interactions [14]. By applying the library screening approach to yeast, Ito *et al.* [17] reported a core dataset of 841 interactions involving 797 proteins. Uetz *et al.* also performed a library screen using their set of baits and preys and reported 692 interactions [16].

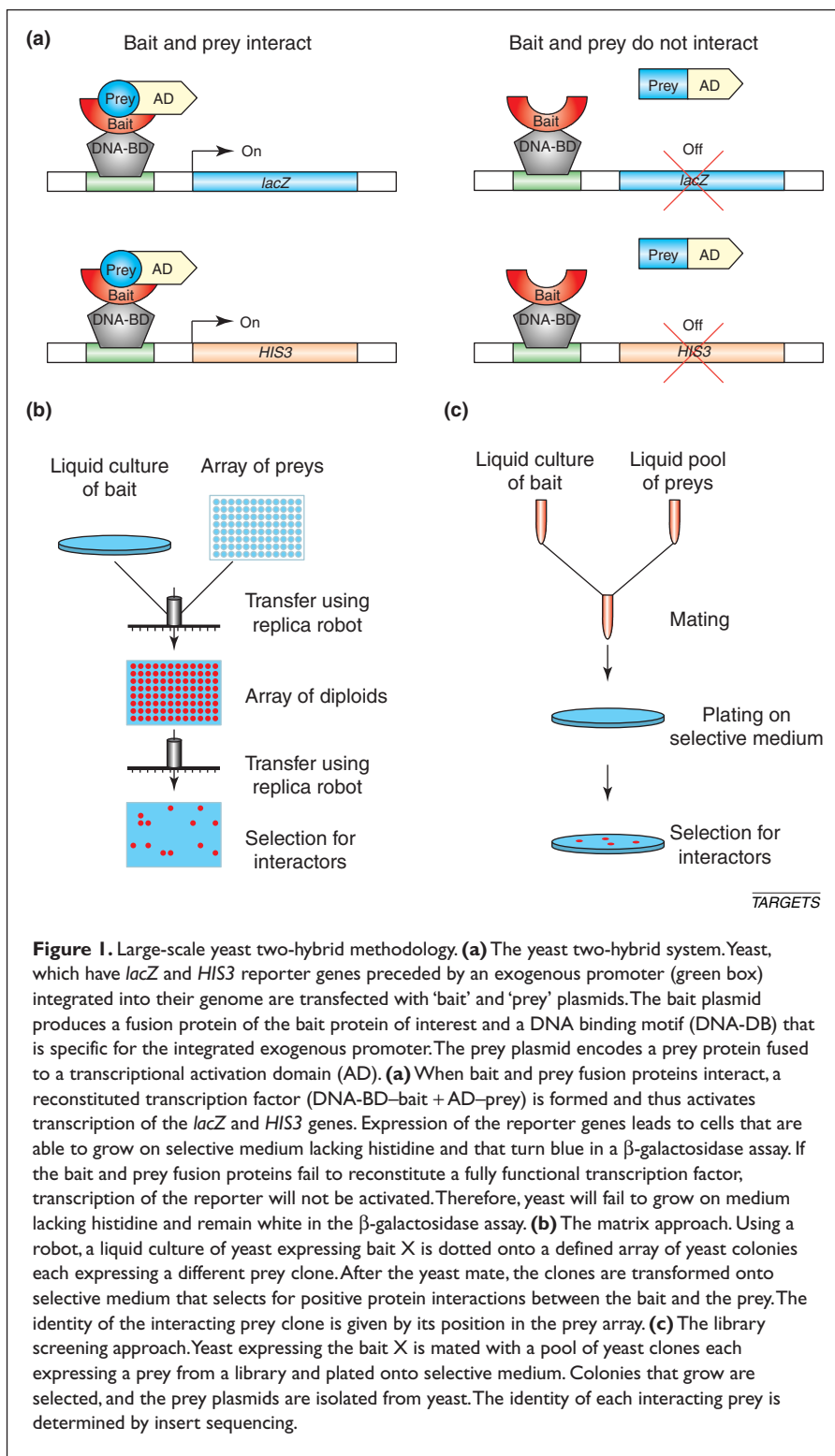
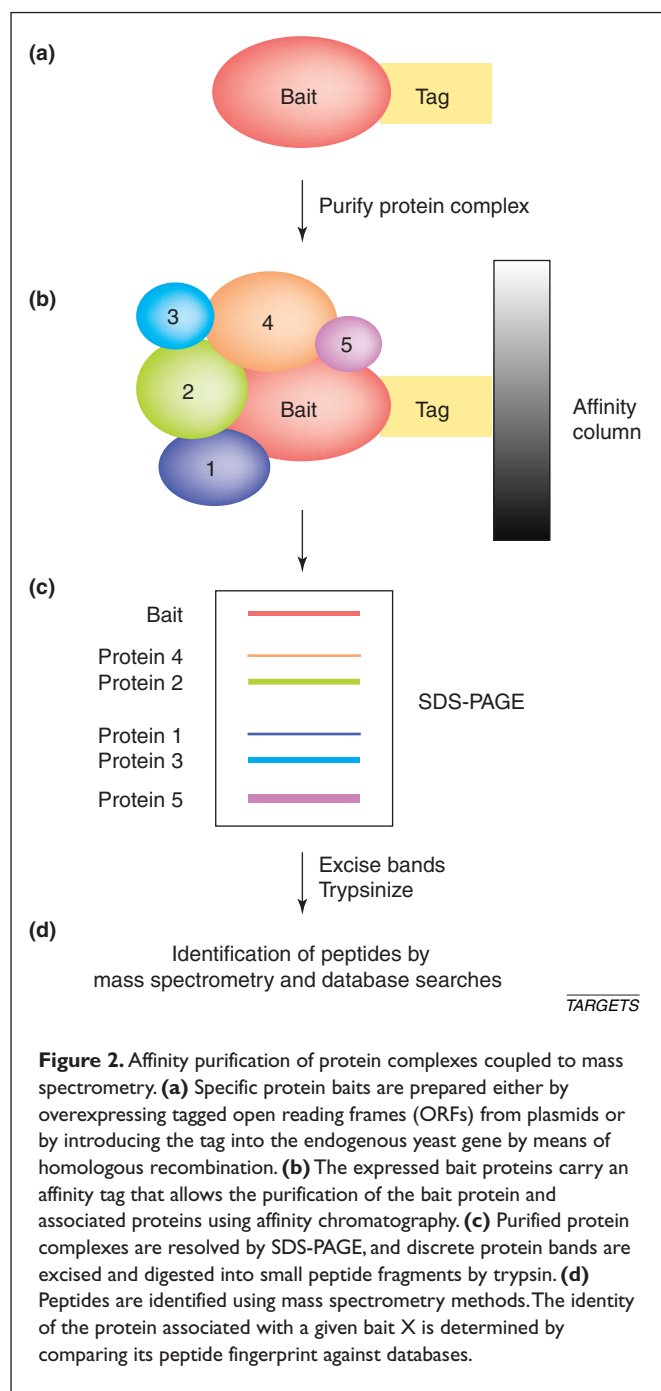


Figure 1. Large-scale yeast two-hybrid methodology. **(a)** The yeast two-hybrid system. Yeast, which have *lacZ* and *HIS3* reporter genes preceded by an exogenous promoter (green box) integrated into their genome are transfected with 'bait' and 'prey' plasmids. The bait plasmid produces a fusion protein of the bait protein of interest and a DNA binding motif (DNA-BD) that is specific for the integrated exogenous promoter. The prey plasmid encodes a prey protein fused to a transcriptional activation domain (AD). **(a)** When bait and prey fusion proteins interact, a reconstituted transcription factor (DNA-BD–bait + AD–prey) is formed and thus activates transcription of the *lacZ* and *HIS3* genes. Expression of the reporter genes leads to cells that are able to grow on selective medium lacking histidine and that turn blue in a β -galactosidase assay. If the bait and prey fusion proteins fail to reconstitute a fully functional transcription factor, transcription of the reporter will not be activated. Therefore, yeast will fail to grow on medium lacking histidine and remain white in the β -galactosidase assay. **(b)** The matrix approach. Using a robot, a liquid culture of yeast expressing bait X is dotted onto a defined array of yeast colonies each expressing a different prey clone. After the yeast mate, the clones are transformed onto selective medium that selects for positive protein interactions between the bait and the prey. The identity of the interacting prey clone is given by its position in the prey array. **(c)** The library screening approach. Yeast expressing the bait X is mated with a pool of yeast clones each expressing a prey from a library and plated onto selective medium. Colonies that grow are selected, and the prey plasmids are isolated from yeast. The identity of each interacting prey is determined by insert sequencing.

TARGETS

Compared with the matrix approach, the library approach has the advantage of using both full-length ORFs and ORF fragments. Thus, a library screen can uncover interactions that only take place between single domains (e.g. when an interacting protein contains both a domain that interacts



with the bait and a domain that inhibits this interaction). By contrast, matrix approaches have no bias towards abundant proteins, unlike cDNA libraries, where abundant interactors can be isolated many times, potentially obscuring rare interactors.

Detection of complexes by affinity purification coupled to mass spectrometry

In contrast to the yeast two-hybrid system, which largely detects binary interactions, affinity purification coupled to

mass spectrometry relies on the selective purification of entire protein complexes from the cell, followed by separation of its subunits and their identification by mass spectrometry (Fig. 2). Several different approaches for expression, purification and detection have been described: (1) overexpression from high-copy vectors [19] versus the use of endogenous expression levels [18,26]; (2) direct purification of endogenous proteins by multidimensional liquid chromatography [26] versus affinity purification using one [19] or two [18] affinity tags; and (3) different mass spectrometry strategies for detection and identification of the complex partners [18,19].

In an approach termed 'TAP' (tandem affinity purification) a sequence encoding tandem calmodulin binding protein and protein A affinity tags was directly fused to yeast ORFs by homologous recombination [18]. A total of 589 tagged baits were used to isolate associated complexes by two rounds of affinity purification. Purified complexes were resolved on one-dimensional gels, bands corresponding to protein components were isolated and the proteins were identified by MALDI-TOF (matrix-assisted laser desorption/ionisation-time of flight) mass spectrometry. The 589 baits used in this study resulted in the identification of 98 known and 134 new complexes. Analysis of the interactions led the authors to propose new cellular roles for 344 proteins, including 231 proteins of unknown function. In a different approach, termed HMS-PCI (high-throughput mass spectrometric protein complex identification), Ho *et al.* [19] overexpressed baits tagged with FLAG (a commercially available epitope tag) and purified associated complexes using biochemical methods followed by tandem mass spectrometry. After removal of nonspecifically interacting proteins (false-positives), 493 baits yielded 3617 interactions. However, it should be noted that because only a single purification step was used the data can be assumed to represent associations between proteins in a co-purified complex rather than protein-protein interactions.

Proteome chips

In contrast to yeast two-hybrid or affinity purification methods that detect protein interactions occurring *in vivo*, proteome chips detect interactions *in vitro* (Fig. 3). In this approach, proteins are overexpressed as fusions to tandem glutathione-S-transferase (GST) and polyhistidine affinity tags and purified in a high-throughput fashion [27]. The purified proteins are attached to the surface of glass slides by means of the polyhistidine affinity tag, resulting in a uniform orientation away from the glass surface and preserving their folded conformation. Zhu *et al.* [20] applied proteome chips to the study of protein-protein interactions by expressing 5800 yeast ORFs and printing the purified

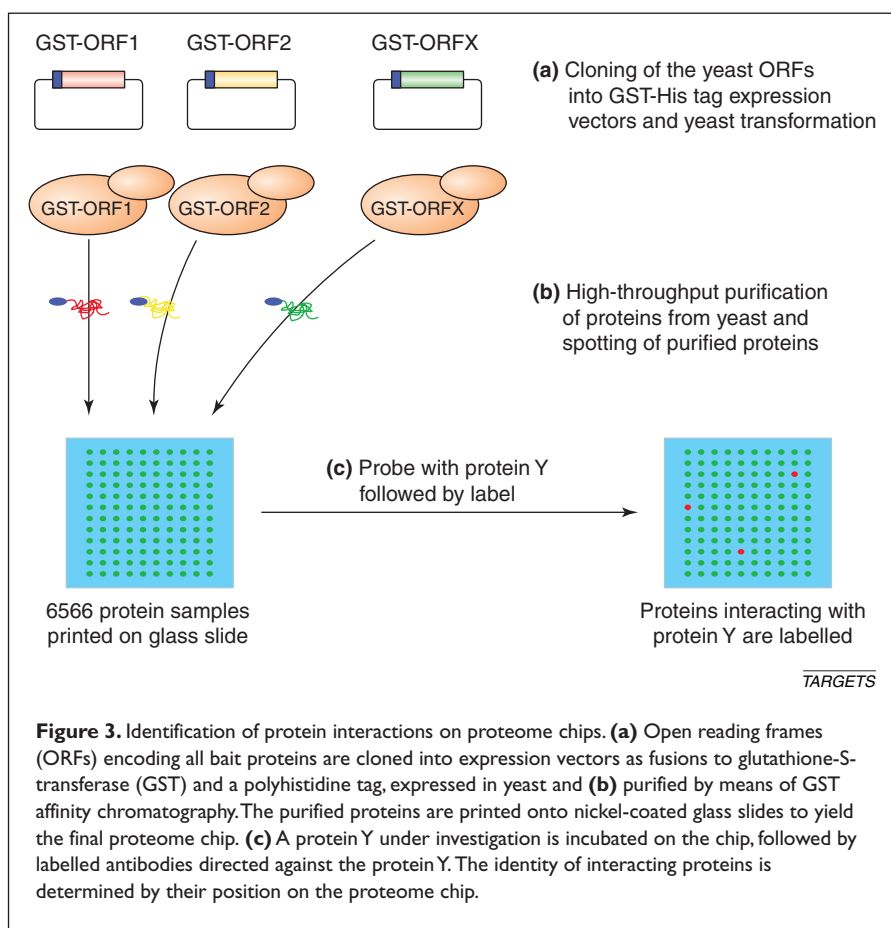
fusion proteins onto glass slides. The array was screened using calmodulin as a bait, and six known and 33 potentially new interaction partners were identified. To demonstrate the wide applicability of proteome chip screening, the authors also used liposomes to screen for proteins that interact with the lipid phosphoinositide.

Although no global interaction map has been created using proteome chips, the technology holds much promise. The ability to screen not only for protein–protein interactions but also for protein–RNA, protein–DNA or protein–lipid interactions clearly complements and extends the yeast two-hybrid and affinity purification methods.

What do we learn from protein–protein interaction maps?

Generally, there are two ways to assess the relative value of large-scale interaction maps. Maps that have been created using different methods can be compared and interactions that are found in both datasets are scored as reliable interactions. For example, comparison of the core datasets of two independent large-scale two-hybrid projects revealed an unexpectedly small overlap: only 20% of all interactions were found in both datasets [17].

Alternatively, the data from large-scale studies can be compared to the bulk of interactions that have been extracted from the literature (Box 2). This is based on the assumption that interactions reported in the literature have been validated carefully. A recent study used these comparisons as benchmarks to compare the available data from large-scale screens [21]. Their findings highlight several problems associated with current approaches: despite the fact that high-throughput methods in yeast have generated 80,000 interactions to date, only a small fraction (2400 interactions) are supported by more than one method. The authors ascribe this to the fact that the methods differ substantially in the type of interactions that are detectable. For example, methods based on the purification of complexes often miss signaling pathways and transport pathways, whereas yeast two-hybrid screens under-represent interactions between proteins involved in translation. Moreover, yeast two-hybrid screens using the library approach and complex purification methods are biased



towards highly abundant proteins, although yeast two-hybrid approaches are less biased than affinity purification methods. By contrast, matrix-based yeast two-hybrid screens do not suffer from a bias towards abundant proteins because defined interactions partners from a pool of known proteins are tested against each other.

Von Mering *et al.* also compared the coverage of different methods and their relative accuracy [21]. Compared with yeast two-hybrid based approaches, complex detection achieves a much higher coverage with fewer baits: in the HMS-PCI approach [19] only 493 yeast baits were needed to link 25% of the yeast proteome. By contrast, spurious, non-reproducible interactions that might arise during the purification steps pose a problem. For a given bait, complex purifications have to be carried out at least twice to separate reproducible interactions from non-specific background. When using TAP methodology, 70% of all interactions fall into the ‘reproducible’ category [18], whereas for the dataset created by HMS-PCI reproducible interactions only make up 20% of all interactions identified [19]. Thus, the high coverage achieved in the affinity purification approach by Ho *et al.* is offset by lower accuracy.

Box 2. Interaction databases

BIND and PreBIND (<http://www.bind.ca/>)

The Biomolecular Interaction Network Database (BIND) stores descriptions of interactions, molecular complexes and pathways. PreBIND can be used to selectively search the literature for protein interactions involving the protein of interest.

GRID (<http://biodata.mshri.on.ca/grid/servlet/Index>)

The General Repository of Interaction Datasets (GRID) is a database of genetic and physical interactions and contains interaction data from many sources, including several genome- and proteome-wide studies, the Munich Information Center for Protein Sequences (MIPS) database, and BIND.

YPD (<http://www.incyte.com/sequence/proteome/databases/YPD.shtml>)

The Yeast Proteome Database (YPD) contains information on all the characterized proteins of *Saccharomyces cerevisiae*.

OSPREY (<http://biodata.mshri.on.ca/osprey/servlet/Index>)

The Osprey Network Visualization System is used for graphically representing physical and genetic biological interactions. Osprey is coupled with the GRID.

PRONET (<http://www.myriad-pronet.com>)

Myriad's ProNet program provides researchers with the ability to rapidly identify large numbers of novel protein–protein interactions. ProNet has the capacity to discover protein–protein interactions regardless of the source of the proteins.

Interactions that are detected using proteome chips are more difficult to compare with either two-hybrid or affinity purification methods because they detect protein interactions that have formed *in vitro*. This might be advantageous because purified proteins can be used to directly assess binary interactions. Zhu *et al.* note that they were able to detect 50% of all calmodulin interactors reported in the literature, which might suggest a higher coverage of interactions than using either two-hybrid or affinity purification studies [20]. However, a reliable comparison of proteome chips with two-hybrid or affinity purification studies will ultimately require the generation of a yeast proteome-wide interaction dataset using proteome chip technology.

Thus, it should be noted that although all methods show great promise for generating comprehensive protein interaction maps, significant improvements have to be made before this goal is reached. Currently, no method covers more than 60% of the yeast genome and all methods are still plagued by shortcomings, such as high rates of false-positives and false-negatives, in addition to bias

towards certain protein classes. Currently, these drawbacks can only be overcome by combining the output of as many independent methods as possible, scoring only interactions that are found in datasets of at least three methods. Only in this way can an accuracy of ~80% be reached [21]. As methods improve and new approaches are being adapted to high-throughput use (see below), the coverage and quality of interaction maps will improve to the point where accurate prediction of protein interactions will be possible.

Increasingly, data on protein interactions are also collected and annotated in different databases (Box 2). Similar to the NCBI's GenBank, these databases provide an interface for querying protein interactions in an easy and intuitive way, and also provide many bioinformatics tools to explore and compare the wealth of data derived from different large-scale protein interaction screens.

Conclusion and future directions

The complexity of protein interactions present in differing cell types, in different times of the cell cycle and under differing environmental conditions renders an exhaustive characterization of the proteome unfeasible. Currently, the goal of proteomics is to provide a benchmark to allow the reliable prediction of protein interaction partners that cooperate in a given cellular task. This benchmark will continue to be scrutinized through conventional genetic and biochemical analyses, and this 'yin' and 'yang' between proteomic interaction techniques and conventional genetic and biochemical analysis will generate progress for both disciplines. Conventional genetic and biochemical analyses will serve to evaluate the validity of particular proteomic techniques while new avenues of research for conventional analysis will continue to be developed.

At this time, if single proteomic datasets are compared to existing conventional genetic and biochemical data, the validity of any single proteomic technique is debatable [21]. However, when specific protein interactions can be confirmed by two or three separate proteomic techniques, the quality of the data are unquestionably superior. Therefore, to generate a better predictor of protein–protein interactions, the limitations of existing techniques must be realised and either overcome or be complemented by additional proteomic techniques. For example, available methods have shown a limitation in the isolation and analysis of membrane proteins [28]. Therefore, techniques such as the membrane-based split-ubiquitin system [29,30], which is able to detect protein interactions among proteins situated within membranes, could possibly produce data to fill in the existing knowledge gaps.

Other detection systems that can be useful in further proteomic studies are the bioluminescence resonance energy

transfer (BRET) and fluorescence energy transfer (FRET) methodologies [31]. In the BRET method, the bait protein is fused to *Renilla* luciferase whereas the prey would be fused to another fluorescent protein [i.e. a member of the green fluorescent protein (GFP) family of proteins]. In the presence of the substrate for luciferase, if a protein–protein interaction occurs the resonance energy of the luciferase signal is transferred to the fluorescent tag of the prey protein, which emits an additional signal that is detected. If no protein interaction occurs, only the luciferase signal is detected. In the FRET system a similar transfer of resonance energy occurs; however, the input energy is derived from an external monochromatic wavelength that activates a fluorophore, thus replacing the luciferase. An advantage of the FRET system compared with BRET is the ability for direct visualization of protein–protein interactions within living cells. Unfortunately in the FRET system, owing to the overlapping absorption and emission spectra of the existing GFP mutants attached to the prey, direct excitation of the acceptor fluorophore by the light used to excite the donor often complicates the interpretation of the results.

As genome sequences continue to be completed and the amount of interaction data increases, a growing number of computational methods for predicting protein interactions continue to emerge (Box 3). These techniques compare the presence or absence of genes within genomes, the spatial relationships among genes (i.e. gene neighborhoods, gene fusion events and co-evolution among protein pairs) to assess the possibility of protein interactions. Furthermore, when the molecular structure of two proteins is known, the molecular prediction (or docking problem) of protein interactions can be scrutinized. Therefore, as more genomic, structural and protein interaction data become available the ability to predict protein interactions *in silico* will be strengthened.

Systems that do not directly detect or attempt to predict protein–protein interactions can also be useful in corroborating existing protein–protein interaction data. Synthetic lethal screening in *S. cerevisiae*, which tests for lethal deletion combinations within haploids of otherwise viable single deletion mutants, has been an effective means of grouping genes into cellular pathways [32]. Also, techniques such as comparative mRNA expression analysis where genes are grouped according to their expression profiles strengthens a putative protein–protein interaction pair [33]. Furthermore, a prerequisite for protein–protein interactions is that the two protein partners are localized to the same cellular compartment. Recent studies in *S. cerevisiae* have determined the cellular localization of 2744 yeast proteins via high-throughput immunolocalization of tagged yeast gene products [34]. Lastly, techniques to

Box 3. Computational methods used to predict protein interactions

Phylogenetic profiles

The presence or absence of genes within entire genomes is used as a predictor of the need of corresponding protein pairs to perform a particular task. Disadvantages include the fact that an entire genome sequence is required for comparisons, essential genes common to most organisms cannot be compared and this method often suggests a functional role but does not imply direct interactions between proteins.

Conservation of gene neighborhoods

Bacterial genomes tend to be organized into regions that encode functionally related proteins (i.e. operons), and thus the organization of the genes in various bacterial genomes is used to predict functional relationships. However, at present, this method can only be applied to bacterial genomes.

Gene fusion events

Interactions between proteins can be deduced if two separate genes (A and B) in one organism are fused into one gene (AB) in another organism. However, this method is restricted to shared domains on distinct proteins, which are often two characteristics that are not easily determined.

Co-evolution among protein pairs

This method is similar to phylogenetic profiles but takes into account the entire structure of phylogenetic trees. In certain cases, corresponding phylogenetic trees of interacting proteins tend to show a greater degree of similarity than do non-interacting proteins. On the negative side, it is often difficult to obtain good high-quality sequences to perform multiple sequence alignments, which should include two or more sequences from the same species to produce the phylogenetic tree.

characterize the presence and/or effects of protein modifications, such as glycosylation and phosphorylation, will enable researchers to better understand the dynamics and requirements of protein–protein interactions [35]. Towards this goal the DIP (Database of Interacting Proteins) has been expanded to LiveDIP, which describes protein interactions by protein states and state transitions [36].

With the completion of genome sequences and therefore the acquisition of the genes involved in the development of organisms and their maintenance and growth, a new era of biology has emerged. The proteomics branch of this revolution has generated tremendous amounts of data that help to improve our understanding of protein interactions and thus provide insight into complex cellular processes. Utmost scrutiny, impeccable documentation and creative advances will grant proteomics a fruitful future.

RESEARCH FOCUS

Acknowledgements

We thank John Miller and Peter Uetz for helpful discussions, and Ulrich Hübscher for his support. The I.S. group is supported by Zürcher Krebsliga, Gebert-Rüf Foundation, Olga Mayenfisch Foundation, Walter Honegger Foundation, Bonizzi-Theler Foundation, EMDO Foundation, Novartis Foundation, Sassella Foundation, Fonds für medizinische Forschung, Kommission für Technische Innovation (KTI, Nr. 5343.2 SUS) and the Swiss National Science Foundation (31-58798.99).

References

- 1 Fleischmann, R.D. *et al.* (1995) Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd. *Science* 269, 496–512
- 2 The human genome sequencing consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409, 860–921
- 3 Venter, J.C. *et al.* (2001) The sequence of the human genome. *Science* 291, 1304–1351
- 4 Goffeau, A. *et al.* (1996) Life with 6000 genes. *Science* 274, 546–547
- 5 The *C. elegans* sequencing consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282, 2012–2018
- 6 Adams, M.D. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287, 2185–2195
- 7 The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408, 796–815
- 8 Gardner, M.J. *et al.* (2002) Sequence of Plasmodium falciparum chromosomes 2, 10, 11 and 14. *Nature* 419, 531–534
- 9 Hall, N. *et al.* (2002) Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13. *Nature* 419, 527–531
- 10 Hyman, R.W. *et al.* (2002) Sequence of Plasmodium falciparum chromosome 12. *Nature* 419, 534–537
- 11 Waterston, R.H. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562
- 12 McCraith, S. *et al.* (2000) Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc. Natl. Acad. Sci. U. S. A.* 97, 4879–4884
- 13 Flajolet, M. *et al.* (2000) A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* 242, 369–379
- 14 Rain, J.C. *et al.* (2001) The protein-protein interaction map of *Helicobacter pylori*. *Nature* 409, 211–215
- 15 Walhout, A.J. *et al.* (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* 287, 116–122
- 16 Uetz, P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627
- 17 Ito, T. *et al.* (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4569–4574
- 18 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147
- 19 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183
- 20 Zhu, H. *et al.* (2001) Global analysis of protein activities using proteome chips. *Science* 293, 2101–2105
- 21 von Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403
- 22 Zhu, H. *et al.* (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genet.* 26, 283–289
- 23 Fields, S. and Song, O. (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340, 245–246
- 24 Finley, R.L., Jr and 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
- 25 Auerbach, D. *et al.* (2002) The post-genomic era of interactive proteomics: facts and perspectives. *Proteomics* 2, 611–623
- 26 Link, A.J. *et al.* (1999) Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17, 676–682
- 27 MacBeath, G. and Schreiber, S.L. (2000) Printing proteins as microarrays for high-throughput function determination. *Science* 289, 1760–1763
- 28 Stagljar, I. and Fields, S. (2002) Analysis of membrane protein interactions using yeast-based technologies. *Trends Biochem. Sci.* 27, 559–563
- 29 Stagljar, I. *et al.* (1998) A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. *Proc. Natl. Acad. Sci. U. S. A.* 95, 5187–5192
- 30 Wittke, S. *et al.* (1999) Probing the molecular environment of membrane proteins *in vivo*. *Mol. Biol. Cell* 10, 2519–2530
- 31 Boute, N. *et al.* (2002) The use of resonance energy transfer in high-throughput screening: BRET versus FRET. *Trends Pharmacol. Sci.* 23, 351–354
- 32 Tong, A.H. *et al.* (2001) Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* 294, 2364–2368
- 33 Hughes, T.R. *et al.* (2000) Functional discovery via a compendium of expression profiles. *Cell* 102, 109–126
- 34 Kumar, A. *et al.* (2002) Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719
- 35 Shaywitz, A.J. *et al.* (2002) Analysis of phosphorylation-dependent protein-protein interactions using a bacterial two-hybrid system. *Sci. STKE* 1, PL11
- 36 Duan, X.J. *et al.* (2002) Describing biological protein interactions in terms of protein states and state transitions: the LiveDIP database. *Mol. Cell. Proteomics* 1, 104–116

Conference reports

TARGETS is pleased to publish the highlights from international conferences. Conference participants who wish to cover a particular meeting should contact:

Dr Chris Watson, Drug Discovery Today Publications, Elsevier, 84 Theobalds Road, London, UK WC1X 8RR
e-mail: TARGETS@drugdiscoverytoday.com