

Mapping protein–protein interactions with combinatorial biology methods

Joelle Pelletier* and Sachdev Sidhu†

Extremely diverse, DNA-encoded libraries of peptides and proteins have been constructed that include a linkage between each polypeptide and the encoding DNA. Library members can be selected by virtue of a particular binding specificity, and their protein sequence can be deduced from the sequence of the cognate DNA. Such combinatorial biology methods have proven invaluable in both identifying natural protein–protein interactions and also in mapping the specificities and energetics of these interactions in fine detail.

Addresses

*Université de Montréal, Département de Chimie, 2900 Edouard-Montpetit, Montréal, Québec H3C 3J7, Canada; e-mail: joelle.pelletier@umontreal.ca

†Department of Protein Engineering, Genentech, Inc., 1 DNA Way, South San Francisco, CA 94080, USA; e-mail: sidhu@gene.com

Current Opinion in Biotechnology 2001, 12:340–347

0958-1669/01/\$ – see front matter

© 2001 Elsevier Science Ltd. All rights reserved.

Abbreviations

BRET bioluminescence resonance energy transfer
DHFR dihydrofolate reductase
FACS fluorescence-activated cell sorting
FRET fluorescence resonance energy transfer
GFP green fluorescent protein
SH2/3 Src-homology 2/3 domain

Introduction

In February 2001, the sequence of the human genome was reported with over 90% coverage, revealing an estimated 31,000 proteins, most of which are of unknown function [1,2]. Elucidation of the genome sequence sets the stage for a still more daunting task: the assignment of biological function to the tens of thousands of newly discovered proteins. The identification of protein–protein interactions can help to determine the biological function of novel proteins, by situating them relative to other proteins in cellular pathways or functional classes (see [3•]).

An ideal general strategy for the detection of protein–protein interactions would be selective enough to reject non-specific interactions (or ‘false positives’) yet sensitive enough to ensure the detection of all biologically relevant interactions, thus avoiding ‘false negative’ results. This ideal strategy would be equally effective with all proteins and would allow for the rapid screening of very large combinatorial sequence spaces. Although such an idealized strategy does not exist, the efforts of numerous researchers have produced powerful strategies for mapping protein–protein interactions.

The first generation of strategies developed for the identification of protein–protein interactions involved the use of

proteins with no linkage to their encoding DNA. For example, protein-affinity chromatography, coimmunoprecipitation of complexed proteins, affinity blotting and chemical crosslinking are all physical methods for the *in vitro* identification of interactions (see the article by Xenarios and Eisenberg in this issue pp 334–339). These systems are generally simple to run and are routinely applied in the validation of interactions determined by library-based strategies (described below). However, being inherently low-throughput, they are not readily amenable to the combinatorial approaches that are needed to rapidly and accurately identify the interactions occurring in a complex organism.

In the past decade or so, the creative use of molecular biology has given rise to numerous combinatorial biology methods that link proteins to their encoding DNA. In this review, we focus on combinatorial biology methods that use DNA-encoded libraries to identify and analyze protein–protein interactions. We first describe the various methods, and then discuss how they have been applied to problems of interest: identifying natural protein–protein interactions and, subsequently, mapping the specificities and energetics of these interactions in fine detail [4•].

Combinatorial biology methods

Combinatorial biology depends on the ability to link polypeptides to their encoding DNA. The methods for generating DNA-encoded polypeptide libraries can be divided into two groups. *In vitro* methods use libraries in which polypeptides are accessible to exogenous ligands. These libraries can be used in direct *in vitro* binding selections with immobilized target ligands to enrich for polypeptides with particular binding specificities and affinities (Figure 1). In contrast, *in vivo* methods use polypeptide libraries that are expressed inside living cells along with a target protein of interest. An interaction between a particular library member and the target protein is detected by virtue of an effect on the host cell, such as a selective growth advantage, or changes to a physical property of the host cell (Figure 2). In these figures and throughout the text, ‘X’ and ‘Y’ identify the interacting proteins of interest, whereas ‘A’ and ‘B’ identify components of the detection system.

In vitro methods

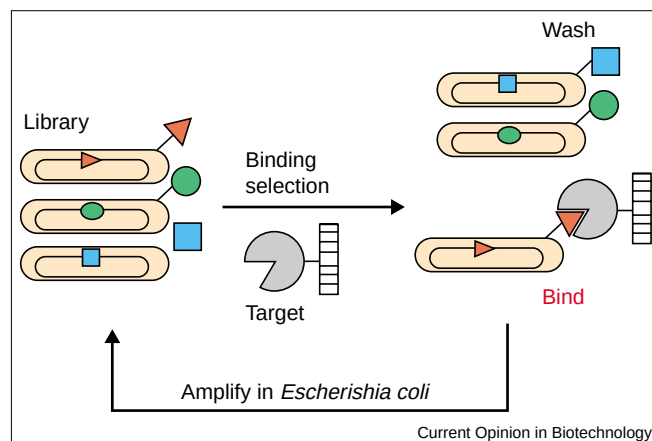
Phage display is the predominant method for *in vitro* combinatorial biology (reviewed in [5•]). The method stems from the observation that polypeptides fused to certain bacteriophage coat proteins are displayed on the surfaces of phage particles that also contain the cognate DNA. Libraries of ‘fusion phage’ can be rapidly sorted to obtain

clones with desired binding properties and phage can be readily amplified by passage through a bacterial host (Figure 1). Phage display was first demonstrated with the *Escherichia coli*-specific M13 bacteriophage [6] and this remains the most popular platform. However, several other *E. coli* phage have since been adapted for phage display [7,8] and eukaryotic systems have also been developed [9]. In addition, a related ‘peptides-on-plasmids’ system has been developed in which proteins are linked directly to the plasmid that encodes them [10]. With current optimized methods, M13 phage display libraries containing up to 10^{12} members can be constructed [5*].

Another approach to DNA-encoded libraries is provided by cell-surface display in which polypeptides are displayed directly on bacterial or yeast cells that also contain the encoding plasmid DNA (reviewed in [11] and in the article by Wittrup in this issue pp 395–399). Although binding selections with these systems involve the use of live cells, they can be considered *in vitro* methods because the polypeptides are on the outside of the cells where they are accessible for binding to exogenous ligands. Thus, these libraries can be used in selections analogous to those with phage display or, alternatively, they can be sorted using fluorescence-activated cell sorting (FACS).

In the methods described above, library diversity is limited by the fact that library construction involves a transformation step in which recombinant DNA is introduced into host cells. However, display methods based on cell-free translation have been developed to circumvent the limiting transformation step and thus enable library diversities in excess of 10^{13} members (see the article by

Figure 1

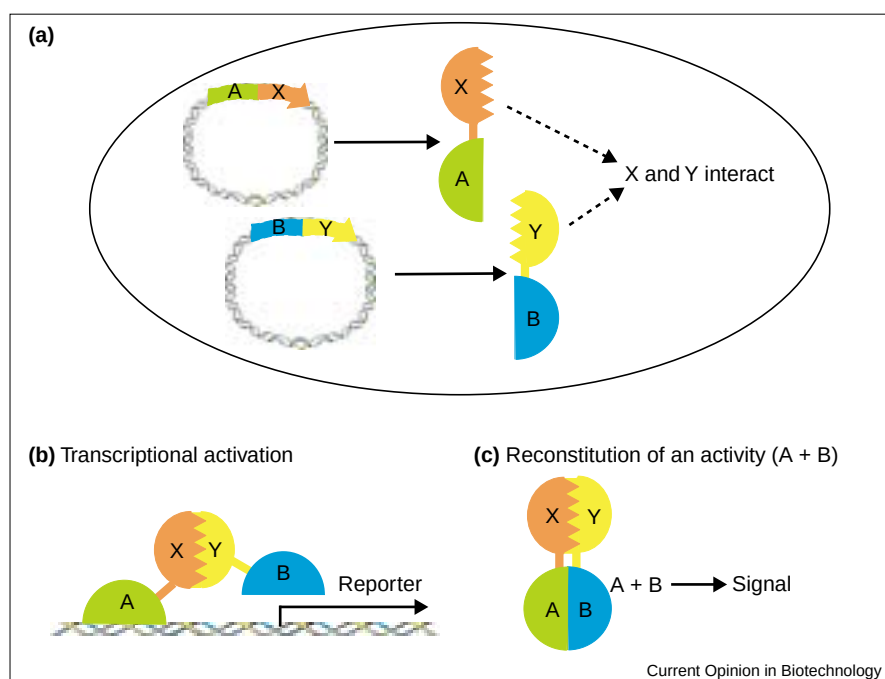


In vitro selection with phage display. Proteins are displayed on the surfaces of phage particles (orange ovals) that also encapsulate the encoding DNA. The phage-displayed library is exposed to an immobilized target; non-binding phage are removed by washing while bound phage are retained. Bound phage are then eluted and amplified by infection of an *Escherichia coli* host. Pools of amplified phage can be used in additional rounds of selection to further enrich for binding clones or, alternatively, the sequences of displayed proteins can be determined by sequencing the encapsulated DNA. Other *in vitro* display methods follow similar principals of binding selections followed by clonal amplification.

Amstutz *et al.* in this issue pp 400–405). In these systems, *in vitro* transcription and translation are used to produce protein–mRNA complexes in which polypeptide libraries are linked to their encoding mRNA, either non-covalently through a stalled ribosome [12] or covalently through a

Figure 2

In vivo selection systems. (a) DNA encoding the target (X, red) and bait (Y, yellow) proteins is fused to either part (A or B) of the reporter system. Co-expression in a cell yields the fusion proteins A–X and B–Y. (b) The yeast two-hybrid system. Interaction between X and Y results in functional pairing of the DNA-binding domain (A) and the transcriptional activator (B) allowing transcription of a reporter gene and leading to a detectable phenotype. (c) The protein fragment complementation assay. Interaction between X and Y results in reconstitution of an active enzyme (A–B), which gives rise to a detectable signal such as fluorescence or biosynthesis of an essential growth factor.



puromycin molecule [13]. These protein–mRNA complexes can be used in functional selections to enrich for binding proteins, and reverse transcription is used to convert the associated mRNA to DNA which can then be amplified with the polymerase chain reaction.

***In vivo* methods**

In vivo methods can be defined as those in which the protein–protein interaction under study is detected inside a living cell. The most widely used *in vivo* method for the detection of protein–protein interactions is the yeast two-hybrid system (Figure 2a,b). Interaction between the two proteins of interest (the ‘target’ and the ‘bait’) results in functional assembly of the two components of a signaling system: a DNA-binding domain and a transcriptional activation domain. As a result, transcriptional activation of a reporter gene occurs leading to a detectable phenotype, which can be screened for or directly selected. Although this describes the ‘standard’ yeast two-hybrid assay, it has given rise to a plethora of variations (reviewed most recently in [14]; see also [15,16]). These include the use of various DNA-binding domains and transcriptional activation domains or the requirement for a bridging molecule, such as RNA or a small ligand, to allow interaction (termed the three-hybrid assay). In addition, so-called ‘reverse’ two-hybrid strategies allow identification of components that disrupt protein–protein interactions (reviewed in [17]). Finally, two-hybrid systems in alternative hosts have been proposed to overcome the limitation of detecting interactions exclusively in the yeast nucleus [18*].

The above strategies rely on the transcriptional activation of a reporter gene, which can present certain disadvantages: first, the detected signal is not the direct result of the studied interaction, and second, any molecules that affect transcription compromise detection. An alternative is detection based on the complementation of enzyme fragments (Figure 2c). Reconstitution of the fragments to yield active enzyme is obligatorily directed by interaction between the target and the bait proteins; the enzyme activity is directly detected. The best-characterized fragment complementation assay is based on the reconstitution of the enzyme dihydrofolate reductase (DHFR). Interaction between the target and bait proteins can be selected for as a result of cellular propagation [19,20*] or detected by fluorescence of a bound inhibitor [20*]. The DHFR fragment complementation assay is not restricted to a particular cell type or compartment, allowing for detection of interactions in a variety of meaningful cellular contexts.

An advantage of fluorescent signaling is the potential for sorting cells using the high-throughput capacity of fluorescent-assisted cell sorting (FACS). As in the case of the DHFR fragment complementation assay, FACS can be applied to the following strategies. A recent twist to fragment complementation is the directed non-covalent reassembly of two fragments of green fluorescent protein (GFP), which allows the direct detection of protein–protein

interactions by fluorescence [21]. An interesting variant to the use of GFP fragments is their coupling to a self-excising ‘intein’ protein [22]. This results in the potentially advantageous covalent reassembly of GFP. Here, two intein fragments were genetically fused to the GFP fragments, and further fused to the proteins of interest. Protein interaction allows reconstitution not of GFP, initially, but rather of the intein. Reconstitution of the intein results in its self-excision and the concomitant covalent ligation of the two GFP fragments, yielding full-length, active GFP. Another variant of fragment complementation relies on weakening the classical, spontaneous complementation of β -galactosidase fragments, where interaction between fused proteins of interest enhances complementation [23]. A wide range of detection protocols exist for the detection of β -galactosidase activity, including those yielding luminescent and fluorescent signals [24]. The fact that the fragments can weakly complement each other in the absence of interacting partners suggests that β -galactosidase complementation may be more appropriately used with pairs of well-behaved partners than in library screening, as background signal is likely to interfere with the detection of weak interactions.

In the recently developed protein recruitment systems (reviewed in [25,26]), the first protein of interest (X) is genetically fused to a membrane localization signal (A). The second protein of interest (Y) is fused either to Sos or to Ras, the membrane localization of both these proteins is a requirement for activation. Activation leads to a selectable trait such as temperature sensitive growth (Sos) or signal transduction leading to transcription of a reporter such as the bioluminescent luciferase (Ras). Recruitment strategies have been demonstrated in yeast and in mammalian cells. The Sos recruitment system is commercialized as Cytotrap™ (Stratagene); its initial short-comings [25] are being addressed by continuing improvements [27]. These systems have been used to screen a library against a single protein of interest (e.g. [28]), but have not been applied to combinatorial screening.

Detection strategies for use in bacteria can offer many advantages relative to yeast and mammalian cells, such as ease of manipulation, the speed of assay and low cost. These recently reviewed approaches [29] include a strategy based on the complementation of two fragments of adenylate cyclase, which allows cyclic AMP synthesis and detection of the result of a signaling cascade [30]. Strategies have also been reported based on the interruption of the dimerization of transcriptional activators and repressors; in these systems dimerization is re-established by interaction of the fused proteins of interest [31,32].

Alternative screening strategies involve the use of light energy transfer, in which the approach of an energy donor and its acceptor result in a detectable pattern of light emission. In both fluorescence and bioluminescence resonance energy transfer (FRET and BRET, respectively) the proteins of interest (X and Y) are genetically fused either

to two fluorescent proteins A and B, in the case of FRET, or to one bioluminescent protein (A) and one fluorescent protein (B), in the case of BRET. For example, a bioluminescent luciferase and GFP may be used. It is necessary for the emission spectrum of A to overlap with the excitation spectrum of B. Thus, the interaction between X and Y, in the cell, allows for proximity-dependant energy transfer between the energy donor A and its acceptor B. This event is recorded either as a decrease in the fluorescence emission of A or as an increase in that of B. FRET and BRET have been successfully applied to the difficult task of detecting the oligomerization of integral membrane proteins *in situ*, among other applications [33].

Mapping protein–protein interactions

Combinatorial biology methods have proven invaluable both in defining natural protein–protein interactions and in rapidly mapping the specificities and energetics of these interactions in fine detail.

Identifying naturally occurring protein–protein interactions

The past year has been marked by many reports of studies to build comprehensive maps of protein–protein interactions within model organisms such as yeast, bacteria and viruses (reviewed in [34]). These impressive achievements, for the most part, used matrix-based approaches in which large sets of predefined polypeptides were screened against each other. This approach is distinct from true library screening where a complex DNA library is screened and the relevant proteins are identified after a hit is obtained (as clearly explained in [18•]). In parallel, progress has also been made in mapping protein–protein interactions in *Drosophila* (recently reviewed in [35]), but not at a high-throughput level. The close genetic linkage between the human and *Drosophila* proteomes indicates that the maps built for this simpler organism will speed mapping of the human proteome.

In an impressive body of work, Uetz *et al.* [36••] reported a high-throughput yeast two-hybrid screen of most yeast proteins (approximately 6000) against each other. These researchers cloned 87% of yeast genes as fusions, first to a DNA-binding domain (see Figure 2a) and then to a transcriptional activation domain. Yeast transformants carrying the first fusion were arranged in an array format and each was mated against a pool of transformants carrying the latter fusion. Using this strategy, 8% of yeast gene products were mapped in interactions. In addition, a smaller and more labor-intensive array screen was performed in which single fusions (rather than a library) were tested against the array. In this case, almost 50% of the gene products tested yielded at least one interaction, but only 3% of the yeast genome was tested in this manner.

An independent, large-scale project by Ito *et al.* [37,38••] has recently reported more complete coverage of the yeast interaction map. The researchers cloned approximately 95% of all yeast genes into both fusion libraries required

for two-hybrid screening. This study yielded approximately 4500 possible interactions among 3300 proteins, or roughly 50% of the gene products tested with almost full coverage of the genome. However, after eliminating the likely false positives, the authors suggest that 400 of these interactions may be biologically significant.

Interestingly, neither of these studies identified more than 15% of previously published interactions [37], suggesting that the map of protein–protein interactions in yeast may be more complex than expected. Further improvements in two-hybrid screening, including computationally assisted validation, will continue to improve the frequency of biologically significant hits and enable more complete mapping [34].

A combinatorial yeast two-hybrid screen of the 266 vaccinia virus proteins was reported: the required fusions with both a DNA-binding domain and an activation domain were obtained, and all possible pair-wise matings were performed in an array format [39]. Out of the 70,000 combinations, 37 interactions were identified, including 28 putative novel interactions. Although this work yielded new and informative material leading to a better understanding of viral functions, there remains much to be done in order to obtain the complete vaccinia virus interaction map because of a high level of false-negative results.

Comprehensive mapping of the interactions that occur among hepatitis C virus proteins is an apparently simple task: the single polyprotein encoded by the virus is processed into approximately 10 proteins only. However, classical yeast two-hybrid experiments failed to reveal interactions. Flajolet *et al.* [40] took advantage of the small size of the virus genome to screen random genomic fragments. They improved the two-hybrid selection protocol by preselecting only in-frame fragments that show no autoactivation and in this way were able to identify several interactions, both known and novel.

These improvements paved the way for the first protein–protein interaction map reported for a prokaryote, that of the bacterium responsible for some forms of stomach ulcers, *Helicobacter pylori* [41••]. The researchers modified the yeast two-hybrid strategy to screen 261 *H. pylori* bait proteins against more than 10⁶ target (or prey) protein fragments. Focusing on protein fragments, rather than on entire proteins, increased the number of combinations that needed to be verified. However, the strategy is advantageous as many proteins are not suitable for two-hybrid screens — they may, for example, give rise to false-positive signals or be insoluble — whereas their constituent fragments may be well suited to the task. Indeed, the researchers connected almost 50% of the organism's proteins through their interactions. Although very labor-intensive, the strategy combined many improvements leading to a lower number of false-positives and false-negatives and should be generally applicable to any genome of interest.

In addition to whole-proteome screening, yeast two-hybrid approaches have also been recently applied to the identification of interactions occurring in highly complex subsets of proteins, such as signal transduction pathways, RNA splicing proteins, cytoskeletal proteins [42] and proteins involved in cellular trafficking pathways [43]. A complete study of the interactions involving 27 proteins of vulval development in the nematode *Caenorhabditis elegans* is noteworthy [44*]. The authors screened a matrix of 27 proteins expressed as DNA-binding domain fusions against a *C. elegans* genomic library of activation domain fusions, representing approximately 20,000 gene products. The authors identified 148 interactions among 124 proteins, including 50% of all previously reported interactions, indicating a relatively low level of false-negatives.

The DHFR protein fragment complementation assay holds much promise in complementing the two-hybrid strategies and becoming a generally applicable strategy for mapping interactions in a variety of relevant cellular contexts. In a demonstration of its use in mammalian cells, it enabled the determination of the mechanism of induced dimerization of the transmembrane erythropoietin receptor as well as the *in vivo* quantitation of interactions [20*,45]. The combinatorial power of the strategy (as illustrated below) suggests that applications to large-scale proteome screening are forthcoming.

Resonance energy transfer methods (FRET and BRET) are being increasingly applied to the detection of protein–protein interactions [46]. Recent examples include the mapping of interactions between two transport receptors and proteins of the nuclear pore complex [47] and applications to the detection of oligomerization of G-protein-coupled transmitter receptors [33]. These approaches can also allow for the determination of the subcellular locus of interaction, by fluorescence microscopy [48]. FRET and BRET are commonly applied to small panels of interacting proteins, but should be amenable to high-throughput detection of protein–protein interactions [49].

Binding partners for orphan receptors and ligands have also been identified using cDNA libraries displayed on either M13 phage [50–53] or other bacteriophage [54,55]. Phage-displayed cDNA libraries share the limitation that is common to most strategies presented here, namely that not all eukaryotic proteins can be displayed efficiently with any particular display system. It is possible that different display systems will prove useful for the display of different subsets of proteins. However, certain proteins types, such as integral membrane proteins, are likely to be unsuitable to any phage-display method.

Mapping binding specificities and energetics

Protein–protein interactions can be roughly divided into two basic types: those that involve binding contacts spread over large, non-contiguous surfaces and those that involve the recognition of a continuous, linear stretch of amino

acids. Combinatorial approaches are particularly useful for mapping the latter type of interaction because it is quite easy to generate large libraries of naïve peptides that effectively mimic linear epitopes. Indeed, phage display was first used to identify linear binding epitopes for monoclonal antibodies and this remains a major use for the technology [56]. By contrast, the analysis of large interfaces is more challenging because libraries must be constructed in the context of large, folded proteins that may not be readily amenable to particular combinatorial biology techniques. However, effective methods have been developed for dealing with an impressively diverse range of protein–protein interactions.

Many intracellular functions are controlled and mediated by protein domains that bind continuous, linear epitopes in other proteins (reviewed in [57]). There are several distinct families of these domains that have specific requirements for the types of ligand they recognize, but combinatorial biology methods have been developed to analyze most of the major interaction types. SH3 and WW domains, which bind to proline-rich sequences, have been studied with phage-display peptide libraries [58,59]. Similarly, *in vitro* phosphorylated phage libraries have been developed to study the binding specificities of SH2 and PTB domains, which recognize phosphotyrosine-containing peptides [60,61]. PDZ domains recognize the extreme C termini of proteins and have been studied using peptide libraries displayed as C-terminal fusions in either a peptides-on-plasmids system [62] or a specialized phage-display system [63]. All these studies produced small, linear peptides that bound the target proteins with high specificity and affinity. These peptides could be used as potent inhibitors of natural protein–protein interactions. These sequences could also be used to identify natural binding partners and to assign biological functions to protein binding domains.

To study interactions that involve noncontiguous binding epitopes, libraries are constructed by introducing mutations into protein domains, either randomly or in a site-directed manner. Following binding selection and sequence analysis of selected clones, the contribution of each protein residue to binding can be assessed on the basis of sequence conservation, because the degree of conservation should be proportional to functional importance. Using a special phage-displayed library in which residues were preferentially allowed to vary only as the wild-type or alanine, it was demonstrated that combinatorial libraries can be used to calculate binding energetics for individual sidechains with accuracies approaching those obtained with conventional point mutagenesis [64**]. Protein–protein interactions have also been analyzed *in vivo* using a fragment complementation assay, which provided the added advantage of simultaneously varying both sides of an interface in a library-against-library approach; this enabled the simultaneous analysis and optimization of both members of a leucine zipper heterodimer [65**,66]. In principle, any combinatorial biology method could be adapted to study binding

epitopes and, indeed, promising results have also been obtained with two-hybrid systems [67,68].

Conclusions

To date, yeast two-hybrid systems have provided the predominant *in vivo* method to study protein–protein interactions, whereas the predominant *in vitro* method has been M13 phage display, at least partly because these were among the earliest methods developed. However, numerous distinct *in vivo* and *in vitro* methods to study interactions have been developed over the past decade, many of which are suitable for automation and high-throughput screening. These methods are likely to be complementary, with each being suitable to a particular subset of proteins and problems. Library sizes and construction methods are constantly improving, thus increasing the diversity that can be accessed and also improving the library quality; parallel improvements in instrumentation and robotics help to increase screening efficiency. Validation is the current bottleneck to positive identification of interacting pairs; computational developments and database construction (as reviewed by Xenarios and Eisenberg in this issue pp 334–339) help both in validation and analysis of the complex interaction networks that are being drawn. As a further extension, the above strategies can also be exploited in identifying ligands that are required for given interactions to occur or which disrupt interactions. The ultimate goal, to draw full interaction maps illustrating the temporal and spatial relationships between all proteins, is not far in the future.

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. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, FitzHugh W *et al.*: Initial sequencing and analysis of the human genome. *Nature* 2001, 409:860-921.
2. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA *et al.*: The sequence of the human genome. *Science* 2001, 291:1304-1351.
3. Walhout A, Vidal M: Protein interaction maps for model organisms.
 - *Nat Rev Mol Cell Biol* 2001, 2:55-62.
 This clearly written review of the strategies for building maps of protein–protein interactions defines the language and the considerations of the field of proteomics. A comprehensive and recent article.
4. Thorner J, Emr SD, Abelson JN (Eds): *Applications of Chimeric Genes and Hybrid Proteins, Part C: Protein–Protein Interactions and Genomics*, vol 328. Academic Press; 2000.
5. Sidhu SS, Lowman HB, Cunningham BC, Wells JA: Phage display
 - for selection of novel binding peptides. *Methods Enzymol* 2000, 328:333-363.
 The authors provide optimized protocols for the construction of phage-displayed libraries. By electroporating highly concentrated *E. coli* with high concentrations of ultrapure DNA, libraries containing in excess of 10^{10} members can be constructed with a single electroporation reaction. Libraries with diversities greater than 10^{12} can be constructed by combining multiple electroporation reactions.
6. Smith GP: Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* 1985, 228:1315-1317.
7. Santini C, Brennan D, Mennuni C, Hoess RH, Nicosia A, Cortese R, Luzzago A: Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage λ . *J Mol Biol* 1998, 282:125-135.
8. Ren Z, Black LW: Phage T4 SOC and HOC display of biologically active, full-length proteins on the viral capsid. *Gene* 1998, 215:439-444.
9. Possee RD: Baculoviruses as expression vectors. *Curr Opin Biotechnol* 1997, 8:569-572.
10. Cull MG, Miller JF, Schatz PJ: Screening for receptor ligands using large libraries of peptides linked to the C terminus of the Lac repressor. *Proc Natl Acad Sci USA* 1992, 89:1865-1869.
11. Georgiou G, Stathopoulos C, Daugherty PS, Nayak AR, Iverson BL, Curtiss R III: Display of heterologous proteins on the surface of microorganisms: from the screening of combinatorial libraries to live recombinant vaccines. *Nat Biotechnol* 1997, 15:29-34.
12. Hanes J, Plückthun A: *In vitro* selection and evolution of functional proteins by using ribosome display. *Proc Natl Acad Sci USA* 1997, 94:4937-4942.
13. Roberts RW, Szostak JW: RNA–peptide fusions for the *in vitro* selection of peptides and proteins. *Proc Natl Acad Sci USA* 1997, 94:12297-12302.
14. Fashena SJ, Serebriiskii I, Golemis EA: The continued evolution of two-hybrid screening approaches in yeast: how to outwit different preys with different baits. *Gene* 2000, 250:1-14.
15. Vidal M, Legrain P: Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* 1999, 27:919-929.
16. Drees BL: Progress and variations in two-hybrid and three-hybrid technologies. *Curr Opin Chem Biol* 1999, 3:64-70.
17. Vidal M, Endoh H: Prospects for drug screening using the reverse two-hybrid system. *Trends Biotechnol* 1999, 17:374-381.
18. Legrain P, Selig L: Genome-wide protein interaction maps using
 - two-hybrid systems. *FEBS Lett* 2000, 480:32-36.
 A minireview that thoroughly discusses the distinctions between matrix-based approaches and library screening. Although the paper focuses mainly on the yeast two-hybrid system, it also describes alternative strategies.
19. Pelletier JN, Campbell-Valois FX, Michnick SW: Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci USA* 1998, 95:12141-12146.
20. Remy I, Michnick SW: Clonal selection and *in vivo* quantitation of
 - protein interactions with protein-fragment complementation assays. *Proc Natl Acad Sci USA* 1999, 96:5394-5399.
 Although this study was not of combinatorial interaction screening, it clearly illustrates that protein fragment complementation assays have the potential for high-throughput combinatorial interaction mapping. Advantages of the strategy include detection of protein–protein interactions in a variety of meaningful cellular contexts and the ability to quantitate interactions.
21. Ghosh I, Hamilton A, Regan L: Antiparallel leucine zipper-directed protein reassembly: application to the green fluorescent protein. *J Am Chem Soc* 2000, 122:5658-5659.
22. Ozawa T, Nogami S, Sato M, Ohya Y, Umezawa Y: A fluorescent indicator for detecting protein–protein interactions *in vivo* based on protein splicing. *Anal Chem* 2000, 72:5151-5157.
23. Rossi F, Charlton CA, Blau HM: Monitoring protein–protein interactions in intact eukaryotic cells by β -galactosidase complementation. *Proc Natl Acad Sci USA* 1997, 94:8405-8410.
24. Blakely BT, Rossi FM, Tillotson B, Palmer M, Estelles A, Blau HM: Epidermal growth factor receptor dimerization monitored in live cells. *Nat Biotechnol* 2000, 18:218-222.
25. Aronheim A: Protein recruitment systems for the analysis of protein–protein interactions. *Biochem Pharmacol* 2000, 60:1009-1013.
26. Aronheim A: Protein recruitment systems for the analysis of protein–protein interactions. *Methods* 2001, 24:29-34.
27. Huang W, Wang SL, Lozano G, de Crombrughe B: cDNA library screening using the SOS recruitment system. *Biotechniques* 2001, 30:94-100.
28. Takemaru KI, Moon RT: The transcriptional coactivator CBP interacts with β -catenin to activate gene expression. *J Cell Biol* 2000, 149:249-254.

29. Ladant D, Karimova G: Genetic systems for analyzing protein–protein interactions in bacteria. *Res Microbiol* 2000, 151:711-720.
30. Karimova G, Pidoux J, Ullmann A, Ladant D: A bacterial two-hybrid system based on a reconstituted signal transduction pathway. *Proc Natl Acad Sci USA* 1998, 95:5752-5756.
31. 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.
32. Joung JK, Ramm EI, Pabo CO: A bacterial two-hybrid selection system for studying protein–DNA and protein–protein interactions. *Proc Natl Acad Sci USA* 2000, 97:7382-7387.
33. Bouvier M: Oligomerization of G-protein-coupled transmitter receptors. *Nat Rev Neurosci* 2001, 2:274-286.
34. Uetz P, Hughes RE: Systematic and large-scale two-hybrid screens. *Curr Opin Microbiol* 2000, 3:303-308.
35. Stanyon CA, Finley RL Jr: Progress and potential of *Drosophila* protein interaction maps. *Pharmacogenomics* 2000, 1:417-431.
36. 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 culmination of a large body of work, this article details two complementary approaches to matrix-based determination of protein–protein interactions. The first strategy is medium-throughput, with a very high success rate (approximately 50% of preselected proteins yielded at least one interaction), but only 3% of the genome was tested this way as the strategy is very labor-intensive. The second strategy is high-throughput, allowing essentially all 6000 yeast proteins to be combinatorially tested, with a comparatively modest success rate (approximately 8% of all proteins yielded at least one interaction).
37. Hazbun TR, Fields S: Networking proteins in yeast. *Proc Natl Acad Sci USA* 2001, 98:4277-4278.
38. 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 effort in mapping protein–protein interactions reported to date. The extent of false-positive and false-negative results is discussed and the putative interactions detected are critically evaluated.
39. 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.
40. Flajole 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.
41. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, Simon S, Lenzen G, Petel F, Wojcik J, Schachter V *et al.*: The protein–protein interaction map of *Helicobacter pylori*. *Nature* 2001, 409:211-215.
- This article describes the first protein–protein interaction map for a prokaryote. The authors present improvements to the yeast two-hybrid strategy in order to achieve a versatile high-throughput screening strategy. The main novel feature is the use of a modest library of full-length proteins screened against very large libraries of protein fragments. The authors clearly and succinctly describe the steps required in building complex libraries of 'targets' and 'baits', determination of the selective pressure required for each individual bait and processing of the raw results.
42. Meng JJ, Ip W: A yeast two-hybrid approach for probing cytoskeletal protein interactions. *Methods Mol Biol* 2001, 161:255-268.
43. Stephens DJ, Banting G: The use of yeast two-hybrid screens in studies of protein–protein interactions involved in trafficking. *Traffic* 2000, 1:763-768.
44. 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.
- The authors developed a thorough approach to mapping interactions in a discrete – yet still large – area of the *C. elegans* proteome. Technical improvements are described for the experimental protocol and for analysis of the putative detected interactions. The high success rate in identifying both previously known and novel interactions is encouraging with respect to undertaking even broader screens.
45. Remy I, Wilson IA, Michnick SW: Erythropoietin receptor activation by a ligand-induced conformation change. *Science* 1999, 283:990-993.
46. de Angelis DA: Why FRET over genomics? *Physiological Genomics* 1999, 1:93-99.
47. Damelin M, Silver PA: Mapping interactions between nuclear transport factors in living cells reveals pathways through the nuclear pore complex. *Mol Cell* 2000, 5:133-140.
48. Li HY, Ng EK, Lee SM, Kotaka M, Tsui SK, Lee CY, Fung KP, Wayne MM: Protein–protein interaction of FHL3 with FHL2 and visualization of their interaction by green fluorescent proteins (GFP) two-fusion fluorescence resonance energy transfer (FRET). *J Cell Biochem* 2001, 80:293-303.
49. Olsen MJ, Stephens D, Griffiths D, Daugherty P, Georgiou G, Iverson BL: Function-based isolation of novel enzymes from a large library. *Nat Biotechnol* 2000, 18:1071-1074.
50. Fransen M, Van Veldhoven PP, Subramani S: Identification of peroxisomal proteins by using M13 phage protein VI phage display: molecular evidence that mammalian peroxisomes contain a 2,4-dienoyl-CoA reductase. *Biochem J* 1999, 340:561-568.
51. Hufton SE, Moerkert PT, Meulemans EV, de Bruine A, Arends JW, Hoogenboom HR: Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *J Immunol Methods* 1999, 231:39-51.
52. Shanmugavelu M, Baytan A, Chesnut J, Bonning B: A novel protein that binds juvenile hormone esterase in fat body tissue and pericardial cells of the tobacco hornworm *Manduca sexta* L. *J Biol Chem* 2000, 275:1802-1806.
53. Cochrane D, Webster C, Masih G, McCafferty J: Identification of natural ligands for SH2 domains from a phage display cDNA library. *J Mol Biol* 2000, 297:89-97.
54. Sche PP, McKenzie KM, White JD, Austin DJ: Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem Biol* 1999, 6:707-716.
55. Fadok VA, Bratton DL, Rose DM, Pearson A, Ezekewitz RA, Henson PM: A receptor for phosphatidylserine-specific clearance of apoptotic cells. *Nature* 2000, 405:85-90.
56. Scott JK, Smith GP: Searching for peptide ligands with an epitope library. *Science* 1990, 249:386-390.
57. Pawson T, Scott JD: Signaling through scaffold, anchoring, and adaptor proteins. *Science* 1997, 278:2075-2080.
58. Linn H, Ermekova KS, Rentschler S, Sparks AB, Kay BK, Sudol M: Using molecular repertoires to identify high-affinity peptide ligands of the WW domain of human and mouse YAP. *Biol Chem* 1997, 378:531-537.
59. Sparks AB, Rider JE, Hoffman NG, Fowlkes DM, Quillam LA, Kay BK: Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLC γ , Crk, and Grb2. *Proc Natl Acad Sci USA* 1996, 93:1540-1544.
60. Dente L, Vetriani C, Zucconi A, Pelicci G, Lanfranccone L, Pelicci PG, Cesareni G: Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine-containing peptides. *J Mol Biol* 1997, 269:694-703.
61. Gram H, Schmitz R, Zuber JF, Baumann G: Identification of phosphopeptide ligands for the Src-homology 2 (SH2) domain of Grb2 by phage display. *Eur J Biochem* 1997, 246:633-637.
62. van Huizen R, Miller K, Chen DM, Li Y, Lai ZC, Raab RW, Stark WS, Shortridge RD, Li M: Two distantly positioned PDZ domains mediate multivalent INAD- phospholipase C interactions essential for G-protein-coupled signaling. *EMBO J* 1998, 17:2285-2297.
63. Fuh G, Pisabarro MT, Li Y, Quan C, Lasky LA, Sidhu SS: Analysis of PDZ domain-ligand interactions using carboxyl-terminal phage display. *J Biol Chem* 2000, 275:21486-21491.
64. Weiss GA, Watanabe CK, Zhong A, Goddard A, Sidhu SS: Rapid mapping of protein functional epitopes by combinatorial alanine scanning. *Proc Natl Acad Sci USA* 2000, 97:8950-8954.
- The surface of human growth hormone that interacts with its receptor was subjected to a combinatorial alanine-scan. Statistical analysis of the DNA sequences of binding clones was used to calculate the effect of individual alanine substitutions as a change in free energy relative to that of the wild-type, and the results were found to be in excellent agreement with free-energy data previously determined by conventional alanine-scanning mutagenesis. The combinatorial method is very rapid because it circumvents the need for protein

purification and analysis, and the general strategy should be applicable to any protein that can be displayed on phage.

65. Pelletier JN, Arndt KM, Plückthun A, Michnick SW: An *in vivo* library-versus-library selection of optimized protein–protein interactions. *Nat Biotechnol* 1999, 17:683-690.

The authors demonstrate an efficient *in vivo* selection for heterodimeric pairs that should be applicable to a wide-range of protein–protein interactions. Two leucine zipper libraries were fused to murine DHFR and introduced into an *E. coli* host that could only survive if an interaction between two library members reconstituted DHFR activity. Multiple rounds of competitive selection allowed for the isolation of novel leucine zipper heterodimers.

66. Arndt KM, Pelletier JN, Müller KM, Alber T, Michnick SW, Plückthun A: A heterodimeric coiled-coil peptide pair selected *in vivo* from a

designed library-versus-library ensemble. *J Mol Biol* 2000, 295:627-639.

67. Shaywitz AJ, Dove SL, Kornhauser JM, Hochschild A, Greenberg ME: Magnitude of the CREB-dependent transcriptional response is determined by the strength of the interaction between the kinase-inducible domain of CREB and the KIX domain of CREB-binding protein. *Mol Cell Biol* 2000, 20:9409-9422.
68. Karimova G, Ullmann A, Ladant D: Protein–protein interaction between *Bacillus stearothermophilus* tyrosyl-tRNA synthetase subdomains revealed by a bacterial two-hybrid system. *J Mol Microbiol Biotechnol* 2001, 3:73-82.