Improving yeast two-hybrid screening systems

Manfred Koegl and Peter Uetz

Abstract

Yeast two-hybrid (Y2H) screening methods are an effective means for the detection of protein–protein interactions. Optimisation and automation has increased the throughput of the method to an extent that allows the systematic mapping of protein–protein interactions on a proteome-wide scale. Since two-hybrid screens fail to detect a great number of interactions, parallel high-throughput approaches are needed for proteome-wide interaction screens. In this review, we discuss and compare different approaches for adaptation of Y2H screening to high-throughput, the limits of the method and possible alternative approaches to complement the mapping of organism-wide protein–protein interactions.

Keywords: Yeast two-hybrid; protein–protein interactions; interaction proteome

INTRODUCTION

To understand the function of a protein, it is useful to know to which other proteins it can bind. For decades, this simple idea has been motivating researchers to look for binding partners of their favourite proteins. Since the biochemical isolation of protein complexes is a tedious and demanding process, alternative methods to find potential binding partners are welcome. Yeast two-hybrid (Y2H) screening [1] has emerged as the most successful of these methods, and has been quickly and widely accepted by the research community. The method has been automated and used in several large-scale projects, including the first drafts of protein interaction maps for humans and several model organisms. In the following, we compare different approaches for adaptation of the method to high-throughput processing, discuss the limits of the method, ways to select reliable interactions from the mass of the screening data and alternative approaches to complement the mapping of organism-wide protein–protein interactions.

HOW THE Y2H SYSTEM WORKS

The basic idea of all two-hybrid methods is to split a protein in two halves that do not work independently, but will work if they can be somehow brought together again. When the two fragments are expressed as fusion proteins ('hybrids') with two other proteins that have sufficient affinity for each other, the two parts of the split protein are combined again and its function is restored.

In the most common application of this idea a transcription factor is split into the separate domains that harbour (i) the DNA-binding activity and (ii) the transcriptional activation function (Figure 1). The reconstitution of the transcription factor is detected by the transcriptional activation of a reporter gene. Commonly used reporters generate a colorimetric or fluorescent readout, or allow growth on selective media. For example, in an yeast strain that lacks a functional *HIS3* gene the wild-type *HIS3* gene as a reporter allows for the selection of interaction-positive colonies in histidine-free medium.

Corresponding author. Dr M. Koegl, Translational Research, and Genomics and Proteomics Core Facilities, German Cancer Research Center, Im Neuenheimer Feld 515, 69120 Heidelberg, Germany. Tel: +49 6221 42 47 10; Fax: +49 6221 42 47 04; E-mail: m.koegl@dkfz.de

Manfred Koegl has a long-standing interest in protein–protein interactions. He is currently heading the protein interaction core facility at the German Cancer Center in Heidelberg.

Peter Uetz has pioneered large-scale yeast two-hybrid screening for several model organisms and will continue his work in his new position at the J Craig Venter Institute.
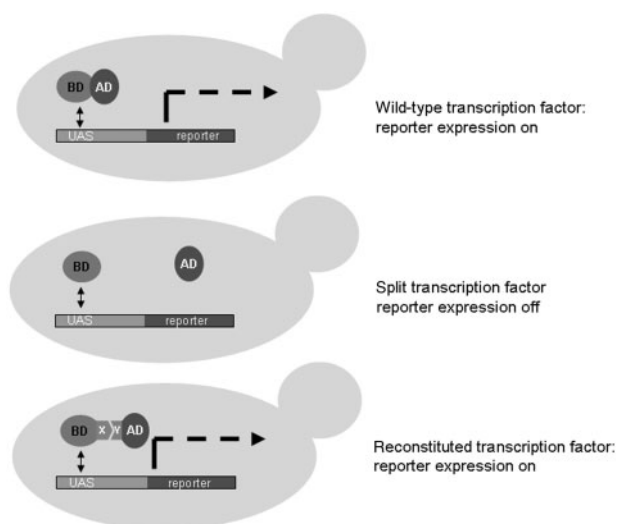
Figure 1: Principle of the Y2H method. Top panel, a yeast cell expressing a transcription factor consisting of an activation domain (AD) and a DNA-binding domain (BD). The BD binds the upstream activating sequence (UAS) and expression of an adjacent reporter gene is activated. Middle panel, the two domains of the transcription factor are expressed separately. The AD is not recruited to the promoter of the reporter gene and transcription is not activated. Bottom panel, two proteins, X and Y, are expressed as fusions with the AD and BD. Interaction of X with Y leads to recruitment of the AD to the promoter. The activation of transcription from the reporter is interpreted as a readout for the interaction of X with Y. The protein fused to the BD is generally referred to as the bait, the protein fused to the AD called the prey.

FISHING IN POOLS OF cDNAS

Reporters that result in selective cell growth allow the enrichment of positive colonies against a background of negative cells. Using this method, complex libraries can be screened for interacting 'prey' proteins with a 'bait' protein of interest. In the early applications of Y2H screening, cDNA pools were based on oligo (dT) or random primed cDNAs prepared from the mRNA of diverse tissues, and cloned into a plasmid suitable for Y2H screening [2, 3]. In case of yeast or prokaryotes, fragmented genomic DNA can be used instead of cDNA [4, 5]. For screening, library clones are pooled, and yeast cells harbouring interacting bait and prey proteins are enriched by use of reporters such as HIS3.

Screening of pooled libraries has been the typical use of the Y2H system in academic labs aiming at the isolation of binding partners for a protein of interest. As suggested by Figure 2, this method is still widely

and successfully applied. A disadvantage of libraries created by the cloning of pools of DNA fragments is the uncontrolled fashion in which the coding sequences of the inserts are attached to the coding sequence of the split transcription factor. In many cases, the hybrid protein will be expressed in the wrong reading frame or from the 5′ or 3′ untranslated regions of the mRNA. The resulting non-natural proteins provide a rich source for non-specific interactions that often litter the results of Y2H screens, and add to the number of false positives that occur in Y2H screens. To minimize false positives, the molecular details of the method, such as the reporter gene constructs and expression vectors for hybrid proteins, have been fine-tuned in many aspects (reviewed in [6–9]), significantly reducing the noise of non-biological interactions. For an initial filtering of the raw interaction data, several technical parameters from Y2H screens are useful. These include the number of different reporters activated by an interaction event, and the level of reporter gene activation. Interactions that do not get past the hurdle of these criteria are usually not reported in publications although some authors have argued that all raw data (including possible false positives) should be released so that they can be used for further improvement of filtering strategies [10].

ARRAYS OF PREYS

Automation of the rate-limiting steps of the method, such as plating of cells for the selection of positives, picking of positive clones and determination of the interaction signal allows taking Y2H screening to a larger scale, including systematic analyses of protein–protein interactions of whole organisms. As shown in Figure 2, such systematic screens make up a sizeable proportion of currently reported protein–protein interactions identified by Y2H methods. A prerequisite of large, systematic Y2H screening is the availability of cDNA clones encompassing the coding regions of the bait proteins in a suitable vector. Collections of individually cloned cDNAs comprising the full-length open reading frame (ORF) of the mRNA are currently being generated for several species (reviewed in [11–13]), in part in dedicated efforts to provide resources for Y2H screening [14–16]. The use of recombinational cloning systems facilitates the shuttling of the coding systems between vectors, such that the ORFs can be readily transferred to plasmids appropriate for the expression of fusions
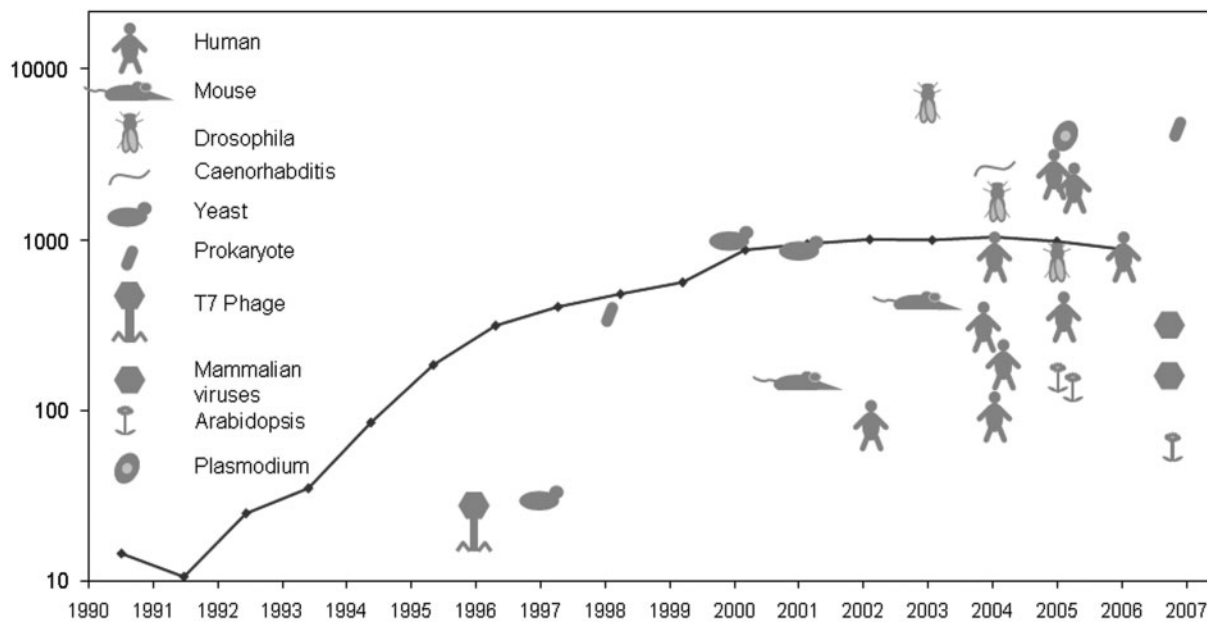
Figure 2: Chronological representation of published Y2H-based data. Solid line, number of papers found in PubMed in any field using 'two hybrid' as a search term, per year. This number is a rough approximation of scientific papers using the method. Very few papers use the term in different contexts than the Y2H system and its use in the identification of novel protein–protein interactions, as can be seen from the searches in the years before the method had been established. A caveat is that not all papers will report the use of the method in the searchable fields in PubMed. Symbols display large systematic interaction studies using the Y2H method, their position on the Y-axis represents the number of high-confidence interactions reported. Note that the numbers are not strictly comparable, since the selection criteria for high-confidence interaction differ among studies. Sources, sorted by year of publication are, 1996: [26], 1997: [19], 2000: [17], 2001: [18, 20, 70], 2002: [36], 2003: [22, 71], 2004: [30, 31, 33, 34, 38, 72], 2005: [23, 24, 27, 29, 32, 39, 73], 2006: [28, 35], 2007: [21, 40, 74].

with the DNA-binding domain or activation domain in yeast. Such 'ORFeome' collections allow a novel strategy for Y2H screening: instead of enriching interacting clones from a mixed pool, the individual clones are tested one by one for an interaction with the bait protein. Typically, the cDNA collection is presented in an arrayed form, and each position in the array is tested pair-wise for interaction signals with a bait protein (see also Figure 3).

This approach has several advantages to the screening of pooled preys (see also Table 1):

- The identity of the arrayed proteins is known, such that it is not necessary to isolate and sequence the library insert.
- The absence of fusions that are in the wrong reading frame or correspond to non-coding DNA avoids interaction signals from non-natural peptides.
- The library is normalised with respect to the representation of each protein. This is in stark contrast to classical cDNA libraries, in which cDNAs from highly expressed mRNAs are

overrepresented and cDNAs from lowly expressed mRNAs are rare, which results in difficulties to find interactions with lowly expressed mRNAs.

- Pair-wise tests for interactions are more sensitive than screens of large cDNA pools, probably because weak signals can be distinguished from background more easily. For example, the number of interactions found by Uetz *et al.* [17] was much larger than the number found by the pooled library used by Ito *et al.* [18] when the same bait was used. However, random libraries regularly find more interactions than array screens because they include fragments that may interact while full-length proteins may not (e.g. Fromont-Racine *et al.* [19] versus Uetz *et al.* [17] or Rain *et al.* [20] versus Parrish *et al.* [21]).

The number of pair-wise tests in such a matrix screen increases with the square of the number of proteins in the matrix. This is the reason why in practice, most large-scale projects have initially screened mini-pools of clones, rather than protein pairs and then further analysed them by sequencing
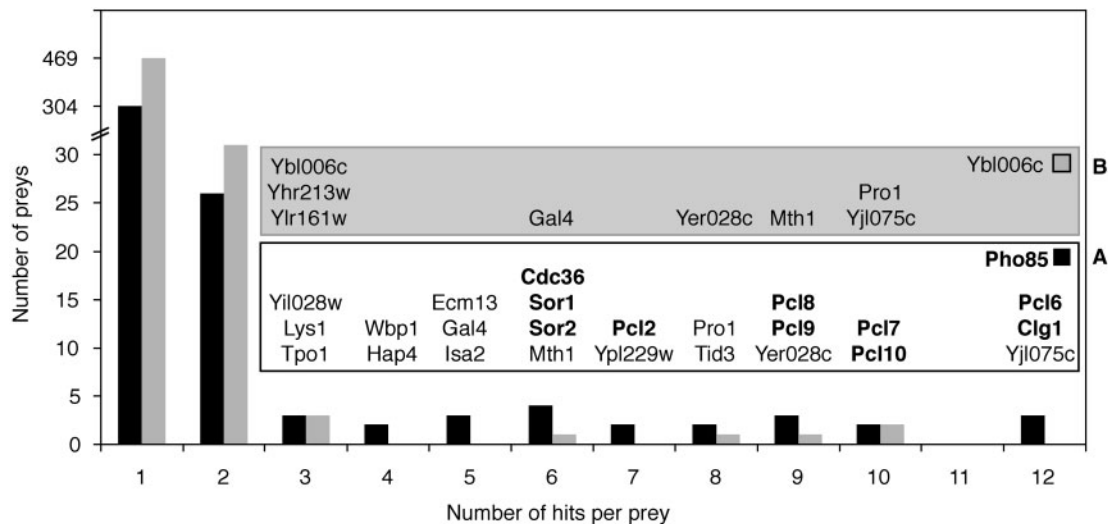
Figure 3: Reproducibility and specificity of Y2H screens. Results of two independent Y2H screens of a genome-wide yeast prey array with PHO85 (black) and YBL006C (grey) as baits. The two screens were repeated 14 and 12 times, respectively. The X axis represents the number of times a given prey was found, the Y axis represents the number of proteins that were found with the respective frequencies. The 14 PHO85 screens generated a total of 354 distinct positive proteins of which 304 were found only once, 26 were found twice, 3 were found 3 times and so on (black bars). Only positives that were found in at least 4 screens were considered as reproducible. For example, PCL6 was found in 12 out of 14 screens. Fourteen preys were reproducible but not 'specific' as they were also found in many (here [3] 50) screens with other baits and are thus considered as unspecific false positives (protein names in regular type). Proteins that were found with fewer than 50 baits were considered as 'true positives' (protein names in **bold** type). 'Reproducibility' and 'specificity' are powerful criteria to identify 'true' positives as is shown by the PCL proteins which are cyclins known to bind to the PHO85 kinase (CLGI is a non-PCL-type cyclin). The biological significance of PHO85's interaction with CDC36 (a transcription factor), SORI, and SOR2 (Sorbitol dehydrogenases) is not known. In contrast to PHO85, YBL006C is a bait that did not produce any reproducible preys despite the fact that it yielded a total of 508 positive preys in all screens combined, all of which are considered as unspecific and thus 'false positive' proteins (i.e. these proteins were found in many screens using unrelated baits). Note that the three proteins that were found three times (including YBL006C itself) may be weak interactors of YBL006C (which actually may form homodimers) because they were very specific (each of them was found with only 2 baits).

[17–19, 22, 23] or selective pair-wise tests [17]. More recently, several studies used smart pooling strategies [79], pools of baits [24] or preys [25] which were de-convoluted to obtain individual protein pairs after mating and selecting these pools.

LARGE–SCALE PROTEIN INTERACTION SCREENING

Both approaches, screening of pooled libraries as well as matrix-type screening of arrayed cDNA libraries, have been automated and used for large-scale interaction maps (Figure 2). An early project dealing with the intra-viral protein interactions of the bacteriophage T7 showed that large protein interaction mapping projects are feasible [26]. This first step was soon followed by large-scale protein–protein interaction mapping projects for bacteria (*Helicobacter*

pylori [20], *Campylobacter jejuni* [21]), yeast [17–19], plants [27], human viruses [28], *Plasmodium falciparum* [29] and higher eukaryotes (*Caenorhabditis elegans* [30], *Drosophila melanogaster* [22, 31, 32]). Several protein interaction networks for human proteins have been generated for specific areas of interest, such as signal transduction and biochemical pathways [33–35], protein families [36–39], subcellular structures or virus–host interactions [40]. Two groups have recently reported unbiased large interaction screens with the goal of outlining the first draft of the human interactome [23, 24]. All these data have proven to be rich sources of biologically relevant information.

ASSESSMENT OF Y2H DATA

In classical projects, Y2H-based data were only published once the interactions had been tested and

Table I: Preys in cDNA pools versus arrays of preys

	Pools	Arrays
Detection of interactions	Selective growth of positive clones→ enrichment from large pools	Pair-wise tests
Clone identification	Sequencing of the library insert	Position on the array encodes the identity of the insert
Library complexity (typically)	Several million	Few to thousands
Libraries screened (typically)	Randomly cloned cDNA fragments	Individually cloned full-length ORFs
Number of tests in systematic screens	Number of screens required is directly proportional to the number of baits (but more clones need to be analysed per screen)	Number of tests required increases with the square of the number of proteins to be analysed
Promiscuous preys	Recognised upon repeated screening of the library. Cannot be removed from the pool	Recognised upon repeated screening of the library, and removed.
Saturation	Hard to approach (e.g. ref. 78: saturation is reached in >500 screens)	Saturation can be approached in a few screens.

confirmed in independent experiments. This has simply not been possible for large-scale Y2H experiments, since the acceleration in data production by Y2H analysis has not been matched yet by the improvements of 'confirming' methods, such as co–immunoprecipitations. Thus, the bad news is that the new data sources are afflicted with uncertainties that need to be taken into consideration for their use. The good news is that the sheer mass of data allows the selection of reliable data by quantitative, partially statistical criteria. Such criteria mainly include the reproducibility of the interaction and the definition and exclusion of promiscuous interactors, as outlined in the following two sections.

TECHNICAL VERSUS BIOLOGICAL ARTEFACTS

For the discussion of artefacts and their elimination, it is helpful to distinguish technical artefacts, in which an interaction signal is generated by events other than a protein–protein interaction, from biological artefacts, where proteins truly interact, but only when artificially co-expressed [41]. For example, proteins may interact in a Y2H assay without ever being naturally expressed in the same cell. In contrast to technical artefacts, biological artefacts are genuine interactions of bait and prey, and cannot be eliminated by technical controls. In fact, when tested in alternative protein interaction assays, biological false positives will mostly be confirmed. Also, it is hard to define false positives with certainty, since it is impossible to give experimental proof that two proteins do under no instances bind to each other.

CRITERIA FOR SELECTING RELIABLE INTERACTIONS

We will discuss five categories of selection methods for reliable interactions, which are based on (i) the reproducibility of interactions, (ii) the promiscuity of interaction partners, (iii) network topology, (iv) comparisons with external data and (v) evolutionary conservation of interaction partners.

- **Reproducibility**: Most technical artefacts are either reproducible, or rare. Rare artefacts can arise e.g. from mutations that artificially generate interaction signals. The likelihood that a rare event occurs twice independently in cells harbouring cDNAs from the same protein is extremely low. Thus, the removal of interactions that are not reproduced within the data set can be used to weed out such rare technical artefacts [22, 30, 39, 42–44].

- **Promiscuity**: Reproducible artefacts are e.g. interaction signals that arise from non-specific binding of the prey to the bait protein chimera. Such artificial activators of the reporter genes become apparent as 'promiscuous' preys when a library or an array is repeatedly screened, since they appear to bind to a great number of unrelated baits (Figure 3). These artefacts can be eliminated from the data set by removing all preys that display promiscuities above a threshold level. The cut-off for promiscuity, i.e. the cut-off line for how many interaction partners are allowed before a protein has to be considered promiscuous, is an arbitrary number. Low cut-off values for exclusion from the data set will increase the number of reliable interactions in the remaining data, but at the expense of increasing the rate of false negative interactions [22, 39].

- **Topology**: The definition of a cut-off promiscuity value as a criterion to exclude interactions from
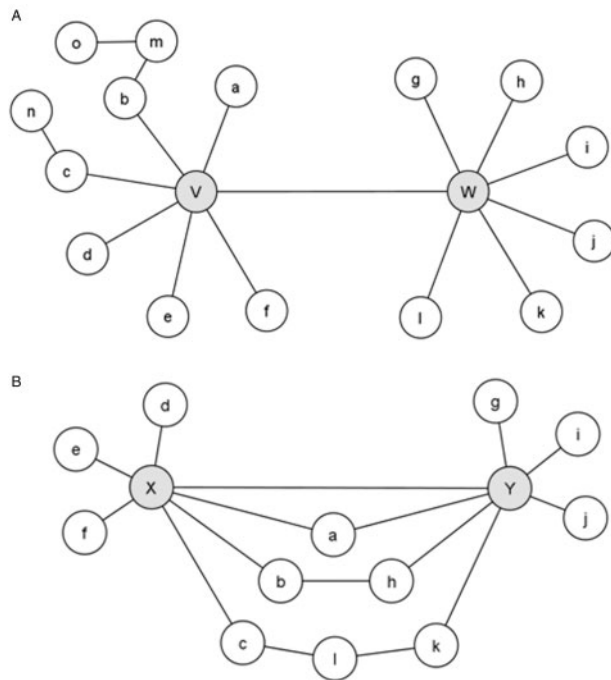
Figure 4: Network topologies can be used to enrich for relevant interactions. Two hypothetical pairs of interacting proteins, V–W and X–Y are shown. The promiscuities of the proteins are equal in both examples. In the top panel, there is no alternative path from V to W. In the bottom panel, the existence of several alternative paths with short path lengths between X and Y lends a higher confidence to this interaction than to the interaction of V and W.

the data set is problematic. Many proteins have a large number of genuine natural binding partners, and will be erroneously excluded from the network based on their apparent promiscuity. When the interaction network is large enough, or can be integrated with external data sets into an existing larger network, topology metrics can be used to correct for that. These metrics test whether the binding partners of a protein are connected to each other. For example, the number of common binding partners of an interaction pair is a positive indicator of interaction reliability (see Figure 4 for an example) [22]. More complex algorithms that calculate weighted alternative path lengths for protein pairs to derive confidence measures [45, 46], or that score local topologies [47] or clusters [24, 44], have been shown to be useful in the selection of relevant interactions.

- **Indirect support**: Comparisons with external data sets have shown that proteins that bind to each other have a higher than average likelihood to be involved in related cellular functions, are more

likely to be expressed at the same time, and to interact genetically with each other [23, 48–51]. These criteria are most useful to assess the overall quality of a data set, and to test the usefulness of selection criteria [52].

- **Conservation**: Lastly, interactions have been shown to be more likely if they are conserved in evolution, as evidenced by paralogous or homologous interacting proteins [24, 39, 48].

LIMITATIONS OF THE Y2H SYSTEM

Many natural protein–protein interactions cannot be detected using the Y2H method. Some proteins do not interact in the environment of the yeast nucleus, such as proteins of the secretory compartments that require oxidative conditions or glycosylation for proper folding. Integral membrane proteins are unlikely to work in the context of reconstituted transcription factor. Many interactions are triggered by post-translational modifications not available in yeast. Other proteins, such as active tyrosine kinases, are toxic to yeast when expressed to high levels, and cannot be used as baits. For these reasons, the rate of interactions not detectable by the Y2H is substantial (e.g. [18, 53]). Rajagopala *et al.* [54] estimated that their array-based Y2H screens found only 23% of previously known interactions involving motility proteins of *Treponema pallidum*. When data from another screen in *Campylobacter jejuni* were added, this fraction rose to 33%. However, many additional interactions were found.

But at least within the limits of the method, it would be desirable that screens be exhaustive, i.e. that they identify all interactions that can be identified by use of the Y2H method. Screens of pooled libraries can only asymptotically approximate saturation. Given that those libraries have complexities of several millions, and weakly expressed proteins are underrepresented, most screens are subsaturating. In contrast, array-based Y2H screens can theoretically be comprehensively screened. However, comparisons of the presently available data sets for yeast (see Figure 3 for an example) [17, 18, 53], fly [22, 31, 32] and man [55] show that in all cases, the overlap of interacting data is minimal, mainly due to the fact that most of the screens are far from exhaustive. Moreover, variations in the details of the Y2H protocol, such as the vectors used, the nature of the re-constituted transcription factor and the libraries screened, have a great impact on the

interactions that can be retrieved. Evidently, due to the relatively low detection rate of the Y2H system, other methods will be needed to approach the complete mapping of human proteome, or that of model organisms. Apart from biochemical fractionation of protein complexes followed by mass spectrometry to analyse their components, several other methods may be apt for the task.

ADDING EDGES: ALTERNATIVE PROTEIN INTERACTION ASSAYS

At the time of inception of the Y2H method, arrayed libraries were not available for screening in pair-wise interaction tests. Interacting protein pairs had to be isolated from complex mixtures of proteins or from complex libraries, and one of the great advantages of Y2H screening compared to other interaction tests was its ability to enrich for clones of interacting proteins from a large pool. The availability of ORFeome collections and the development of methods that allow thousands of pair-wise interaction tests in parallel make this advantage somewhat obsolete. Additional methods now become applicable to matrix-type interaction screens, although their advantages or disadvantages will only become clearer when more data is available. Three of them are discussed subsequently.

PROTEIN AND PEPTIDE MICROARRAYS

Microarrays have led to a tremendous parallelisation in the analysis of nucleic acids. For proteins, microarray technology (reviewed in [56]) is still in an earlier stage of development. Problems with the expression, purification, storage and stability of large sets of native proteins still severely hamper progress in the field. To date, proteome-wide arrays useful for protein interaction studies have been generated only for yeast proteins. In a pioneering project, these arrays were used to identify novel calmodulin-binding proteins [57]. Apart from yeast proteins, protein interaction studies using protein microarrays have been centred on particular protein families or domains, such as the SH2 domain [58] or the PDZ domain [59]. Possibly, the use of nucleic acid programmable protein arrays (NAPPA) may provide a route for the cost-effective generation of protein chips useful for the study of protein–protein interactions [60]. For NAPPA chips, DNA molecules are spotted that guide the *in situ* production of

recombinant proteins by a coupled *in vitro* transcription and translation reaction. The expressed tagged proteins are captured by specific antibodies spotted onto the same spot as the DNA. The immobilised array can be probed for binding with an alternatively tagged soluble protein. In a proof-of-concept experiment, this method has been used to detect protein–protein interactions among 29 human replication initiation proteins [60].

PROTEIN COMPLEMENTATION AND ALTERNATIVE TWO-HYBRID ASSAYS

Using a similar principle as Y2H tests, protein complementation assays (PCAs) use two proteins tagged with two fragments of a reporter protein (reviewed in [61]). Upon interaction of the proteins, the two fragments can reconstitute the active reporter protein, providing a readout for the interaction. A great variety of proteins lend themselves to use in PCAs [61]. Some of them have direct read-outs, such as luciferase enzymes [62, 63], other have indirect read-outs, such as the split ubiquitin ([64] reviewed in [65]) or the split tobacco etch virus (TEV) system [66]. The split ubiquitin system is well matured, and has been applied in a successful effort to map hundreds of protein–protein interactions involving yeast membrane proteins [67]. In the split TEV system, the split enzyme is a protease from TEV. This amino acid motif recognised by this protease is absent in mammalian proteins, such that it can be expressed in a mammalian cell without inflicting damage on the cellular proteins. Activation of TEV causes the liberation of a transcription factor from an inactive complex, which can be read out directly by reporter proteins (Figure 5A). Another alternative two-hybrid assay that has been shown to be amenable to cDNA library screening is the MAPPIT system, which is based on the interaction-dependent activation of STAT transcriptional regulators by a chimeric receptor coupled with transcriptional reporters [68].

QUANTITATIVE CO-PRECIPITATION USING LUCIFERASE-TAGGED PROTEINS

The most straightforward concept to test for a protein–protein interaction is to purify one of the proteins, and test for the presence of the other.
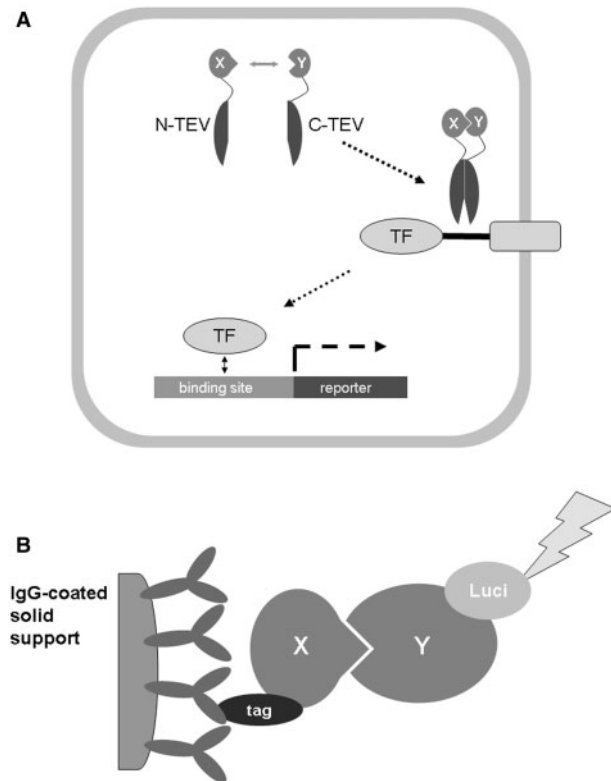
Figure 5: (**A**) The split TEV system as an example for a protein-fragment complementation assay. Two proteins, X and Y, are expressed in fusion with the amino- and C-terminal fragments of the protease TEV. Upon binding of X to Y the two fragments of the protease unite, and the enzyme gains activity. A transcription factor is connected to a membrane anchor via linker that has the amino acid recognition site of TEV. Cleavage of the linker by the reconstituted TEV releases the transcription factor from its membrane anchor to activate a reporter gene in the nucleus. (**B**) Precipitations of luciferase-tagged proteins [69]. Here, the first protein X of an interacting pair is tagged with an epitope recognised by an antibody, the second protein is expressed as a fusion protein with luciferase. Both proteins are co-expressed in a mammalian cell. Extracts are allowed to bind to a solid support coated with an antibody to the tag on the protein X. Unbound material is washed away, and the retained luciferase activity is taken as a measure for the binding of Y to X.

In un-biased assays using mass spectrometric analyses of the co-precipitated material, such protocols have been the basis of the discovery of many protein–protein interactions, and used in large-scale projects for yeast [75, 76] and mammalian [77] protein complex mapping. For more straightforward detection of the binding partner, proteins can be fused to more easily detectable proteins, such as luciferase. In this case, pair-wise interactions are tested in

dedicated assays. Barrios-Rodiles *et al*. [69] miniaturized this assay and applied it to the analysis of protein–protein interactions in TGF-β signal transduction. This method is quick and cost-effective enough to allow for proteome-wide interaction screens. As in the MAPPIT and the split TEV system, interactions are isolated from a physiological environment, which is beneficial when interactions need to be tested after a regulatory event, e.g. cytokine stimulations (Figure 5B).

OUTLOOK

Despite a large number of interactions deposited in specialised databases there is still no complete interactome available for any organism. Driven by current efforts in large-scale Y2H screening, the availability of ORFeome collections and novel methods for the detection of protein–protein interactions, we expect such interaction maps to become available for a few model organisms in the near future. Overlapping protein interaction data sets gathered by independent methods will increase the confidence in those interactions that are detected by more than one method. Large-scale affinity purification projects coupled with mass spectrometry analyses will also complement the map of protein interactions and elucidate the composition of complexes which are stable enough to survive the purification processes.

Prior to the introduction of Y2H screening, *identifying* a potential interaction has been the rate-limiting step in many projects, and the minimal merit of the method is that the rate-limiting step has been shifted to confirming an interaction's significance. With the availability of *confirmed* protein interaction data in public databases, this obstacle will be removed as well, and the rate-limiting step will be shifted towards understanding the biological function of the interactions.

Key Points

- Large-scale Y2H screening projects are currently used to build the first proteome-wide binary protein–protein interaction maps.
- Availability of proteome-wide repositories of expression clones facilitates protein interaction screens by Y2H and other methods such as automated co-purification assays.
- Statistical filtering of large protein interaction data sets allows to define high-confidence protein interaction data.
- Novel methods are waiting in the wings and will increasingly contribute to the comprehensive mapping of protein–protein interactions.

References

1. Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature* 1989;**340**:245–6.

2. Chevray PM, Nathans D. Protein interaction cloning in yeast: identification of mammalian proteins that react with the leucine zipper of Jun. *Proc Natl Acad Sci USA* 1992;**89**:5789–93.

3. Staudinger J, Perry M, Elledge SJ, *et al.* Interactions among vertebrate helix-loop-helix proteins in yeast using the two-hybrid system. *J Biol Chem* 1993;**268**:4608–11.

4. Chien CT, Bartel PL, Sternglanz R, *et al.* The two-hybrid system: a method to identify and clone genes for proteins that interact with a protein of interest. *Proc Natl Acad Sci USA* 1991;**88**:9578–82.

5. Yang X, Hubbard EJ, Carlson M. A protein kinase substrate identified by the two-hybrid system. *Science* 1992;**257**:680–2.

6. Bartel P, Chien CT, Sternglanz R, *et al.* Elimination of false positives that arise in using the two-hybrid system. *Biotechniques* 1993;**14**:920–4.

7. Serebriiskii IG, Golemis EA. Two-hybrid system and false positives. Approaches to detection and elimination. *Methods Mol Biol* 2001;**177**:123–34.

8. Vidalain PO, Boxem M, Ge H, *et al.* Increasing specificity in high-throughput yeast two-hybrid experiments. *Methods* 2004;**32**:363–70.

9. Gietz RD. Yeast two-hybrid system screening. *Methods Mol Biol* 2006;**313**:345–71.

10. Hart GT, Ramani AK, Marcotte EM. How complete are current yeast and human protein-interaction networks? *Genome Biol* 2006;**7**:120.

11. Brizuela L, Richardson A, Marsischky G, *et al.* The FLEXGene repository: exploiting the fruits of the genome projects by creating a needed resource to face the challenges of the post-genomic era. *Arch Med Res* 2002;**33**:318–24.

12. Brasch MA, Hartley JL, Vidal M. ORFeome cloning and systems biology: standardized mass production of the parts from the parts-list. *Genome Res* 2004;**14**:2001–9.

13. Temple G, Lamesch P, Milstein S, *et al.* From genome to proteome: developing expression clone resources for the human genome. *Hum Mol Genet* 2006;**15**(Spec No 1):R31–43.

14. Hudson JR, Jr, Dawson EP, Rushing KL, *et al.* The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res* 1997;**7**:1169–73.

15. Parrish JR, Limjindaporn T, Hines JA, *et al.* High-throughput cloning of *Campylobacter jejuni* ORFs by in vivo recombination in Escherichia coli. *J Proteome Res* 2004;**3**:582–6.

16. Lamesch P, Li N, Milstein S, *et al.* hORFeome v3.1: a resource of human open reading frames representing over 10,000 human genes. *Genomics* 2007;**89**:307–15.

17. Uetz P, Giot L, Cagney G, *et al.* A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 2000;**403**:623–7.

18. Ito T, Chiba T, Ozawa R, *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc Natl Acad Sci USA* 2001;**98**:4569–74.

19. Fromont-Racine M, Rain JC, Legrain P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* 1997;**16**:277–82.

20. Rain JC, Selig L, De Reuse H, *et al.* The protein-protein interaction map of *Helicobacter pylori*. *Nature* 2001;**409**:211–5.

21. Parrish JR, Yu J, Liu G, *et al.* A proteome-wide protein interaction map for *Campylobacter jejuni*. *Genome Biol* 2007;**8**:R130.

22. Giot L, Bader JS, Brouwer C, *et al.* A protein interaction map of *Drosophila melanogaster*. *Science* 2003;**302**:1727–36.

23. Rual JF, Venkatesan K, Hao T, *et al.* Towards a proteome-scale map of the human protein-protein interaction network. *Nature* 2005.

24. Stelzl U, Worm U, Lalowski M, *et al.* A human protein-protein interaction network: a resource for annotating the proteome. *Cell* 2005;**122**:957–68.

25. Zhong J, Zhang H, Stanyon CA, *et al.* A strategy for constructing large protein interaction maps using the yeast two-hybrid system: regulated expression arrays and two-phase mating. *Genome Res* 2003;**13**:2691–9.

26. Bartel PL, Roecklein JA, SenGupta D, *et al.* A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet* 1996;**12**:72–7.

27. Hackbusch J, Richter K, Muller J, *et al.* A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension home-odomain proteins. *Proc Natl Acad Sci USA* 2005;**102**:4908–12.

28. Uetz P, Dong YA, Zeretzke C, *et al.* Herpesviral protein networks and their interaction with the human proteome. *Science* 2006;**311**:239–42.

29. LaCount DJ, Vignali M, Chettier R, *et al.* A protein interaction network of the malaria parasite *Plasmodium falciparum*. *Nature* 2005;**438**:103–7.

30. Li S, Armstrong CM, Bertin N, *et al.* A map of the interactome network of the metazoan *C. elegans*. *Science* 2004;**303**:540–3.

31. Stanyon CA, Liu G, Mangiola BA, *et al.* A *Drosophila* protein-interaction map centered on cell-cycle regulators. *Genome Biol* 2004;**5**:R96.

32. Formstecher E, Aresta S, Collura V, *et al.* Protein interaction mapping: a *Drosophila* case study. *Genome Res* 2005;**15**:376–84.

33. Colland F, Jacq X, Trouplin V, *et al.* Functional proteomics mapping of a human signaling pathway. *Genome Res* 2004;**14**:1324–32.

34. Goehler H, Lalowski M, Stelzl U, *et al.* A protein inter-action network links GIT1, an enhancer of huntingtin aggregation, to Huntington's disease. *Mol Cell* 2004;**15**:853–65.

35. Lim J, Hao T, Shaw C, *et al.* A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. *Cell* 2006;**125**:801–14.

36. Nakayama M, Kikuno R, Ohara O. Protein-protein interactions between large proteins: two-hybrid screening using a functionally classified library composed of long cDNAs. *Genome Res* 2002;**12**:1773–84.

37. Vollert CS, Uetz P. The phox homology (PX) domain protein interaction network in yeast. *Mol Cell Proteomics* 2004;**3**:1053–64.

38. Lehner B, Sanderson CM. A protein interaction framework for human mRNA degradation. *Genome Res* 2004;**14**: 1315–23.

39. Albers M, Kranz H, Kober I, et al. Automated yeast two-hybrid screening for nuclear receptor-interacting proteins. *Mol Cell Proteomics* 2005;**4**:205–13.

40. Calderwood MA, Venkatesan K, Xing L, et al. Epstein-Barr virus and virus human protein interaction maps. *Proc Natl Acad Sci USA* 2007;**104**:7606–11.

41. Cusick ME, Klitgord N, Vidal M, et al. Interactome: gateway into systems biology. *Hum Mol Genet* 2005; **14**(Spec No. 2):R171–81.

42. Uetz P, Hughes RE. Systematic and large-scale two-hybrid screens. *Curr Opin Microbiol* 2000;**3**:303–8.

43. Uetz P. Two-hybrid arrays. *Curr Opin Chem Biol* 2002;**6**: 57–62.

44. Sharan R, Suthram S, Kelley RM, et al. Conserved patterns of protein interaction in multiple species. *Proc Natl Acad Sci USA* 2005;**102**:1974–9.

45. Chen J, Hsu W, Lee ML, et al. Discovering reliable protein interactions from high-throughput experimental data using network topology. *Artif Intell Med* 2005;**35**: 37–47.

46. Chen J, Hsu W, Lee ML, et al. Increasing confidence of protein interactomes using network topological metrics. *Bioinformatics* 2006;**22**:1998–2004.

47. Saito R, Suzuki H, Hayashizaki Y. Construction of reliable protein-protein interaction networks with a new interaction generality measure. *Bioinformatics* 2003;**19**: 756–63.

48. Deane CM, Salwinski L, Xenarios I, et al. Protein interactions: two methods for assessment of the reliability of high throughput observations. *Mol Cell Proteomics* 2002;**1**: 349–56.

49. von Mering C, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 2002;**417**:399–403.

50. Deng M, Sun F, Chen T. Assessment of the reliability of protein-protein interactions and protein function prediction. *Pac Symp Biocomput* 2003:140–51.

51. Bader JS, Chaudhuri A, Rothberg JM, et al. Gaining confidence in high-throughput protein interaction networks. *Nat Biotechnol* 2004;**22**:78–85.

52. Suthram S, Shlomi T, Ruppin E, et al. A direct comparison of protein interaction confidence assignment schemes. *BMC Bioinformatics* 2006;**7**:360.

53. Reguly T, Breitkreutz A, Boucher L, et al. Comprehensive curation and analysis of global interaction networks in *Saccharomyces cerevisiae*. *J Biol* 2006;**5**:11.

54. Rajagopala SV, Titz B, Goll J, et al. The protein network of bacterial motility. *Mol Syst Biol* 2007;**3**:128.

55. Ramirez F, Schlicker A, Assenov Y, et al. Computational analysis of human protein interaction networks. *Proteomics* 2007;**7**:2541–52.

56. Hall DA, Ptacek J, Snyder M. Protein microarray technology. *Mech Ageing Dev* 2007;**128**:161–7.

57. Zhu H, Bilgin M, Bangham R, et al. Global analysis of protein activities using proteome chips. *Science* 2001;**293**: 2101–5.

58. Jones RB, Gordus A, Krall JA, et al. A quantitative protein interaction network for the ErbB receptors using protein microarrays. *Nature* 2006;**439**:168–74.

59. Stiffler MA, Chen JR, Grantcharova VP, et al. PDZ domain binding selectivity is optimized across the mouse proteome. *Science* 2007;**317**:364–9.

60. Ramachandran N, Hainsworth E, Bhullar B, et al. Self-assembling protein microarrays. *Science* 2004;**305**: 86–90.

61. Michnick SW, Ear PH, Manderson EN, et al. Universal strategies in research and drug discovery based on protein-fragment complementation assays. *Nat Rev Drug Discov* 2007;**6**:569–82.

62. Ozawa T, Kaihara A, Sato M, et al. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal Chem* 2001; **73**:2516–21.

63. Paulmurugan R, Umezawa Y, Gambhir SS. Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies. *Proc Natl Acad Sci USA* 2002;**99**:15608–13.

64. Johnsson N, Varshavsky A. Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci USA* 1994;**91**: 10340–4.

65. Lehming N. Analysis of protein-protein proximities using the split-ubiquitin system. *Brief Funct Genomic Proteomic* 2002;**1**:230–8.

66. Wehr MC, Laage R, Bolz U, et al. Monitoring regulated protein-protein interactions using split TEV. *Nat Methods* 2006;**3**:985–93.

67. Miller JP, Lo RS, Ben-Hur A, et al. Large-scale identification of yeast integral membrane protein interactions. *Proc Natl Acad Sci USA* 2005;**102**:12123–8.

68. Eyckerman S, Verhee A, der Heyden JV, et al. Design and application of a cytokine-receptor-based interaction trap. *Nat Cell Biol* 2001;**3**:1114–9.

69. Gavin AC, Bosche M, Krause R, et al. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 2002;**415**:141–7.

70. Krogan NJ, Cagney G, Yu H, et al. Global landscape of protein complexes in the yeast *Saccharomyces cerevisiae*. *Nature* 2006;**440**:637–43.

71. Bouwmeester T, Bauch A, Ruffner H, et al. A physical and functional map of the human TNF-alpha/ NF-kappa B signal transduction pathway. *Nat Cell Biol* 2004;**6**:97–105.

72. Barrios-Rodiles M, Brown KR, Ozdamar B, et al. High-throughput mapping of a dynamic signaling network in mammalian cells. *Science* 2005;**307**:1621–5.

73. Suzuki H, Fukunishi Y, Kagawa I, et al. Protein-protein interaction panel using mouse full-length cDNAs. *Genome Res* 2001;**11**:1758–65.

74. Suzuki H, Saito R, Kanamori M, et al. The mammalian protein-protein interaction database and its viewing system that is linked to the main FANTOM2 viewer. *Genome Res* 2003;**13**:1534–41.

75. Lehner B, Semple JI, Brown SE, *et al*. Analysis of a high-throughput yeast two-hybrid system and its use to predict the function of intracellular proteins encoded within the human MHC class III region. *Genomics* 2004;**83**: 153–67.

76. de Folter S, Immink RG, Kieffer M, *et al*. Comprehensive interaction map of the *Arabidopsis* MADS Box transcription factors. *Plant Cell* 2005;**17**:1424–33.

77. Lawit SJ, O'Grady K, Gurley WB, *et al*. Yeast two-hybrid map of *Arabidopsis* TFIID. *Plant Mol Biol* 2007;**64**:73–87.

78. Kaltenbach LS, Romero E, Becklin RR, *et al*. Huntingtin interacting proteins are genetic modifiers of neurodegeneration. *PLOS Genetics* 2007;e82.

79. Jin F, Avramova L, Huang J, Hazbun T. A yeast two-hybrid smart-pool-array system for protein-interaction mapping. *Nat Methods* 2007;**4**:405–7.