Two-hybrid arrays

Peter Uetz

The two-hybrid system is a genetic method for detecting protein–protein interactions. The assay can be applied to random libraries or arrays of colonies that express defined pairs of proteins. Arrays enable the testing of all possible protein pairs for interactions in a systematic fashion. The array format makes a large number of individual assays comparable and thus greatly simplifies the identification of false positives. Two-hybrid arrays have been used to study interactions among the proteins of yeast, hepatitis C virus, vaccinia virus, *Drosophila*, *Caenorhabditis elegans*, mouse and other species, and have already identified thousands of interactions.

Addresses
Institut für Toxikologie und Genetik, Forschungszentrum Karlsruhe, Postfach 3640, D-76021 Karlsruhe, Germany;
e-mail: peter.uetz@itg.fzk.de

Current Opinion in Chemical Biology 2001, **6**:57–62

1367-5931/01/$ – see front matter

Abbreviations

GGA	Golgi-localized, γ-ear-containing, ARF-binding protein
HCV	hepatitis C virus
MS	mass spectrometry
ORF	open reading frame
PCR	polymerase chain reaction

Introduction

As protein interactions are an essential part of all life, numerous methods have been developed to study them [1]. The two-hybrid assay has proved to be one of the most efficient techniques for finding new interactions [2–4]. The procedure is simple, inexpensive, and has the important advantage of being unbiased (i.e. no previous knowledge about the interacting proteins is necessary for a screen to be performed). However, the system also has a reputation for producing a significant number of false positives that require cumbersome analysis to separate the 'wheat' of true interactions from the 'chaff' of false positives.

The advent of complete genome sequences has dramatically changed two-hybrid searches for interacting proteins. Two-hybrid screens of random libraries can be performed much more rapidly when inserts from positive transformants can be identified by sequencing just a few base pairs to identify the encoded proteins. With complete genome sequences at our fingertips, two-hybrid screens can be carried out without any sequencing when known proteins are tested for interactions. Many such individual tests have been published in the literature, but complete genome sequences allow them to be carried out systematically with complete families or functional groups of proteins. However, genome-wide screens have been done only with the whole protein complement of yeast and a few viruses. This review focuses on such genome-wide screens and highlights, from the large number of small-scale array experiments, just a few examples to illustrate their range of applications.

From two-hybrid assays to arrays

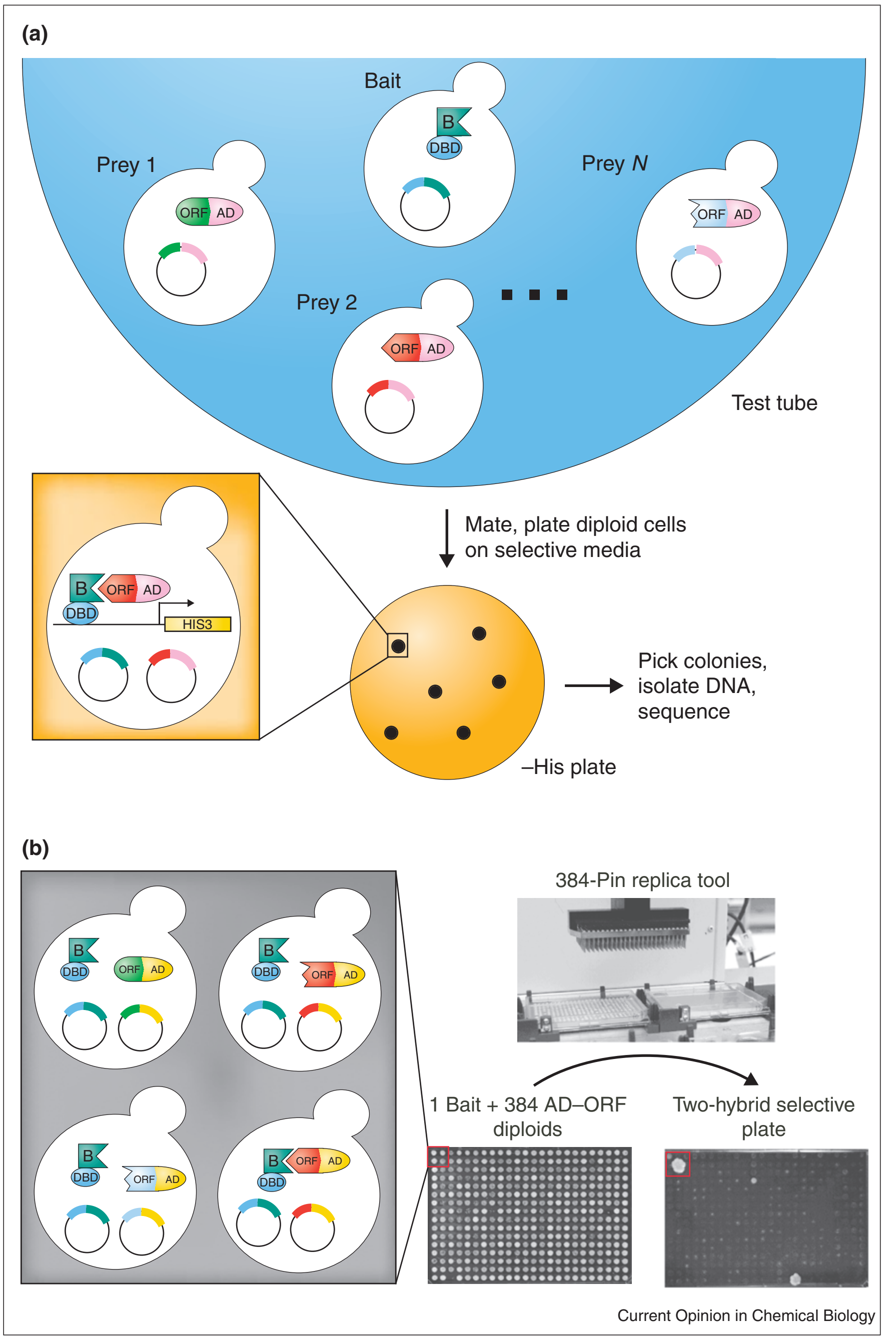
In the yeast two-hybrid system, interactions between two protein fusions are detected through protein–protein interaction-dependent reporter gene activation *in vivo* (Figure 1). This procedure is typically carried out by screening a protein of interest against a random library of potential protein partners via a genetic selection (Figure 1a). Plasmid DNA is recovered from cells expressing interacting proteins and gene identities are determined by DNA sequencing. However, two-hybrid screening can also be done in a colony array format, in which each colony expresses a defined pair of proteins [5,6]. Because the particular protein pair expressed in each colony is defined by its position in the array, positive signals identify interacting proteins without further characterization, thus obviating the need for DNA purification and sequencing. Many studies have used yeast colony 'mini-arrays' for the analysis of certain groups of proteins. Mini-arrays (as opposed to complete genome arrays) express defined sets of proteins that may be related by some common function or some other criterion. The interrogation of a two-hybrid colony array usually involves a mating strategy in which a DNA-binding domain hybrid (often termed the 'bait' protein) is tested against all activation domain hybrids (often termed 'prey' proteins) in a grid pattern. For the sake of simplicity, arrays usually use full-length open reading frames (ORFs) but can also use fragments when interaction domains are being mapped.

The main advantage of two-hybrid arrays is their systematic nature, which may cover all proteins from a whole genome. Complete sets of proteins not only put interactions in the context of other proteins but also allow one to compare the results from many assays. In fact, this allows the identification of two-hybrid false positives (see below). Finally, array screens can be automated to a high degree, which is a prerequisite for large-scale or genome-wide screens.

Applications of small-scale two-hybrid arrays

Small-scale two-hybrid arrays (also called mini-arrays) can be used to study a wide range of biological questions (Tables 1 and 2). Finley and Brent [7] reported one of the first small-scale array experiments to study interactions among the members of a protein family, namely cyclin-dependent kinases, cyclins, and related proteins (cyclin-dependent kinase interactors, Cdis) from *Drosophila* and other species. These authors discovered 19 interactions in just 45 individual tests but plan to extend the screen to the whole *Drosophila* proteome [8].

Figure 1



Current Opinion in Chemical Biology

Principle of two-hybrid library and array screens. **(a)** Typical two-hybrid screens use a library of random DNA or cDNA fused to a transcriptional activation domain (AD), expressed in yeast (‘preys’; circles denote plasmids). The library clones are mated to a strain of opposite mating type that expresses a protein of interest (‘bait’, B) as a fusion to a DNA-binding domain (DBD). If bait and prey interact in the resulting diploid cells, they reconstitute a transcription factor, which activates a reporter gene whose expression allows the diploid cell to grow on selective media (here, without histidine). As an alternative to mating, prey libraries can also be transformed into the bait strain in order to express bait and prey in the same cell. In any case, positive clones have to be picked, their DNA isolated and the encoded plasmids sequenced in order to identify interacting proteins. **(b)** Array screens use defined sets of cloned prey ORFs or fragments thereof that are mated systematically to a certain bait strain. Matings and two-hybrid tests can be automated when large sets of preys have to be assayed, as in the case of whole genomes.

A defined biological process has been studied by Walhout *et al.* [9] who studied vulval development in *C. elegans* by using a miniarray of 29 proteins. Thirteen interactions were discovered including two novel ones. A similar example is the study of cell polarity in yeast. Drees *et al.* [10] screened 68 yeast proteins with various functions in cell polarity, although they used a whole-genome yeast array. Overall, 191 protein interactions involving 110 proteins were detected, including 128 novel ones.

Two-hybrid tests cannot identify protein complexes, which are now routinely characterized by mass spectrometry (MS) [11]. However, two-hybrid assays can be used to map interactions *within* a complex, a task MS cannot easily address. Recently, the yeast proteasome was one of the first protein complexes studied by two-hybrid analysis [12]. In this study, 31 proteasome proteins were screened against a whole-genome yeast array. Altogether, 55 interactions were identified: 21 between components of the proteasome complex and 34 between proteasome proteins and other proteins. Some of the two-hybrid pairs are not direct neighbors in the crystal structure [13], suggesting that some interactions may be nonspecific or that bridging effects account for the two-hybrid results.

Surprisingly, there was no overlap between the non-proteasomal proteins found in the MS analysis and the two-hybrid screens. Further studies are obviously needed to account for this discrepancy.

Finally, two-hybrid arrays can also be used to map interaction domains. For instance, Puertollano *et al.* [14] used two-hybrid arrays expressing nested deletions to show that the cytosolic tail of human mannose 6-phosphate receptor bind to the so-called VHS domain of Golgi-localized, γ-ear-containing, ARF-binding protein (GGA).

Large-scale, but not genome-wide screens: mouse

Many companies and academic groups have started to work on human and mouse protein-interaction maps. However, few results have been published at this early stage. In a pilot study, Suzuki *et al.* [15••] tested 3500 mouse cDNAs for interactions using a new procedure (Figure 2). Altogether, about 12 million protein pairs were tested and among them 145 interactions were found.

Genome-wide two-hybrid arrays

Viral genomes: hepatitis C virus and vaccinia

Surprisingly few viral genomes have been studied systematically for protein interactions, although their small size makes them ideal targets for such screens. Flajolet *et al.* [16] studied interactions among proteins of the hepatitis C virus (HCV). The HCV genome encodes a single polyprotein that is processed into about 10 mature proteins. In this study, all mature virus proteins were tested pairwise against each other. Surprisingly, no interactions were found this way. The authors conclude that protein products from this 'full-length' pre-protein do not work in their two-hybrid assay, most likely because of folding problems. In order to circumvent these problems, they generated random libraries from the HCV genome in both bait and prey vectors. After screening 200 randomly chosen bait clones, five interactions were found, of which three had not been previously reported.

Table 1

Applications for two-hybrid arrays (see text for details).

Application	Example	Refs
Interactions within a protein family	*Drosophila* cell cycle proteins	[7]
Interactions in a process	*C. elegans* vulva development	[9]
Mapping interactions in a complex	Yeast, *C. elegans* proteasomes	[12,28]
Mapping interaction domains	Human M6PR-GGAs	[14]
Whole-genome interaction maps	Vaccinia, yeast	[17•,18•• 20••]

McCraith *et al.* [17•] expressed all 266 open reading frames (>65 amino acids) that are encoded by the 190 kb vaccinia virus genome as two-hybrid bait and prey proteins. Then they tested most of the ~70 000 pairs among them by two-hybrid assays, which returned 37 protein interactions.

Yeast

The first array-based two-hybrid screen of a whole proteome was published in early 2000 [18••]. This paper describes screens of 192 bait proteins against the 6000 yeast prey proteins, resulting in 281 distinct protein pairs.

Ito *et al.* [19,20••] described a similar strategy to search the yeast genome for protein interactions. First, they cloned all yeast ORFs in bait and prey vectors. However, instead of testing each bait clone against all prey clones, Ito *et al.* generated 62 pools with up to 96 baits each. Similarly, they combined up to 96 preys in each of 62 prey pools. These pools were mated to each other in all possible 3844 (62 × 62) combinations. The resulting diploid cells were streaked out on selective media to select for two-hybrid positives. Altogether, more than 15 000 positives were picked and subjected to colony polymerase chain reaction (PCR) and DNA sequencing, resulting in 13 754 sequence reads of baits and preys. From these sequences, 4549 protein interactions were deduced, although only 1533 were found three or more times. A subset of 841 protein pairs was found four or more times and defined as 'core' data set (i.e. as highly reliable).

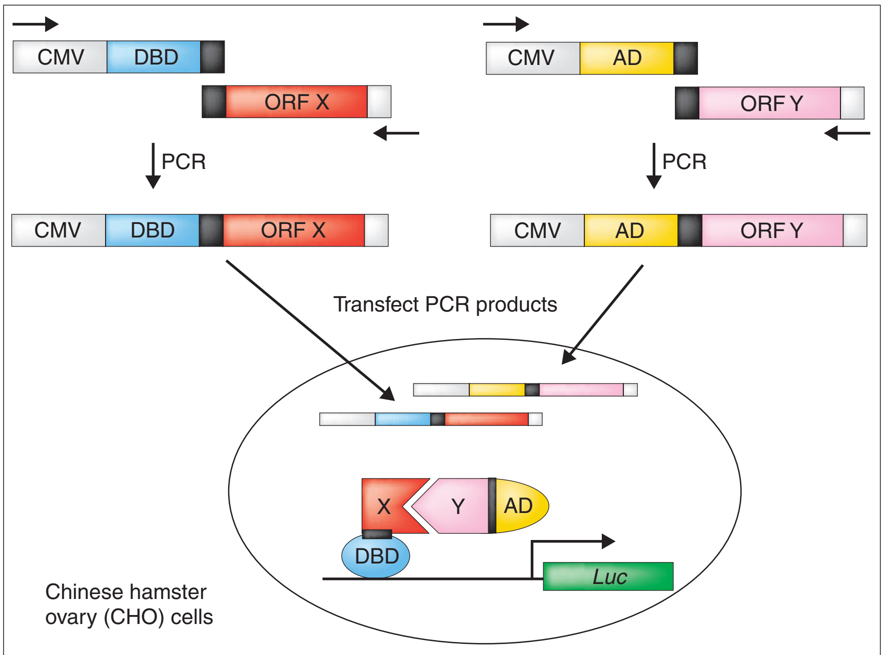
Table 2.

Two-hybrid array screens discussed in this paper.

Organism	Project	Proteins*	Assays*	Interactions*	Refs
Drosophila	Cell cycle proteins	13	45	19	[7]
C. elegans	Vulva development	29	841	8[†]	[9]
Mouse	Whole-genome pilot	~3500	$\sim 12\times10^6$	145	[15••]
HCV	Whole genome	10	~100	0/3[‡]	[16]
Vaccinia	Whole genome	266	~64 000	37[§]	[17•]
Yeast	One by one array	192	~1 150 000	281	[18••]
Yeast	Pool by pool	~6000	~36 000 000	4549/841[‡]	[19,20••]
Yeast	Cell polarity	68	~408 000	191[#]	[10]
Yeast	Proteasome	31	~186 000	55	[12]

*Number of proteins screened, individual two-hybrid assays, and interactions found. [†] Before this study, 11 interactions were already known among the 29 proteins and six of these were found again in the matrix experiment. In addition, two novel interactions were found. [‡] See text for details. [§] 25 of the 266 proteins turned out to be strong transcriptional activators and could not be tested as baits. The remaining 241 baits returned 37 interactions, of which five were detected in both directions (i.e. with both proteins expressed as bait and prey). 13 of the interactions represented homodimers. 28 of the 37 interactions were previously unknown. Eight interactions were among proteins of unknown function. [#] 128 interactions had not been described previously and 44 involved 20 proteins of unknown function.

Figure 2



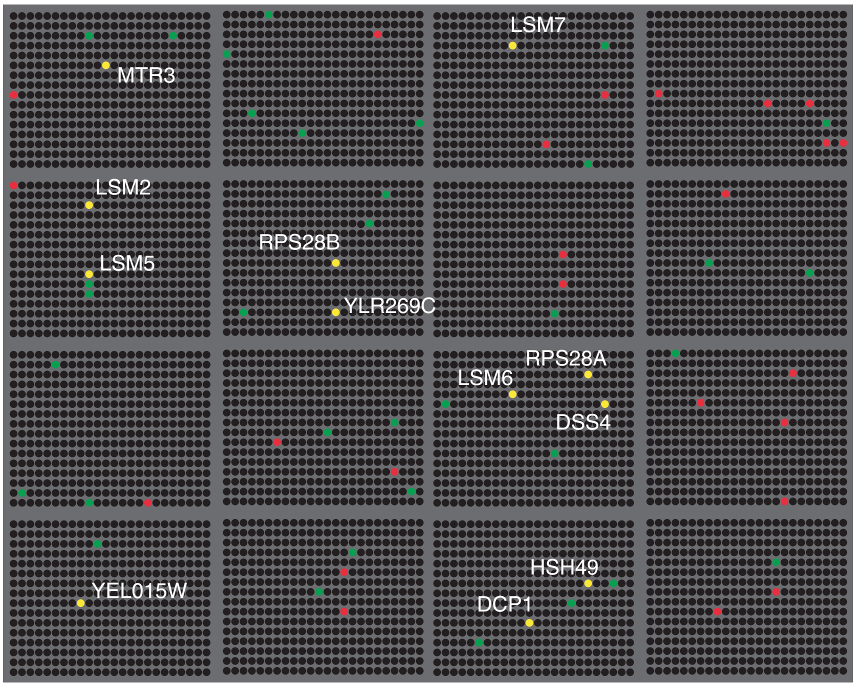
Current Opinion in Chemical Biology

Mammalian two-hybrid system used by Suzuki *et al.* [15••]. About 3500 mouse cDNAs were amplified by PCR and these PCR products (ORF X and ORF Y) mixed with another PCR product that carried a cytomegalovirus (CMV) promoter and either a Gal4 DNA-binding domain (DBD) or a VP16 transcriptional activation domain (AD). Because the two PCR products have overlapping sequences, they can be fused into one DNA fragment by a secondary PCR reaction using primers at the ends of the individual fragments. The final PCR fragments were transfected into tissue culture cells (CHO-K1) together with a reporter plasmid that carried a luciferase gene (*Luc*). When the encoded proteins interact, the luciferase reporter gene is transcribed and its activity can be measured as fluorescence. All 3500 × 3500 protein combinations were tested. To speed up the screening procedure, various numbers of baits and preys were co-transfected (i.e. pooled), and positive signals were later deconvoluted to identify interacting proteins.

Arrays and two-hybrid false positives

As with DNA microarrays, two-hybrid arrays allow a comparison of each individual assay with multiple identical assays. When the first array screens were done, it turned out that most positives are not reproducible when a screen is repeated [18••]. Although the molecular reasons for that are not really understood, simply repeating an array screen identifies those non-reproducible false positives

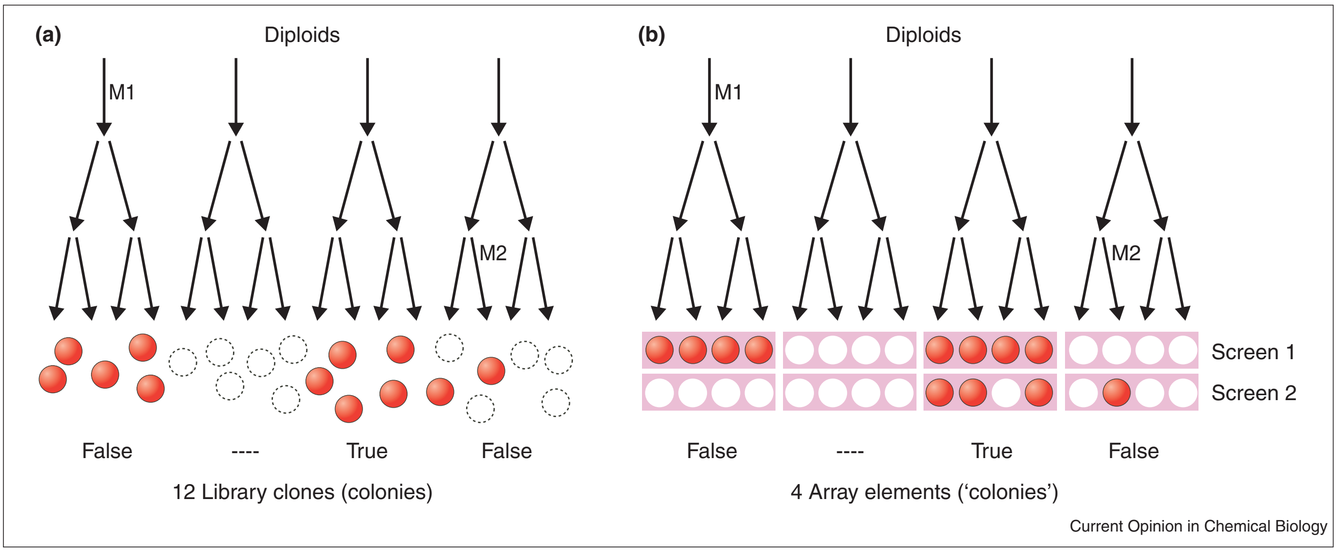
Figure 3



Current Opinion in Chemical Biology

Array screens efficiently eliminate false positives: whole yeast genome, full-length ORF screen using LSM8 as bait. LSM8 is a yeast protein involved in splicing. Positives from two independent screens are shown in red and green. Common positives in both screens are labelled in yellow with gene names attached. Note that the two screens generated 70 positives of which only 12 were reproducible. All non-reproducible positives are considered as false positives. Among the 12 positives, 9 are known to be involved in RNA processing or RNA binding. Another two, RPS28A and B, are ribosomal proteins. YLR269C and YEL015W are proteins of unknown function. DSS4 is the only positive whose association with LSM8 cannot be explained easily because it is a guanine-nucleotide exchange factor for Sec4. Among the 58 non-reproducible positives, only three are known to be involved in RNA processing or translation (i.e. RNA binding). Reproducible positives from Uetz *et al.* [18••], nonreproducible positives based on unpublished data.

Figure 4



Origin and identification of false positives in array screens. One source of false positives may be mutations (M) or other random events of unknown nature. When mutations happen early during the propagation of haploids or diploids (M1), more false positives may result than at later stages (M2). Dotted circles indicate diploids that don't grow after transfer to selective plates. **(a)** In a conventional two-hybrid screen, the identity of positives remains unclear, until their DNA has been sequenced. False positives cannot be identified easily. **(b)** In array screens, mutations are easily identified because it is unlikely that they occur twice in the same array element when a screen is done twice (Screen1 and Screen2 indicate two independent screens). However, 'true' positives should be reproducible in independent screens. In addition, they should be specific for certain baits (i.e. not be found with unrelated baits). Compare also with Figure 2.

(Figure 3). In addition, a number of preys can be found multiple times with unrelated baits. Such positives are bait-unspecific false positives by definition.

Most false positives in conventional library screens cannot be identified without additional experiments, especially when they are found only once in a screen (Figure 4). Two-hybrid arrays avoid such additional experimentation and require verification only for the small number of reproducible positives.

Conclusions

Despite its routine use in thousands of labs, the classical yeast two-hybrid system is certainly not perfect. False positives are still a major concern in conventional screens even though arrays and other modifications help to identify them. Another limitation is the possibility of bridging effects (i.e. endogenous proteins can act as bridging factors and therefore imply a direct interaction, although only an indirect interaction takes place). This affects all homologous systems including the novel system of Suzuki *et al.* who studied mouse proteins in hamster cells.

Other two-hybrid methods such as bacterial systems [21], SOS recruitment [22,23] or split ubiquitin [24] have hardly been used for large-scale screens and therefore cannot be compared in their performance. The same is true for protein chips, which are serious contenders for two-hybrid systems and may even replace them to some extent [25••,26]. Protein chips should be more reproducible than two-hybrid assays because they don't rely on complex biological processes and therefore have fewer variables. However, it remains to be seen if many purified proteins immobilised on a surface behave like proteins expressed in a living cell. A major advantage of two-hybrid arrays and protein chips alike, namely their defined ORF composition, is also a major disadvantage: it turned out that full-length ORFs often do not interact with their partners. The reasons are largely unknown, but often regulatory components such as allosteric modulators or signaling proteins may be missing or inactive in the context of protein fusions. Alternatively, the fusion proteins may not be expressed properly or not localized to the nucleus. When we understand the folding and dynamics and therefore the function of proteins better, we may design our screens in a way that avoids such problems.

Besides such technical issues, it is not clear yet whether screens of random libraries or some more directed approaches such as two-hybrid or peptide arrays are more cost-effective. Without doubt, there is still lots of room for improvements in any methodology, and even simple logistic steps can improve output considerably. With more completely sequenced genomes available, more two-hybrid arrays will certainly be built. However, protein or two-hybrid arrays of multicellular organisms make it necessary to select biologically meaningful subsets of proteins for screening. For example, only proteins expressed in the same tissue, cell or even subcellular compartment may be screened, excluding proteins from other tissues or compartments. In the study of Suzuki *et al.* [15••]

such a selection was not applied and resulted in only one interaction detected per 80 000 individual assays, as compared with one interaction per 4100 assays in a yeast array screen [18••]. Finley and Brent [7] even achieved a positive rate of one in two with their defined set of proteins, simply by careful selection of the proteins.

No matter how efficient, reliable, and cost-effective a method is, we need to keep in mind that two-hybrid arrays are nothing but another way to provide insight into the intricate network of processes in a cell [27]. No technique answers all our questions, and needs to be complemented by a whole slew of other methods to generate the complete picture of life at the molecular level we are all aspiring after.

Acknowledgements

I would like to thank Stan Fields and Michael Pankratz for helpful comments on the manuscript.

References and recommended reading

Papers of particular interest, published within the annual period of review, have been highlighted as:

- • of special interest
- •• of outstanding interest

1. Golemis EA (Ed): *Protein–Protein Interactions – A Molecular Cloning Manual*. Cold Spring Harbor: Cold Spring Harbor Laboratory Press; 2001.

2. Fields S, Song O: **A novel genetic system to detect protein–protein interactions.** *Nature* 1989, **340**:245-246.

3. Drees BL: **Progress and variations in two-hybrid and three-hybrid technologies.** *Curr Opin Chem Biol* 1999, **3**:64-70.

4. Frederickson RM: **Macromolecular matchmaking: advances in two-hybrid and related technologies.** *Curr Opin Biotechnol* 1998, **9**:90-96.

5. Cagney G, Uetz P, Fields S: **High-throughput screening for protein–protein interactions using two-hybrid assay.** *Methods Enzymol* 2000, **328**:3-14.

6. Cagney G, Uetz P: **High-throughput two-hybrid screening using yeast arrays.** *Curr Protocol Protein Sci* 2001, **19.6**:1-12.

7. Finley RL Jr, Brent R: **Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators.** *Proc Natl Acad Sci USA* 1994, **91**:12980-12984.

8. Stanyon CA, Finley RL Jr: **Progress and potential of *Drosophila* protein interaction maps.** *Pharmacogenomics* 2000, **1**:417-431.

9. Walhout AJ, Sordella R, Lu X, Hartley JL, Temple GF, Brasch MA, Thierry-Mieg N, Vidal M: **Protein interaction mapping in *C. elegans* using proteins involved in vulval development.** *Science* 2000, **287**:116-122.

10. Drees BL, Sundin B, Brazeau E, Caviston JP, Chen GC, Guo W, Kozminski KG, Lau MW, Moskow JJ, Tong A *et al.*: **A protein interaction map for cell polarity development.** *J Cell Biol* 2001, **154**:549-571.

11. Mann M, Hendrickson RC, Pandey A: **Analysis of proteins and proteomes by mass spectrometry.** *Annu Rev Biochem* 2001, **70**:437-473.

12. Cagney G, Uetz P, Fields S: **Two-hybrid analysis of the *Saccharomyces cerevisiae* 26S proteasome.** *Physiol Genomics* 2001, **7**:27-34.

13. Groll M, Ditzel L, Lowe J, Stock D, Bochtler M, Bartunik HD, Huber R: **Structure of 20S proteasome from yeast at 2.4 Å resolution.** *Nature* 1997, **386**:463-471.

14. Puertollano R, Aguilar RC, Gorshkova I, Crouch RJ, Bonifacino JS: **Sorting of mannose 6-phosphate receptors mediated by the GGAs.** *Science* 2001, **292**:1712-1716.

15. •• Suzuki H, Fukunishi Y, Kagawa I, Saito R, Oda H, Endo T, Kondo S, Bono H, Okazaki Y, Hayashizaki Y: **Protein–protein interaction panel using mouse full-length cDNAs.** *Genome Res* 2001, **11**:1758-1765.

The first large-scale two-hybrid screen on mammalian proteins. The authors used a novel mammalian two-hybrid system including a luciferase reporter gene for easy measurements. In addition, they completely avoided any cloning steps by simply transfecting PCR fragments.

16. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, Tiollais P, Transy C, Legrain P: **A genomic approach of the hepatitis C virus generates a protein interaction map.** *Gene* 2000, **242**:369-379.

17. • McCraith S, Holtzman T, Moss B, Fields S: **Genome-wide analysis of vaccinia virus protein–protein interactions.** *Proc Natl Acad Sci USA* 2000, **97**:4879-4884.

The first and largest array-based interaction map of a large virus encoding 266 proteins.

18. •• Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.

The first paper that used a whole-genome two-hybrid array. Also one of the first reports in which all yeast proteins were studied systematically on the protein level. This paper also indicated for the first time how arrays can be used to identify two-hybrid false positives.

19. Ito T, Tashiro K, Muta S, Ozawa R, Chiba T, Nishizawa M, Yamamoto K, Kuhara S, Sakaki Y: **Toward a protein–protein interaction map of the budding yeast: a comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins.** *Proc Natl Acad Sci USA* 2000, **97**:1143-1147.

20. •• Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: **A comprehensive two-hybrid analysis to explore the yeast protein interactome.** *Proc Natl Acad Sci USA* 2001, **98**:4569-4574.

The largest two-hybrid study published so far that attempted to test all protein–protein pairs. The paper includes about 4500 protein interactions of which 841 are considered as highly reliable.

21. Hu JC, Kornacker MG, Hochschild A: ***Escherichia coli* one- and two-hybrid systems for the analysis and identification of protein–protein interactions.** *Methods* 2000, **20**:80-94.

22. Aronheim A: **Membrane recruitment systems for analysis of protein–protein interactions.** *Methods Mol Biol* 2001, **177**:319-328.

23. Aronheim A: **Protein recruitment systems for the analysis of protein +/– protein interactions.** *Methods* 2001, **24**:29-34.

24. Stagljar I, Korostensky C, Johnsson N, te Heesen S: **A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*.** *Proc Natl Acad Sci USA* 1998, **95**:5187-5192.

25. •• Zhu H, Bilgin M, Bangham R, Hall D, Casamayor A, Bertone P, Lan N, Jansen R, Bidlingmaier S, Houfek T *et al.*: **Global analysis of protein activities using proteome chips.** *Science* 2001, **293**:2101-2105.

First whole-genome yeast protein chip. All yeast proteins were purified as GST fusion proteins and immobilized on a chip. The array was then probed with calmodulin and phospholipids to detect many new interacting proteins.

26. Zhu H, Snyder M: **Protein arrays and microarrays.** *Curr Opin Chem Biol* 2001, **5**:40-45.

27. Schwikowski B, Uetz P, Fields S: **A network of protein–protein interactions in yeast.** *Nat Biotechnol* 2000, **18**:1257-1261.

28. Davy A, Bello P, Thierry-Mieg N, Vaglio P, Hitti J, Doucette-Stamm L, Thierry-Mieg D, Reboul J, Boulton S, Walhout AJ *et al.*: **A protein–protein interaction map of the *Caenorhabditis elegans* 26S proteasome.** *EMBO Rep* 2001, **2**:821-828.