

UNDERSTANDING GENE AND ALLELE FUNCTION WITH TWO-HYBRID METHODS

Roger Brent

Department of Genetics, Harvard Medical School and Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114; e-mail: brent@opal.mgh.harvard.edu

Russell L. Finley Jr.

Center for Molecular Medicine and Genetics, Wayne State University School of Medicine, Detroit, Michigan 02178; e-mail: rfinley@cmb.biosci.wayne.edu

KEY WORDS: yeast two-hybrid, interaction trap, protein networks, functional genomics

ABSTRACT

Two-hybrid schemes for detecting protein-protein interactions have deepened our understanding of biology by allowing scientists to identify individual important proteins. Recent developments will allow biologists to chart regulatory networks and to rapidly generate hypotheses for the function of genes, allelic variants, and the connections between proteins that make up these networks. Future developments will allow biologists to test inferences about the function of network elements, and allow global approaches to questions of biological function.

CONTENTS

INTRODUCTION	664
YEAST TWO-HYBRID SYSTEMS	666
<i>Generic Yeast Two-Hybrid Systems</i>	666
<i>History</i>	667
<i>Present State of the Art</i>	670
OTHER TRANSCRIPTION-BASED INTERACTION DETECTION METHODS	675
<i>Systems that Detect Interactions Dependent on Third Proteins</i>	675
<i>Systems that Detect Interactions Dependent on Small Molecule Ligands</i>	676
<i>Systems that Allow Selection Against Interactions</i>	677
<i>Two-Bait Systems</i>	677
<i>Mammalian Two-Hybrid Systems</i>	679
<i>One-Hybrid Methods</i>	680

<i>Methods to Measure Protein-RNA Interactions</i>	681
INTERPRETATION OF TWO-HYBRID RESULTS	682
<i>False Positives</i>	682
<i>False Negatives</i>	683
<i>Doctrine for Interpreting Positive and Negative Results</i>	684
THE HYPOTHESIS GENERATING ENGINE	686
<i>Use to Infer Gene and Allele Function</i>	686
<i>Interaction Mating</i>	687
<i>Testing Functional Inferences from Interaction Data</i>	692
<i>Expected Future Developments in Functional Genomics</i>	698
CONCLUSION: A NEW CONTINENT OF GENETIC INFORMATION	699

INTRODUCTION

Saccharomyces cerevisiae was sequenced in 1996 (17, 37). *Caenorhabditis elegans* will be sequenced by the time this article appears. *Homo sapiens* will be sequenced by 2003, but the sequence of its coding genes is moving much faster: as of this writing (1997) some sequence from 80% of its genes is available in commercial databases; and by the time this article appears, it is likely that a similar quantity of gene sequence will be available to the public.

Unfortunately, the sequence of a gene does not always reveal its biological function. For example, from sequence analysis, 30% of yeast genes have known human homologs, 40% are similar enough to other genes in other organisms to suggest function in general, and 30% are unlike any sequence seen before (17, 36). The problem is compounded by the fact that large numbers of genes work together to effect any given biological process, so that to understand biology, it is not sufficient to understand the function of individual genes. Thus, the problem confronting biologists wishing to understand gene function is a daunting one. There are too many genes (and too few biologists) to allow conventional genetic methods (2, 7, 33, 45–49, 52, 53, 56, 81, 89, 90) to establish their biological function within our lifetimes.

Grim as this situation is, it contains two rays of hope. One is that the problem is finite. The human genome may encode 80,000 genes, or 100,000, or 120,000, but it is unlikely to encode 200,000 [for example, see Reference (70)]. The second is that many proteins work by touching one another, either to form lasting functional complexes, or in transient interactions that result in modifications to one of the interacting partners. The biological processes that handle cellular information flow and control cellular decisions, from signal transduction to cell cycle regulation, are largely governed by these protein-protein interactions. If intermediary metabolism is the industrial economy of the cell, then these decision-making networks represent the cell's postindustrial economy (see Figure 1). As we describe below, two-hybrid methods are well suited to dissect the genetic pathways that govern such cellular decisions, and their extension will give insight into pathways that have not yet been discovered.

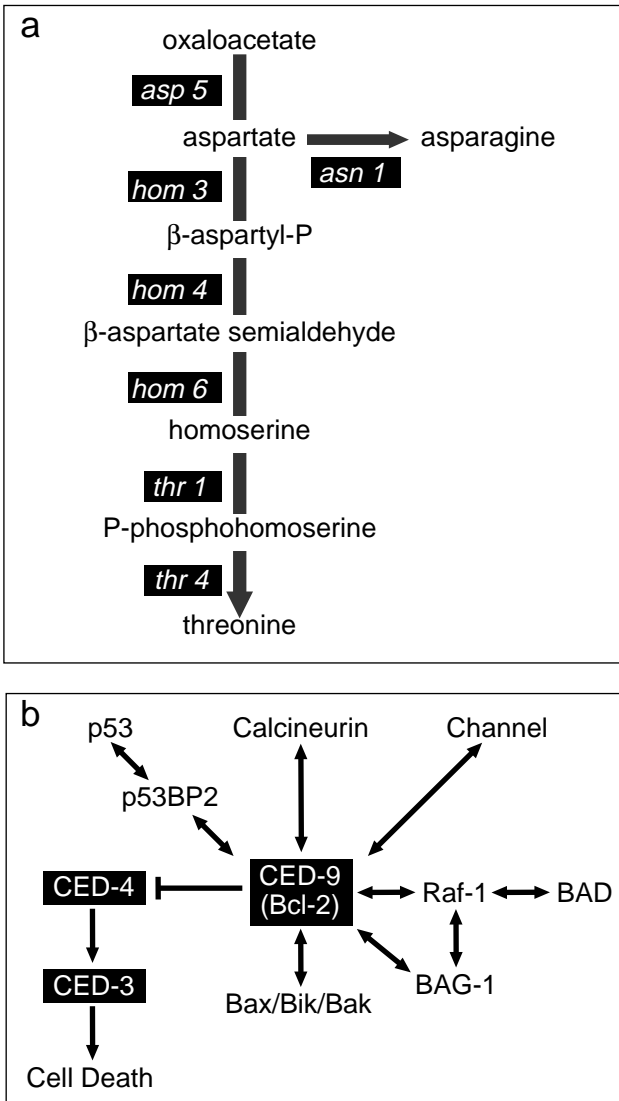


Figure 1 Industrial and postindustrial cellular economies. The *top panel* shows a typical metabolic pathway, in this case the pathway involved in biosynthesis of threonine in *S. cerevisiae* (91). In it, enzymes, here named by the names of the genes that encode them, act successively on small molecule substrates. The *bottom panel* shows a typical decision making pathway, in this case one governing entry into apoptosis in mammalian cells. In it, proteins, here called by the names in common use, act on one another, often by direct protein-protein interaction. The consequence of these interactions is a cellular decision whether to enter apoptosis (77, 91).

Here, we review current uses of simple two-hybrid systems, and we describe some of their more widely used spin-offs. We then describe the emerging use of such systems to chart genetic regulatory networks, to assign function to new genes, to place these into ordered pathways, and to select molecules that can test simple hypotheses about their function. We then describe the possible future uses of two-hybrid and successor methods to provide rough substitutes for some of the methods of classical genetics, and to extend genetic analysis to biological systems so far considered intractable.

YEAST TWO-HYBRID SYSTEMS

Generic Yeast Two-Hybrid Systems

The earliest and simplest two-hybrid system, and the core of many current versions, is an assay *in vivo* for interaction between two specially constructed proteins (27) (Figure 2). The assay is conducted in yeast (*S. cerevisiae*) and uses transcription of yeast reporter genes to measure the protein interaction. One of the two proteins is expressed as a fusion to a DNA-binding domain from a transcription factor, and the other is expressed as a fusion to a transcription

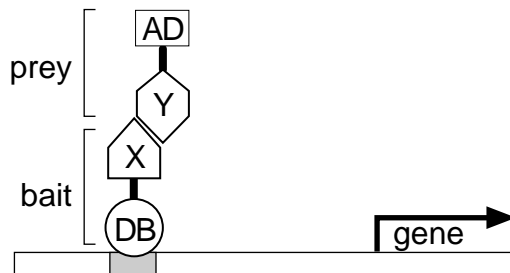


Figure 2 Generic two-hybrid system (27). Interaction of proteins X and Y upstream of a reporter gene in yeast leads to transcription activation. Figure shows that X is part of a chimeric protein, which binds to a site on DNA upstream of a reporter gene (“gene”) by means of a DNA-binding moiety (“DB”). Y is also part of a chimeric protein, which also carries an activation domain (“AD”). The protein interaction that situates the activation domain in the vicinity of the reporter gene stimulates its transcription. Commonly, BD is the DNA-binding domain from either Gal4, or native *E. coli* LexA, and the sites placed upstream of the reporter are Gal4 binding sites or LexA operators, respectively. AD is an activation domain typically derived from Gal4, from VP16, or from B42 (see text). The BD fusion is referred to as the “bait”, since the most common use for it is to fish for or trap library-encoded proteins that interact with a known protein such as X; for this reason, the AD fusion is sometimes referred to as the “fish” or the “prey.” Although the stimulation of transcription by such two-protein complexes to study transcription was anticipated by Ma & Ptashne (64), independent work by Fields & Song (27) developed this general assay for protein-protein interactions.

activation domain. If the fusion proteins interact, they activate transcription of specially designed reporter genes that carry binding sites for the DNA binding partner.

History

TRANSCRIPTION ACTIVATION Development of two-hybrid assays relied on the accumulated understanding of eukaryotic transcription initiation. Experiments done in the mid-1980s established that transcriptional activators are frequently modular, with at least two separable domains: one that binds to DNA, and another that activates transcription (Figure 3). First, experiments in yeast showed that binding to a DNA site in a eukaryote, either by a prokaryotic repressor or by a deletion derivative of a eukaryotic activator, was not sufficient to activate transcription (13, 57). Second, domain swap experiments demonstrated that chimeric proteins that contained a DNA-binding domain fused to portions of eukaryotic transcription activators stimulated transcription. This fact was demonstrated in yeast by expressing a hybrid of the DNA-binding domain from the bacterial repressor LexA and the activation domain from the yeast transcription factor Gal4 (14). The hybrid bound to LexA recognition sites (operators) placed upstream of a *lacZ* reporter gene and activated *lacZ* transcription. Subsequent experiments showed that other chimeric proteins functioned in yeast, including chimeras that consisted of DNA binding domains fused to activation domains randomly encoded by *Escherichia coli* DNA (63).

DNA binding domains Inspection of protein sequences in current databases reveals thousands of proteins with sequence motifs that indicate they bind specific sites on DNA. These proteins (and their DNA binding domains) are often rotationally symmetric dimers or tetramers; in eukaryotes, these typically bind to 10- to 20-bp sites within the promoter or enhancer region of a gene. For example, the two best studied *S. cerevisiae* transcription factors, Gal4 and Gcn4, are dimers that bind recognition sites located within a few hundred base pairs upstream of transcription start points (8, 35, 51). Gal4, Gcn4, and other DNA binding proteins contain identifiable domains that contact these sites. A cursory examination of publicly available protein sequences reveals more than 30 sequence families of such DNA binding domains.

Activation domains Many eukaryotic transcription activators contain at least one activation domain. Various activation domains may work through different mechanisms, but all of them are thought to function by interacting directly or through intermediary proteins with RNA polymerase II- or III-associated proteins in the vicinity of the transcription start site [reviewed in (76)]. Activation domains have in common the property that they do not need to be positioned precisely at the promoter in order to activate (51a, 96).

It is now clear that the modular organization of eukaryotic gene regulatory proteins is a general rule, and that most eukaryotic transcription activators possess distinct DNA binding and transcription activation domains. Yeast two-hybrid systems take advantage of this modularity.

DEVELOPMENT INTO TWO-HYBRID METHODS The efflorescence of rough and ready protein engineering that established the modular nature of transcription

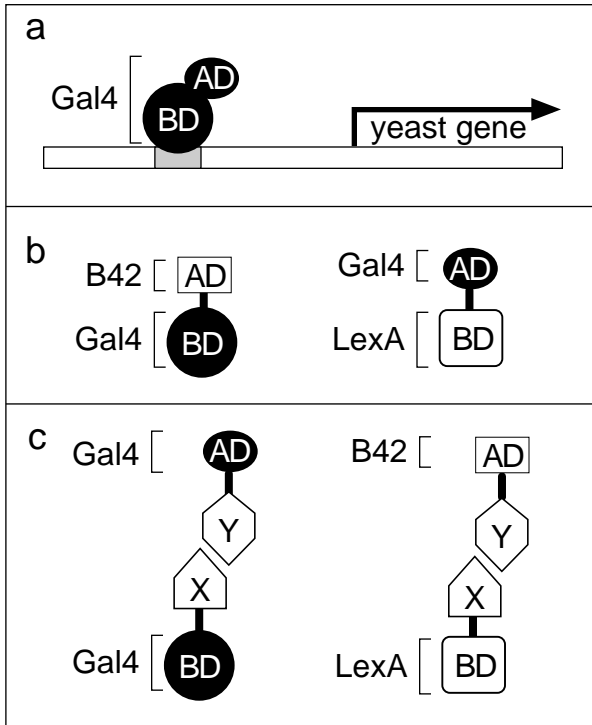


Figure 3 Modular nature of transcription factors. *a.* Gal4 is a yeast transcription factor that binds at specific sites upstream of yeast *GAL* genes to activate transcription. Placement of Gal4 binding sites upstream of other genes or reporters will make them Gal4 responsive. Gal4 is a single polypeptide with separate functional domains: a major transcription activation domain at its carboxy terminus, (AD) and an amino-terminal DNA-binding domain (DB). *b.* Fusion experiments showed that transcription factors are modular. A fusion protein consisting of the Gal4 BD and an acidic activation domain (B42) encoded by *E. coli* sequences can bind to Gal4 sites and activate transcription (63). Likewise, a fusion containing *E. coli* LexA and the Gal4 activation domain can bind to LexA binding sites (LexA operators), placed upstream of a gene and activate transcription (14). *c.* Two-hybrid transcription activators. Fusions of the Gal4 BD with protein X and the Gal4 AD with protein Y can reconstitute an active transcription factor if X and Y interact with each other (27). Fusions of LexA with protein X and the B42 AD with protein Y demonstrate that an active transcription factor can consist entirely of proteins not encoded by the yeast genome (41, 111).

activators provided one inspiration for the two-hybrid methods. Another came from experiments that showed that transcription could be activated by binary protein complexes, in which the DNA binding and activation domains resided on separate polypeptide chains. This was first demonstrated for the naturally occurring protein Oct-1, which possesses a DNA-binding domain that locates it to certain promoters, but only activates those promoters when it is further complexed with a transcription activation domain provided by the Herpes Virus protein, VP16 (88, 95, 96) (this phenomenon is now called “recruitment”). At about the same time, Ma & Ptashne demonstrated that a DNA-bound Gal4 derivative, which could not activate transcription, did activate when it recruited a chimeric protein that contained a Gal4-interacting protein (Gal80) and an activation domain (64). The significance of these experiments lay in their demonstration that the DNA-binding domain and activation domain of a functional transcription factor can reside on different polypeptides, and that transcriptional activity could be reconstituted when those polypeptides interacted. A third source of inspiration for two-hybrid methods came from efforts to use chimeric transcription activators and specialized reporter genes to reveal or confer transcription activation on oncoproteins and regulatory proteins from higher eukaryotes (42, 59), and to use the transcription activation as a phenotype to identify from cDNA libraries proteins that modified transcription activity of those chimeric proteins (EA Golemis & R Brent, unpublished), and thus might interact with them genetically.

However, it was independent work by Fields & Song that resulted in the first general genetic assay for protein interactions (27). These workers measured the interaction between two yeast proteins involved in regulating the *SUC2* gene, Snf1 and Snf4, by expressing them as chimeras. One chimera contained the DNA-binding domain (DB) of Gal4 at the amino terminus of Snf1, and the other contained an activation domain (AD) from Gal4 at the amino terminus of Snf4 (Figure 4). Interaction between the two chimeric proteins brought the activation domain to a *lacZ* reporter containing Gal4 binding sites and was detected when β -galactosidase encoded by *lacZ* caused a colony to turn blue on X-Gal indicator plates. Because Snf1 and Snf4 are not transcription factors, these experiments uncoupled the study of interaction from the requirement that either underivatized partner affect transcription. They also showed that the two-hybrid assay provides a genetic means to measure protein-protein interactions in vivo, and to identify domains and structural features of proteins involved in these interactions. Moreover, the two-hybrid assay suggested a way to isolate new proteins, and the DNA that encodes them, based on their ability to interact with known proteins. Reduction of these ideas to practice took a long time [for example, see Reference (18), discussed below], since it took time to devise systems where the transcription phenotypes were selectable, and to construct high quality interaction libraries (18, 24, 41, 101, 111).

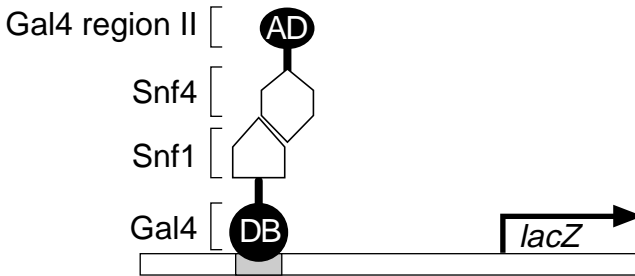


Figure 4 The first two-hybrid assay (27). Fields & Song showed that interaction between a fusion protein that contained the Gal4 DNA-binding region fused to Snf1, and a fusion protein that contained Snf4 fused to the Gal4 C-terminal activation region (region II) resulted in transcription of a Gal4-responsive reporter gene.

Present State of the Art

FINDING INTERACTING PARTNERS The most established use of two-hybrid methods is in interactor hunts, to isolate new proteins from activator domain-tagged libraries that interact with LexA or Gal4 fusion “baits.” In the first published experiment of this kind, Fields and colleagues (18) used a bait that contained Gal4-DB and activation partners that contained the Gal4-AD. First, they constructed yeast plasmids that directed the synthesis of baits and potential interaction partners with these moieties fused to yeast Sir proteins, which regulate transcriptional silencing. They expressed pairwise combinations of the fusion proteins and detected interactions when the *lacZ* reporter was activated. In this manner they showed that Sir4 could interact with itself and with a number of Sir4 deletion derivatives.

Chien et al then showed that this assay could be used to identify new interacting proteins. They transformed a yeast strain that expressed a bait protein, Sir4 fused to Gal4-DB, with a plasmid library expressing yeast proteins fused to Gal4-AD, and identified transformants that turned blue on the indicator plates. While some of the library plasmids in the blue colonies encoded Gal4, which bound to the Gal4 sites carried on the reporter, others encoded proteins that interact with Sir4, including Sir4 itself (as expected from the previous experiments) and a new Sir4-interacting protein, Sfi1 (as hoped) (18). These results demonstrated the potential of yeast two-hybrid methods to identify new proteins that interact with known proteins, along with the genes or cDNAs that encode them.

The two-hybrid method for isolating new proteins was further developed in several labs (24, 41, 101). At about the same time, Dalton & Treisman (22) used a method with serum response factor (SRF) and reporters containing SRF binding sites to isolate SRF-interacting proteins from a cDNA library expressing

fusions to VP16. Although the SRF bait in these experiments bound DNA on its own, and this experiment was thus not technically a two-hybrid experiment, the conceptual similarity is clear.

The first significant improvement to the system was the inclusion of additional reporters whose expression is required for yeast growth. Reporters derived from the yeast *LEU2* gene or *HIS3* gene enabled researchers to apply a selection for yeast expressing them by plating yeast containing library plasmids on media lacking leucine or histidine, respectively (41, 24, 101). Use of these selectable marker genes transformed the interactor hunt procedure from a screen to a selection for interacting proteins, allowing relatively rare clones to be isolated. Most systems now use both the selectable marker and *lacZ* (Figure 5). By requiring both reporters to be activated during an interactor hunt, the investigator can identify and discard cells that contain library proteins that may bind specifically to the regulatory region of one of the reporters, rather than to the bait.

A typical interactor hunt (1, 3, 29, 31) begins with a strain that contains a *LEU2* or *HIS3* reporter along with a *lacZ* reporter, and that expresses a bait protein of interest fused to LexA or to the Gal4-DB. The strain that contains the bait is then transformed with a plasmid library that expresses cDNA encoded proteins fused to an activation domain. Transformants are generally selected in liquid or on plates, then plated onto selection plates where only yeast expressing the *LEU2* or *HIS3* gene grow. These are then tested for activation of the *lacZ* reporter by placing them onto X-Gal plates or by a filter β -galactosidase assay (1).

The first step in characterizing colonies in which both reporters are active is to show that activation of the reporters is due to the activation tagged protein (sometimes called the “prey” or “fish” protein) and not to a yeast mutation. This is done by reintroducing the library plasmid into the original strain that expresses the bait and recapitulating activation of the reporters. A second important test is to show that the library-encoded protein interacts specifically with the protein of interest and not with, for example, the LexA or Gal4 moiety of the bait, or with other unrelated bait proteins. This is done by introducing the library plasmid into strains expressing other bait fusions and showing that cDNA expression results in activation of the reporters only in strains containing the original bait. The specificity test is often done using an interaction mating assay as described below. This assay is useful in identifying nonspecific interactors, proteins that appear to interact with many different baits.

COMPARING DIFFERENT SYSTEMS During the early 1990s, a number of labs developed two-hybrid systems, each of which differ in their components.

Reporters The three systems in most common use (24, 41, 101) differ significantly in their reporters. All three systems use *lacZ* reporters derived from a *GAL1-lacZ* fusion that has the yeast *GAL1* promoter lacking its upstream

transcription control region, and a small portion of the *GALI* coding region fused in-frame with *lacZ* (105, 110). These *lacZ* reporters have various numbers of either LexA operators or Gal4 binding sites placed upstream of the transcription start site. These reporters exist on derivatives of the original multicopy yeast plasmids or on plasmids that can be integrated into the yeast genome. The reporter phenotypes from the integrated *lacZ* reporters are weaker, typically making these reporters less sensitive to weak interactions.

In addition, as described above, these systems feature a second reporter gene whose transcription restores a nutritional prototrophy. Yeast strains bearing chromosomal *HIS3* or *LEU2* reporters with upstream binding sites for either Gal4 or LexA are in common use. These selectable reporters are highly sensitive to transcriptional activation allowing detection of even weak interactions [for example see Reference (25)]. Their high sensitivity, however, can present a significant problem when using bait proteins that by themselves activate transcription. The *HIS3* reporters introduced by Vojtek et al (101) and Elledge and coworkers (24) provide one way to deal with such activating baits. It is possible to adjust the amount of *HIS3* activation required for histidine prototrophy, and hence the sensitivity of the reporter, by adding a competitive inhibitor of the *HIS3* gene product, 3-aminotriazole (3-AT). Thus, for baits that activate *HIS3*, 3-AT is added to the medium to a level that makes the strain auxotrophic for histidine, and then the activation library can be screened to detect interacting proteins that result in further activation of *HIS3* (24).

Similarly, the sensitivity of the *LEU2* reporters can be adjusted by adding 6-fluoroleucine to the medium (C Denis, personal communication). However, in the system described by Gyuris et al (41), the sensitivity of the *LEU2* and *lacZ* reporters is usually adjusted by choosing versions with different numbers of upstream LexA operators.

Bait expression vectors In the workhorse versions of the three commonly used systems, baits are expressed from plasmids as moieties fused C-terminal to a LexA or Gal4 moiety. Both the Gal4 moiety and the "full-length" LexA moieties used contain sequences that promote dimerization of the fused protein and thus binding to the reporters. The most significant difference between these two domains is that the Gal4 amino terminus contains a nuclear localization signal while LexA does not (86). This difference has several consequences. First, for many LexA fused baits, the nuclear concentration is lower than it would be if the bait were localized to the nucleus, although the resulting nuclear concentration is typically sufficient for the bait, as judged by *in vivo* repression assays (13), to occupy operators more than 50% of the time (39); since the prey is always localized to the nucleus, this fact ensures that the concentration of the prey is usually greater than the concentration of the bait, which allows the

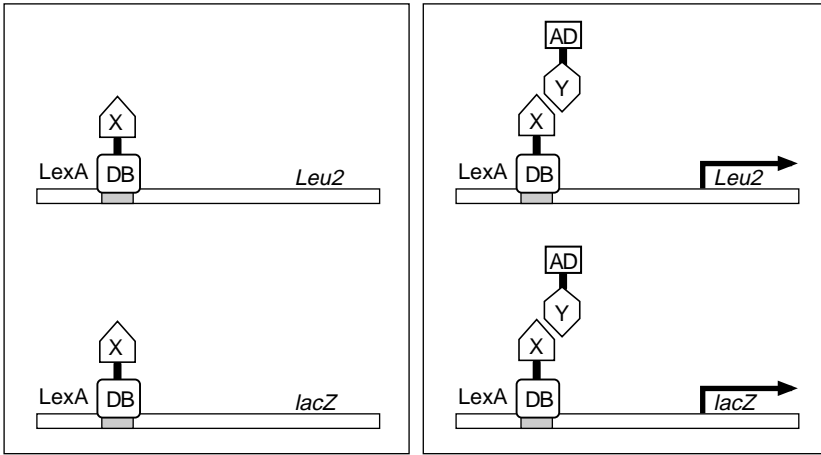


Figure 5 A dual reporter system (41). The *left panel* shows a yeast cell with a bait, LexA-X, upstream of LexAop-*lacZ* and LexAop-*LEU2* reporters. Except for the reporters, the cells lack functional *lacZ* and *LEU2* genes. These reporters are chosen by the investigator to have the desired level of sensitivity to protein interactions. The *right panel* shows that transcription of both reporters is activated by interaction of the bait with a second fusion protein, Y fused to an activation domain. Cells with both reporters active form colonies on medium lacking leucine, and those colonies are blue on X-gal medium. Expression of the Y-AD fusion protein is conditional, and the reporter phenotypes are thus observed only under conditions where Y-AD is expressed.

bait concentration to be neglected in calculations of the approximate affinities these systems can detect [used in (25)]. Second, the typically lower nuclear concentration of LexA baits means that, when expressed from integrating vectors, many LexA fusions are not produced in sufficient amounts to efficiently occupy the reporters, leading to a loss of sensitivity not found for integrating Gal4 bait expression vectors. A number of variant LexA expression vectors exist, including ones that contain nuclear localization sequences, ones that fuse LexA to the C terminus of the bait, and ones that induce the synthesis of the bait only when cells are grown on galactose (12, 34a).

Prey expression vectors The systems differ more significantly in their prey vectors (Figure 6). The prey vector used by Gyuris et al (41) uses the relatively weak activation domain, B42, encoded by *E. coli*, and expresses fused proteins that also contain a nuclear localization sequence and an epitope tag. The B42 activation domain may increase the spectrum of proteins recovered by obviating the toxic effects (squenching) that strong transcription activators have in yeast (34). The system developed by Elledge et al (24), uses Gal4 activation domain II, which is derived from the C terminus of the protein (65). The system

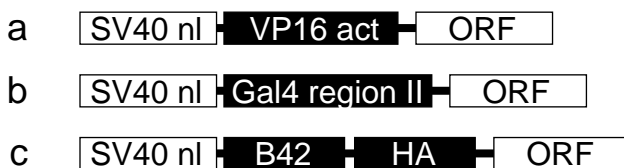


Figure 6 Prey proteins in the most commonly used systems. Figure shows the moieties carried on activation domain expression vectors from the systems described by Hollenberg, Sternglanz & Weintraub (101); by Becherer, Kilburn & Elledge (24); and by Gyuris et al (41). Because of the requirement that open reading frames (ORFs) encoded by cDNA inserted in these vectors be expressed in frame with the fused moieties, all these vectors have the ORF expressed C terminal to their other moieties. In all these systems, preys are localized to the nucleus by addition of a nuclear localization sequence derived from SV40 T antigen. Preys expressed by the Gyuris vector include the HA epitope tag from influenza virus hemagglutinin to facilitate their detection and purification by immunological methods. Preys in the Gyuris system are expressed under the control of the yeast *GALI* promoter whose expression is shut off if the cells are grown on glucose; the prey vector in the Becherer, Kilburn & Elledge system makes the yeast sensitive to cyclohexamide, so that only yeast that have lost the prey plasmid will grow on such medium; both systems thus provide a means to verify that reporter transcription depends on prey. The major difference between preys expressed by these systems lies in the activation domain. The VP16 activation region [residues 401–479] is a stronger activation domain than the C-terminal region activation region II of Gal4 [residues 768–881], and Gal4 region II is a stronger activation domain than B42. Use of B42 may increase the spectrum of proteins recovered in hunts by eliminating the toxicity (squenching) that Gal4 and stronger transcription activators can have on yeast (34).

described by Vojtek et al (101) uses the still more powerful activation domain derived from herpes simplex virus VP16.

The prey expression vectors also differ in whether they allow verification that the reporter phenotypes depend on expression of a protein encoded by the prey plasmid. One way to facilitate such verification is to express the prey conditionally. Gyuris et al (41) pioneered the expression of cDNA-encoded proteins from the intact yeast *GALI* promoter. This promoter is only active in cells grown in galactose but repressed in cells grown in glucose. Use of inducible libraries allows one to identify and eliminate one of the large classes of false positive that arise during a hunt. These occur in cells that, because of genetic or epigenetic changes, activate the reporter genes independently of the library protein. For example, it is common to observe some bait proteins that activate the reporters at a very low level; during a hunt, cells often arise in which more of the bait is expressed; and in which the reporters are active (EA Golemis & R Brent, unpublished). For some baits a large fraction of the positives fall into this class. In the system designed by Gyuris et al, these false positives are identified because the reporters are active even when the cDNA is not expressed, i.e. in glucose. Inducible libraries also offer the advantage of allowing isolation of cDNAs that encode mildly toxic proteins (41).

Unfortunately, these galactose-inducible libraries cannot be used with systems that use the Gal4 DNA-binding domain; in these systems the yeast strain must lack endogenous Gal4, which is needed to activate the *GAL1* promoter. Another approach to the problem, introduced by Elledge and coworkers, is to make prey expression plasmids that carry a counter-selectable marker (e.g. *Cyh^s*, which causes sensitivity to cyclohexamide). Cyclohexamide-resistant host cells that no longer carry the prey plasmid can form colonies on medium containing cyclohexamide, allowing the investigator to select cells that have lost the prey plasmid and to verify that those cells no longer display the transcription phenotypes.

More recently, researchers have combined and modified components of different systems to consolidate some of their advantages into integrated user-friendly systems [for example, see References (12, 31, 55, 58)]

OTHER TRANSCRIPTION-BASED INTERACTION DETECTION METHODS

Systems that Detect Interactions Dependent on Third Proteins

Some protein-protein interactions cannot be detected with standard two-hybrid methods because they require a third molecule not normally available in yeast. For example, the affinity of two proteins that contact each other directly may be enhanced by the expression of a third protein that contacts both (Figure 7a). Elion, Sprague, and Wigler and their coworkers observed this for two yeast signal transduction proteins, Ste7 and Ste11, which interact better in yeast that express Ste5, a protein they both contact; subsequent biochemical experiments confirmed that Ste5, Ste7, and Ste11 exist in a multiprotein complex (20, 66, 75). In another example, Ozenberger & Young (74) demonstrated enhanced dimerization of the mammalian growth hormone receptor (GHR) in cells that also expressed the peptide growth hormone (GH).

Expression of a third protein can also be used to detect interactions between proteins that make no direct contact, but interact solely via a third, bridging protein (Figure 7b). Wigler and colleagues demonstrated, for example, that the signal transduction proteins Ras and Mek interact in a complex bridged by Raf, and that *S. pombe* Scd1 bridges interactions between Cdc42 and scd2, and between Cdc42 and Ras (15, 97).

Finally, expression of a third protein can sometimes facilitate a protein-protein interaction without forming a lasting part of a stable complex (Figure 7c). This situation arises for interactions that depend on posttranslational modifications that are relatively uncommon in yeast. The only such examples we are currently aware of are for interactions that depend on tyrosine phosphorylation. For example, Osborne et al (73) expressed a tyrosine kinase, Lck, while

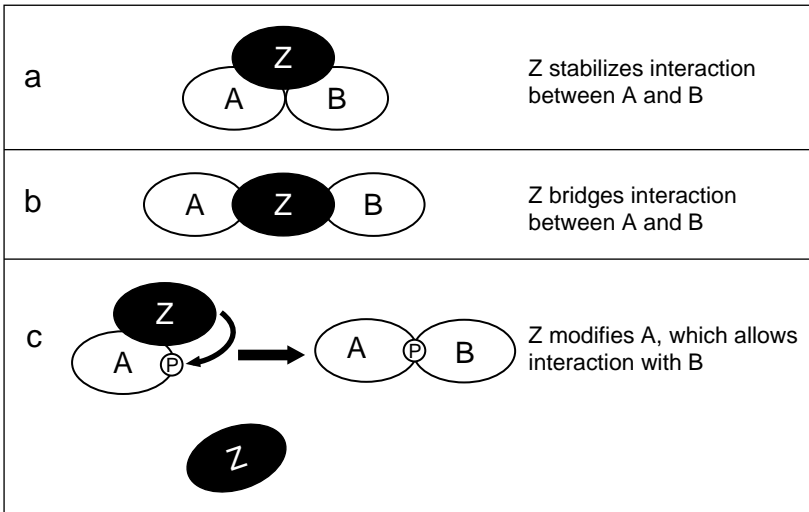


Figure 7 Different kinds of ternary protein complexes. *a.* Stabilizing. Protein Z stabilizes the interaction of A and B. *b.* Bridging. Z contacts A and B with different surfaces; A and B make no direct contact. *c.* Sequential. Z modifies A and dissociates. Modified A then contacts B.

conducting interactor hunts with intracellular portions of receptors as baits. By this means they were able to isolate proteins with immunoreceptor-based tyrosine activation motifs (ITAMs), which interacted with tyrosine phosphorylated forms of the receptors.

The fact that multiprotein complexes can be detected by two-hybrid methods can be exploited in interactor hunts. For example, a third protein can be expressed to facilitate or enhance interactions between a bait protein and library proteins. Alternatively, a hunt can start with a cell that expresses a bait and prey protein that does not interact or that interacts poorly, and an expression library screened for proteins that bridge or enhance their interaction.

The existing results with three-protein systems suggest an obvious cautionary note for interpreting two-hybrid interactions; in any given case, it is conceivable that a detected binary interaction depends on one or more yeast proteins, and thus may not involve direct contacts between the individual partners.

Systems that Detect Interactions Dependent on Small Molecule Ligands

Augmented two-hybrid assays are also useful in the detection of interactions that depend on a third, non-protein, ligand. Lee et al (60), for example, isolated proteins that interact with the thyroid hormone receptor (TR) by conducting two-hybrid hunts either in the presence or absence of thyroid hormone. They

found two classes of TR interacting proteins: those that interacted with TR only in the presence of thyroid hormone, and those that interacted with TR only in the absence of the hormone. Similarly, Chui et al (19), identified library proteins that interacted with the human FK506/rapamycin-binding protein, FKBP12, only in the presence of rapamycin, and Wang et al (102) demonstrated a rapamycin-dependent two-hybrid interaction between FKBP12 and the type I TGF- β receptor.

Licitra & Liu (62) developed an augmented two-hybrid system and used it to identify receptors for small organic ligands like FK506 (Figure 8). They expressed LexA fused to the rat glucocorticoid receptor (GR) in yeast in the presence of an organically synthesized hybrid molecule consisting of a ligand for GR, dexamethasone, and a so-called bait ligand, in this case FK506. They then screened a standard human activation domain fusion library to find proteins that bind FK506, brought to the reporter through the interaction of dexamethasone with LexA-GR. Using this system, they were able to isolate FKBP12 from a cDNA library, thus demonstrating that this approach could be used to identify new receptors for ligands.

Systems that Allow Selection Against Interactions

The products of some reporter genes can be selected against (Figure 9). These counter-selectable reporters are useful in allowing selection of mutations in one or the other hybrid protein that abolish interaction. They may also be useful in the task of identifying reagents that disrupt specific protein interactions, as discussed later.

There are several reporters whose expression is deleterious to yeast growth. Expression of yeast *URA3* reporters (100, 108) kills yeast in the presence of 5-fluoroorotic acid (5-FOA); 5-FOA is converted by the *URA5* gene product into 5-fluorouridine monophosphate, which is converted by the *URA3* gene product to 5-fluorouridine monophosphate; this is eventually converted to 5-fluorodeoxyuridine monophosphate, which competitively inhibits thymidylate synthetase and blocks DNA synthesis. The yeast *LYS2* gene has also been used as a two-hybrid reporter (RL Finley & R Brent, unpublished); selection against *LYS2* expression can be done in media containing alpha-aminoadipate, which again is converted into a product that keeps the cell from growing. A *GAL1* reporter has also been used for negative selection (107); in a *gal10* yeast grown on galactose, in which *GAL1* is expressed, the compound galactose-1-phosphate accumulates and kills the yeast.

Two-Bait Systems

Xu et al (manuscript submitted) described a dual-reporter containing cells in which transcription of selectable and counter-selectable reporters is directed by different baits (Figure 10). Such systems are useful to isolate proteins

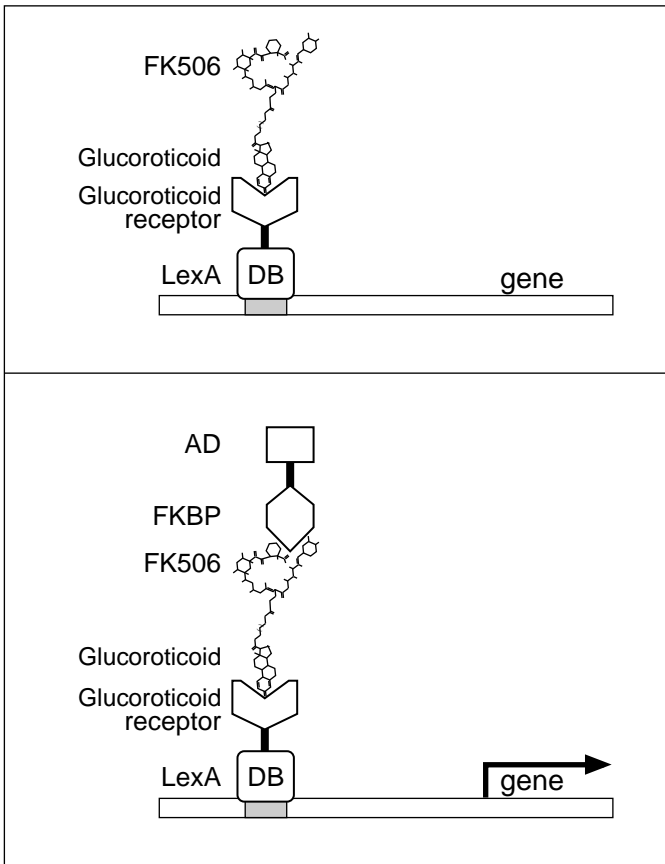


Figure 8 Selecting proteins that interact with small organic molecules. Licitra & Liu (62) showed that a LexA-Glucocorticoid receptor chimera could interact with a synthetic Glucocorticoid-FK506 compound (*top*), and that yeast that contained the LexA-GR fusion and were grown on medium containing the steroid-FK506 chimera could select FKBP12 from a human interaction library by reporter gene activation (*bottom*).

that simultaneously contact the different baits (“bridge proteins”, see above), and those that contact one bait but not the other (“discriminatory proteins”) (56a, 108). Because the different baits can be closely related, and in fact can be allelic variants of the same protein, two-bait systems promise to facilitate the isolation of naturally occurring proteins that preferentially interact with either a wild-type or a disease-state variant of a protein, and of peptide aptamers that interact with either the wild-type or with mutant allelic forms [for example, see Reference (106)].

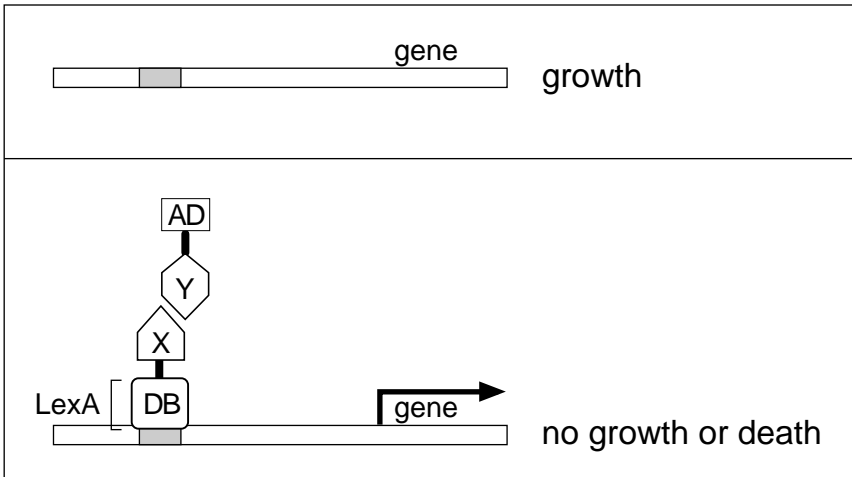


Figure 9 Counterselectable reporters. Figure shows a LexA-operator-containing reporter. Depending on the identity of the reporter gene and the composition of the medium on which the cell is grown, expression of the reporter keeps the cell from growing (LYS2 on α -amino-adipate) or even kills the cell (URA3 on 5-FOA, GAL1 in *gal10*⁻ cell, on galactose).

Mammalian Two-Hybrid Systems

The effectiveness of yeast two-hybrid systems has inspired development of similar systems for detection of protein interactions in mammalian cells. Such systems offer potential advantages, such as detection of interactions that depend on posttranslational modifications that yeast may not provide. Moreover, they might allow identification and monitoring of interactions that change in response to cell stimulation.

Like the yeast systems, the mammalian systems rely on reconstitution of active transcription factors, for example, from two different chimeric proteins that, respectively, contain the yeast Gal4 DNA-binding domain and the VP16 activation domain. Interaction leads to activation of the gene encoding chloramphenicol acetyl transferase (CAT), whose activity is then measured in cell extracts (23,26). To date, mammalian two-hybrid systems have been most useful for examining interactions between known proteins [for example, see Reference (92)]. However, the potential exists for developing working selections that would enable library screening in mammalian cells. For example, two-hybrid reporters have been developed that would allow selection of cells expressing them, including the CD4 cell surface marker, and a gene that confers Hygromycin B resistance (26). Another system has been described that uses the SV40 T antigen as the reporter and replication of a plasmid that requires

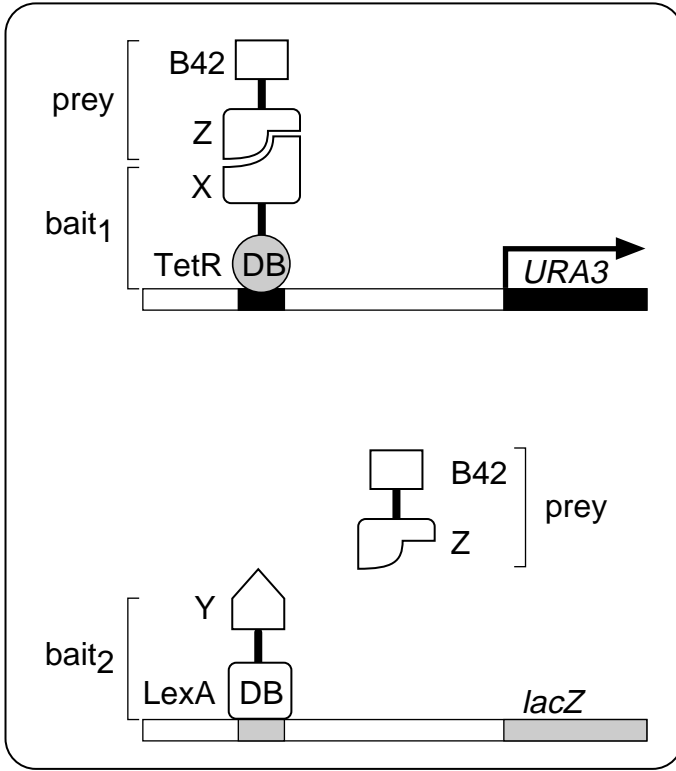


Figure 10 A two-bait system (106, 108) that simultaneously registers two different protein interactions. Figure shows a single cell with two reporters, Tetop-URA3, and LexAop-lacZ. In this cell, a prey protein, Z-B42, interacts with one bait, TetR-X, but not the other, LexA-Y, with expected effects on reporter transcription.

the large T antigen for replication (98) as a readout. Such systems are unlikely to find widespread use until some experimenters are willing to invest in them the work needed to make them rugged and easy to use.

One-Hybrid Methods

One-hybrid systems are methods for identifying DNA-binding proteins that recognize a particular stretch of DNA (54, 61). The bait in this case is a DNA sequence that is placed upstream of a reporter gene like *lacZ*. Libraries identical to those used in the two-hybrid systems, with proteins expressed as fusions to an activation domain, can be screened. Library-encoded proteins that bind the

target DNA result in activation of the reporter. Li & Herskowitz used such a system to identify yeast Orc6, a protein that recognizes a sequence found in a yeast origin of replication (61).

Methods to Measure Protein-RNA Interactions

Other two-hybrid system spin-offs include so-called three-hybrid systems, for detecting protein-RNA interactions (84). In one such system, one of the hybrids is LexA fused to a well-characterized sequence-specific RNA-binding protein from bacteriophage MS2. The second hybrid is an RNA molecule that contains both a test (bait) sequence and a sequence that forms a stem-loop structure recognized by MS2; this chimeric RNA associates with LexA-MS2 molecules bound to sites upstream of a reporter gene. The third hybrid is a protein fused to an activation domain, as in the two-hybrid assay; if this protein binds to the RNA bait, transcription of the reporter gene will be activated. This method has been used successfully in test experiments with a number of known RNA-protein interactions, including the iron regulatory protein binding to iron response elements, the HIV transactivator Tat binding to Tat response elements, and the stem-loop binding protein SLBP binding to sequences involved in regulating histone pre-mRNA processing (84, 103). This three-hybrid system thus has the potential to be used to identify new RNA-binding proteins that recognize known target RNAs, and RNA sequences recognized by known proteins (Figure 11).

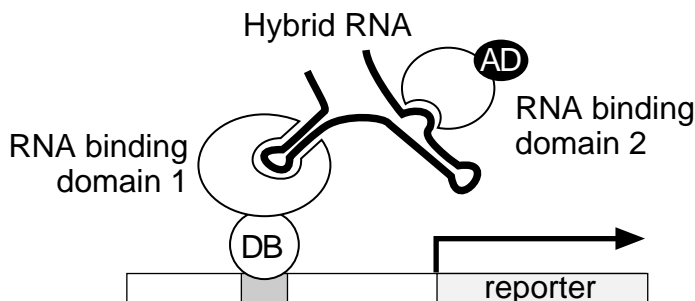


Figure 11 Selecting proteins that interact with RNAs. The first hybrid molecule contains a DNA-binding domain (DB) fused to a well-characterized sequence-specific RNA-binding protein (RNA-binding domain 1). The second hybrid molecule is an RNA containing a sequence that is recognized by RNA-binding domain 1, and a test sequence. The third hybrid contains a transcription activation domain (AD) fused to a test protein. If the test protein interacts with the test RNA sequence (and hence contains RNA-binding domain 2) the reporter will be transcribed. Adapted from Reference (84).

INTERPRETATION OF TWO-HYBRID RESULTS

False Positives

Interactor hunts with current technology still result in proteins that satisfy all the selection criteria, and thus definitely interact with the bait in the yeast nucleus, but that for one reason or another are not highly valued by the investigators who perform the hunt. These proteins are typically referred to as “false positives.” Analysis of two-hybrid experiments benefits from an ability to identify these proteins, and from the realization that they typically fall into different classes, some of which are informative about the biology of the problem under investigation. Classes of false positives include:

1. Interactions that are informative but that could never occur in nature. These include specific interactions between proteins that are normally not expressed in the same cell at the same time. Although not in themselves significant, such interactions can be informative in that they suggest the existence of similar interactions between related molecules that are co-expressed, and in that they can immediately suggest testable hypotheses about the function of the interacting proteins.
2. Interactions that may well occur in nature but that are not informative. Here we define “not informative” in a strict operational sense, to mean “cannot be used to suggest definitive experiments at the time of the finding.” For example, one class of frequently observed protein interactions are those between baits and members of the ubiquitin-dependent proteolysis pathway (38). Although it is possible that some of these interactions are relevant to an understanding of how the protein is usually degraded, and thus to an understanding of the function of the bait, there is often too little that can be done experimentally to follow up the proteolysis idea to make it easy to explore, and for that reason, one cannot determine experimentally if it is relevant.
3. Interactions that presumably do not occur in nature. We and others initially imagined that such interactions might be fairly frequent. For example, we imagined that unrelated proteins might associate using common motifs, such as amphipathic alpha helices, even though the proteins are never present in the same component of the same cell at the same time. Instead, for whatever reasons, with the exception of a relatively small number of “sticky” or “promiscuous” proteins, actual interactions that are clearly false seem to be fairly infrequent. Perhaps the best evidence on this has been collected by Golemis & Serebriiskii in an ongoing survey (38). Analysis of results from almost 100 hunts has revealed that the majority of them resulted in proteins that made sense to the investigators. Among the proteins that did

not make sense to the investigators were many that were isolated in multiple hunts using different baits. These frequently isolated false positives, which included heat shock proteins, ribosomal proteins, and proteosome subunits, may be generally sticky proteins or may activate the reporters in some other way, for example by enhancing the activation ability of a particular bait by making it a more stable protein.

False Negatives

As biologists move from the desire to detect any proteins that interact with their proteins of interest to the desire to survey and catalog all of the protein interactions in pathways and genomes, the problem of false-negative protein interactions has become more important. For a biologist, an undetected interaction might represent a path into the biology of a system that cannot be taken. For a researcher in a pharmaceutical company, an undetected interaction might represent a missed target for small molecule drug discovery.

It is worth noting that many imaginable causes of false negatives have also not proven to be great problems in practice. These imagined causes include the idea that the fragments of proteins encoded by these chimeras might misfold into non-native structures; here, unexpectedly, experience has shown that many proteins are far more modular in structure than would have been predicted in 1980. Other worries have included the fact that two biological processes that affect secreted proteins do not occur in the yeast nucleus: secretory glycosylation, and disulfide bond formation. Here, experience, mostly in industrial settings, and thus slow to appear in the literature, has been good: Although most secreted proteins are glycosylated, in most cases this glycosylation is not necessary for their folding and apparently contributes little to the free energy of their interactions. Correspondingly, although formation of incorrect disulfide linkages can cause a protein to misfold, many disulfide linked proteins function well as baits, suggesting that they are folding correctly and that the disulfides contribute rather to their stability in extracellular environments (M Stahl, personal communication). Failure of the DNA binding and transcription activation moieties to work as predicted is quite rare: Experience has shown that most LexA fusion proteins bind LexA operators in the yeast nucleus (39, 86) and that, in the one case where a moiety on the protein blocked nuclear entry, addition of a nuclear localization sequence to the bait was sufficient to overcome the problem (9). Finally, except in the rare cases where either the bait or prey contains a powerful repression domain (A Reymond, R Brent, unpublished), we know of no cases where the activation domain has failed to activate once moved close to the promoter.

Experience has, however, defined a large number of ways that protein interactions can fail to be detected in two-hybrid systems. First, current systems are unable to detect protein interactions with equilibrium dissociation constants

weaker than 10–50 μM . Many important protein interactions, including, but not limited to, many important enzyme substrate interactions in signaling pathways, many interactions that typically take place in the plane of a membrane, and many cooperative interactions between DNA bound transcription regulatory proteins, are below this threshold. Second, weak interactions result in lower plating efficiencies, leading to their under-detection in standard interactor hunts (25). Third, the use of cDNA libraries brings representation problems, from loss of rare messages to depletion of the amino termini of large proteins, simply from the ways such libraries are often constructed. Fourth, the toxicity caused by activation domains and DNA bound activation domains [squelching, (34)] may lead to underrepresentation of certain protein interactions in systems that use strong activation domains (41). Fifth, the fact that both interacting partners have bulky multidomain moieties fused to their amino termini means that these moieties may block interactions that depend on the amino termini of the partners. Sixth, the requirement that the interacting proteins be properly folded in aqueous environment of the yeast nucleus means that potential interacting proteins that normally span a membrane multiple times are unlikely to fold correctly.

Despite the need to diminish the number of false-negative interactions, there are limits to how much of a decrease will be possible by improvement of existing systems. While we can predict that improvements in reporter genes will result in detection with increased sensitivity (perhaps to detection of interactions with K_{d} s up to 100 μM), such increases in sensitivity will increase the number of false-positive interactions. Similarly, in the context of genome-wide surveys (see below), it is unlikely that systematic surveys will be done to cover all the posttranslational modifications not found in yeast (although it is possible to imagine that systematic interaction surveys might be done with numerous tyrosines in the yeast artificially phosphorylated). Given the constraints imposed by the fact that these assays use yeast as a test tube, a test tube that must survive for the assay to be scored, it is likely that the best approach to capturing interactions missed by two-hybrid experiments will be to supplement them with interaction data from physical methods.

Doctrine for Interpreting Positive and Negative Results

The doctrine for interpretation of positive results is relatively highly developed. Common practice prescribes that the proposed interaction should also be observed by a different technique, such as co-immunoprecipitation of the putative interactors from the appropriate cell or tissue type. These co-precipitation experiments are in vitro experiments: They require breaking cell membranes, and the interactions observed in them occur at different salt concentrations, typically in the presence of detergents, and are thus subject to artifacts. However, at least

these artifacts are likely to be different than those encountered in two-hybrid approaches.

An alternative, easier, and global approach to the identification of potential false positives is to determine whether the interacting proteins are ever expressed in the same cell at the same time. The increasing public availability of expression information, from exhaustive sequencing of libraries from different cells and tissues, and from large-scale protein localization studies [for example, see References (80, 82, 83, 99, 104)], has increased the likelihood that this information may be available without doing any further experiments.

Confirmation of the validity of an interaction can also come from two-hybrid experiments themselves. Probably the strongest such criterion is suppression of an interaction defect. For example, Hardy et al (43) found a yeast protein, Rif1, that interacted strongly with the wild-type but not a missense mutant of Rap1, a transcription factor involved in replication and silencing of telomere transcription. These workers then selected a Rif1 mutation that restored interaction with the mutant Rap1 bait. This Rif1 mutation also suppressed the effect of the yeast Rap1 mutation *in vivo*.

Confirmation of the validity of a given interaction can also come from demonstrating that the interaction is specific and makes biological sense. This is the most important criterion (in use today) for determining the validity of interactions. One simple confidence building technique comes from two-hybrid experiments themselves. Even before the functional genomic implications of two-hybrid experiments were widely appreciated, it was apparent that specific interaction of a protein with a given bait, but not with members of a panel of related and unrelated baits, lends credence to the idea that the interaction is real (41, 44, 101, 111). This confirmatory technique will become more widely used as two-hybrid information becomes more widely available.

Finally, a sometimes-used, but frequently unreliable criterion for validity of an interaction is affinity. Interaction affinity, normally quantified as the equilibrium dissociation constant or K_d , can be crudely estimated from interaction phenotypes when attention is paid to the stoichiometry of the interacting components (41). Recently, Golemis and coworkers showed that it can be better estimated by the use of ordered sets of reporter genes of various sensitivities (25). However, most biologists now realize that many important and specific interactions, from enzyme-substrate interactions to cooperative interactions between gene regulatory proteins, are weak, and that the correlation between the strength of an interaction and its significance is thus imperfect at best.

The reader will have noticed that none of the methods we have listed for determining the significance of a particular interaction is automatic. They are methods that produce results that must be evaluated by the investigator, using arbitrary objective criteria (e.g. a known sequence vs an unknown sequence), using arbitrary subjective criteria (e.g. a tyrosine kinase vs tyrosine

phosphatase), or using criteria evolved from the investigator's knowledge of the particular biological system.

By contrast, the criteria to show that a negative is a true negative are not established at all. The need for such criteria will become more acute in the future, as we generate comprehensive maps of protein connections (4, 30) and desire to extract maximum information from them.

THE HYPOTHESIS GENERATING ENGINE

Use to Infer Gene and Allele Function

During the 1970s, a number of prophetic geneticists described the use of genetic suppression as a window into biology [for example, see Reference (45)]. Given a mutation in a gene, one could isolate unlinked mutations that suppressed the effect of the first mutation. These mutations might lie in proteins that functioned in the same pathway as the first gene. For example, gain-of-function mutations that suppressed recessive mutations might function downstream in the same pathway, while recessive mutations that suppressed other recessive mutations might define genes whose products touched one another. In both cases, identification of second-site suppressors identified genes that were conceptually linked.

In the early 1990s, a clear analogy between two-hybrid experiments and suppressor genetics emerged. Interactor hunts identified proteins that interacted with a bait. These preys could be made into baits, and used in turn to identify new interactors. Here, the linkage was physical rather than conceptual: Although the fact that proteins touch one another does not mean they function in the same process (see above definition of false positives), proteins that touch one another frequently do. The ability of two-hybrid experiments to provide analogs to the relatively sophisticated suppressor genetic technologies opened genetic networks in human and other genetically intractable organisms to pathway analysis.

Interactor analysis has similarities and differences to the suppressor analysis that inspired it (Figure 12). Like suppressor analysis, interactor analysis can identify members of genetic pathways. Like suppressor analysis, the positions and identities of nodes in those pathways would sometimes give clues to the function of pathway members and the pathway as a whole. However, unlike suppressor analysis, interactor analysis cannot reveal causal linkages, only strong physical ones, so proteins that act through third proteins or small molecule intermediates, or that interact with one another more loosely than the detection threshold, are not revealed. Unlike reciprocal temperature-shift experiments on partners identified by suppressor analysis, two-hybrid experiments cannot reveal the order with which components act (Figure 12). These

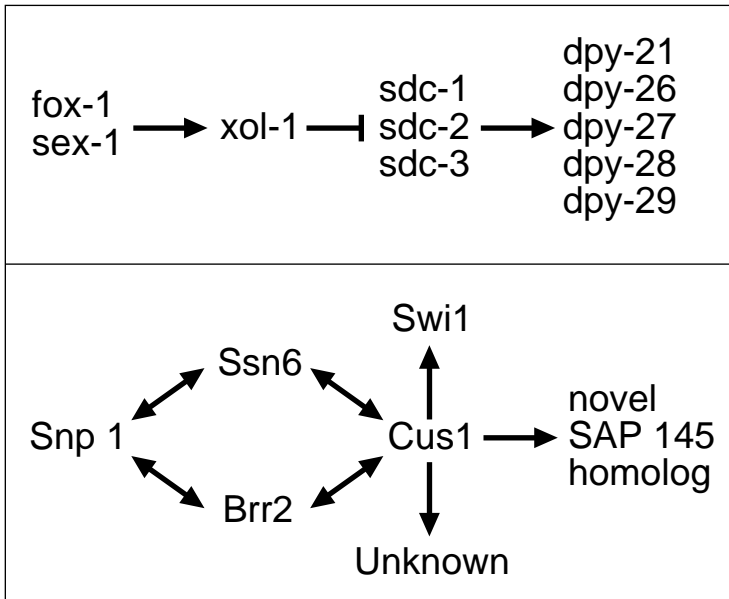


Figure 12 Comparison of genetic pathways and interaction pathways. *Top panel*, adapted from (11), shows a genetic network that governs X-chromosome dosage compensation in the nematode, largely the work of B Meyer and coworkers, and largely drawn from classical genetic methods such as epistasis analysis (e.g. 2). Arrows indicate the direction and sense in which gene products act on one another; arrows with heads indicate positive action, and arrows with flat heads indicate negative action. *Bottom panel* shows a portion of network of interactions among yeast RNA splicing proteins, the work of Fromont-Racine et al (32), which describes one of the most extensive two-hybrid mapping efforts to date. Arrows here indicate protein-protein interaction. Arrows have two heads; that is, two-hybrid data by themselves do not allow inferences about causality.

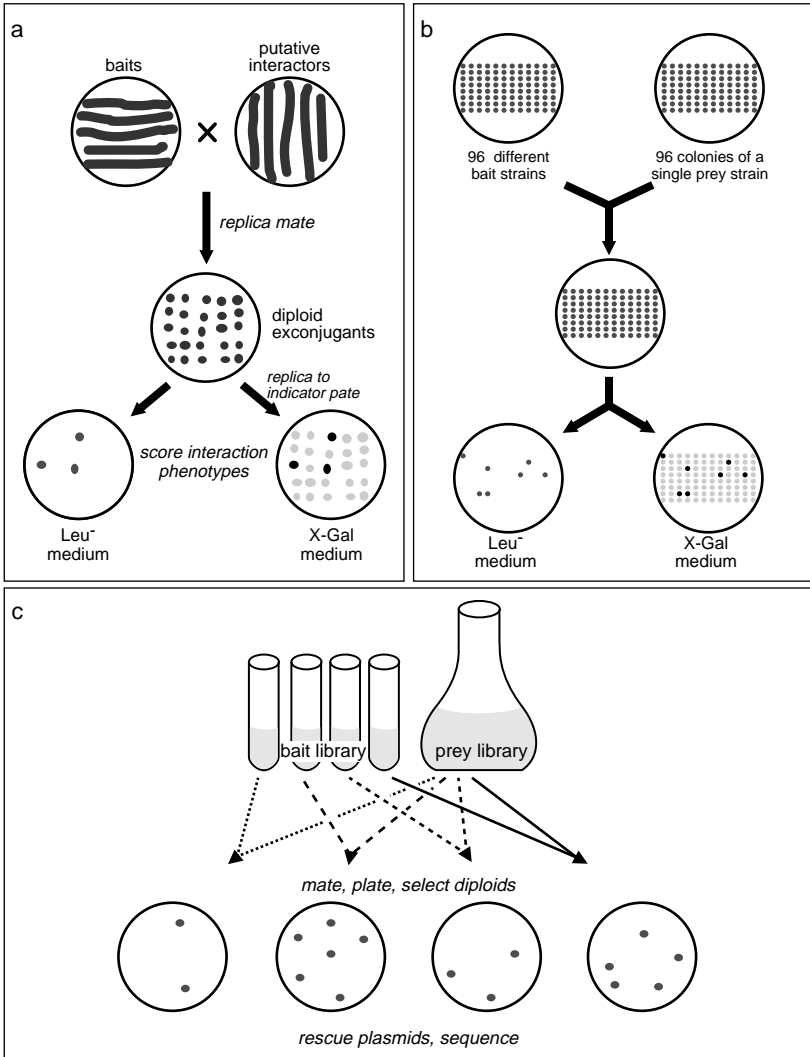
shortcomings are, however, offset by powerful advantages: Two-hybrid experiments can be done systematically, in large-scale, and with the gene products of organisms that wholly lack manipulative genetics.

Interaction Mating

Large-scale two-hybrid experiments rely on a fortunate property of yeast: Haploid yeast exist in two mating types that mate with one another and form diploids if they are physically juxtaposed. The fact that yeast mate affords a very simple way to introduce plasmids encoding potentially interacting proteins into the same nucleus: Simply, mate haploid strains of opposite mating types that contain the potential interactors and determine the transcription phenotypes of the reporters in the resulting diploid. This method has been used to simplify

conventional library screens (5; RL Finley Jr, unpublished). It has also been used to vastly increase the number of individual protein-protein interactions that can be assayed. The increase in assayable interactions afforded by interaction mating experiments makes possible the systematic mapping of interactions in pathways and whole genomes.

Two ways of performing interaction mating experiments have emerged (Figure 13). One (the arrayed library approach) is to mate arrayed cells that



contain known potentially interacting bait proteins against arrayed cells that contain potentially interacting preys. The other (“all against all”) is to mate libraries of cells that contain different baits with libraries of cells that contain preys, select cells in which an interaction is occurring, rescue the DNA encoding the interacting proteins, and determine their identity. We discuss these methods in turn.

Interaction mating with arrayed libraries is extremely simple. There are two variants of it (Figure 13). In one, stripes of bait strains are crossed with stripes of prey strains on rich plates. Patches of diploid cells form at the intersection of these stripes. Diploids are selected by replica plating on appropriate selective media and are then scored for interaction phenotypes (28). In another, a single bait or prey strain is mated, using a replica velvet or multiprong device, to gridded patches of strains of opposite mating type that contain potential partners. Diploids are again selected by plating on appropriate selective media and scored for interaction phenotypes. The entire procedure can be easily automated to facilitate very large-scale mapping of interactions (30).

Interaction mating of libraries against libraries is simpler still. In it, pools of haploid bait cells are mated against pools of haploid prey cells, and cells in which interaction occurs are selected by plating on medium that selects for reporter transcription (Figure 13). In one example of this, Bartel, Fields, and colleagues made libraries of such cells from the bacteriophage T7 genome (4). Because some baits activated transcription, they typically mated small pools of about ten baits against pools that contained the entire prey library. They recovered bait and prey plasmids from diploids in which the reporters were activated and sequenced them to determine interacting pairs. This catalog revealed many known and many unknown but informative interactions between T7 proteins. A more recent

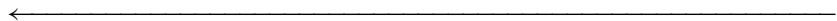


Figure 13 Interaction mating methods. All begin with preys and baits in strains of opposite mating types. *a.* An early array method (28). In it, baits are streaked horizontally, preys are streaked vertically, diploids form in the patches and are selected, and then transferred to plates where the interaction phenotypes are scored. *b.* A more advanced array method (30). In it, 96 strains that contain different baits or different preys are replica mated to 96 patches of cells from the same bait or prey strain. The resulting cells are then plated directly onto selection plates that allow transcription of the reporter genes to be scored. *c.* A “mate-and-sequence” method (4). In it, two pools of cells are transformed with libraries of bait and prey plasmids. The pool of bait plasmids is in turn divided into many smaller aliquots; since many baits activate transcription on their own, this division insures that there will be many individual pools in which the reporters are not adventitiously activated by an activating bait. Each aliquot of the bait pool is mated with the prey pool, and the resulting diploids plated on medium that selects for reporter activation. Bait and prey plasmids from cells that contain interacting plasmids are rescued and sequenced to determine the identity of the interacting proteins. Recently, Fromont-Racine et al (32) have described a higher-throughput variant of this technique in which mating takes place on filters.

study by Fromont-Racine et al (32) generated a similar catalog of interactions between yeast proteins involved in yeast mRNA splicing, which revealed a similar haul of known and previously unknown but suggestive interactions (11). In the future, use of counter-selectable reporter genes may allow depletion of the bait libraries of proteins that activate transcription, and thus potentially obviate the need for the sub-pooling strategy.

Future interaction mating schemes will likely combine arrayed and mate-and-sequence approaches. Currently, it is becoming easier to grid and array libraries and cells transformed with library plasmids, and it is likely that most collections of potentially interacting proteins either will be custom made or will be isolated from libraries and then arrayed. Once libraries are arrayed, whether the matings are done with individual library members against panels, with small numbers of library members against panels, or with the entirety of one library against another will be determined by the person who writes the program for the individual mating experiment, and the actual work will be done by laboratory robots.

Information from interaction mating experiments is a fertile source of testable ideas about gene function. In one example, Reymond and colleagues (69) used interaction mating to guide initial experiments to find function for the Rox/Mnt protein, a putative breast cancer tumor suppressor. Rox is a basic helix loop helix/leucine zipper protein (bHLH/Zip), and in these experiments Reymond mated Rox against a panel of other bHLH and leucine zipper proteins that positively and negatively regulate cell proliferation. These experiments showed that Rox interacted with itself and with Max, a protein that forms heterodimers with Myc to cause cancer. These results suggested a model for Rox function, in which Rox inhibits cancer by sequestering Max. Subsequent experiments in cultured cells have lent support to this model (69; A Reymond, R Brent, unpublished). Here, data from interaction mating supplied a working hypothesis, which the investigators were able to test quickly by conventional methods. The point is not that these data substituted for conventional work to test the ideas, but that they suggested which experiments to do.

Similarly, interaction mating can provide testable guesses about the function of allelic variants of normal proteins. For example, in another study, Reymond & Brent (78) used interaction mating, followed by analysis *in vitro*, to study protein interactions of allelic variants of the p16 tumor suppressor protein. p16 normally binds to cyclin-dependent kinase 4 (Cdk4) and cyclin-dependent kinase 6 (Cdk6), and suppresses their activity. Mutations in its coding sequence are found in families that are predisposed to melanomas and adenocarcinomas (93). Reymond & Brent showed that allelic variants of the p16 tumor suppressor found in cancer-prone pedigrees were deficient in their ability to interact with their targets, Cdk4 and Cdk6. That much was expected. Two results of the

work were not expected, however. One was the finding that one p16 variant, p16-G101W, interacted normally with Cdk6. In this case, the two-hybrid data clearly showed that the deleterious effect of this allele in the population is not due to a defective interaction with Cdk6, but more likely to a defective interaction with Cdk4. These data thus suggest that further investigation of the interaction of this allelic variant with Cdk4 may eventually impact public health, but that its interaction with Cdk6 need be considered no further.

The second unexpected result was that another p16 allele, p16-I49T, was also deficient in interaction with Cdk4. This variant is present in some cancer-prone pedigrees, but is also present in individuals not known to be associated with such pedigrees. The finding that p16-I49T is defective in interaction with Cdk4 and Cdk6 suggests that this allele may also predispose its carriers to increased risk for tumors. Here the test of the hypothesis that p16-I49T is associated with predisposition to cancer will need to be epidemiological, rather than molecular-biological, and will thus be more difficult.

In the above examples, observation of interactions between proteins whose biochemical function was at least somewhat defined allowed formulation of testable hypotheses. However, there are also cases in which knowledge of the function of the interacting proteins is not needed to make functional inferences. For example, we showed that individual binary protein-protein interactions could suggest the existence of ternary protein complexes (28). More complex patterns of binary interactions exist, and, particularly when conjoined with crude affinity data, are likely to allow the generation of important functional inferences. For example, in Figure 14*a*, the proteins may form a five-protein complex, and in 14*b*, protein Z may be a regulated protein kinase.

Whatever information can be gained from analysis of the patterns of protein interactions, integration of interaction data with other biological data clearly adds value to the quality and precision of the inferences that an investigator can formulate. In particular, addition of data about the time and place during development in which interacting proteins are expressed can quickly suggest whether a given interaction is likely to be significant to the process under study. Because interaction information will be stored in databases, as much of the other types of information that can be used with it to generate functional inferences (e.g. sequence, time, and place of expression) is already, the effort to systematize the use of this information to make testable hypotheses is likely to be more than an interesting epistemological problem. Rather, the need to systematize generation of inferences from such data is likely to present students of data integration with an important early test case.

In those cases where the function of proteins and allelic variants depends on their interactions with other proteins, two-hybrid experiments can cast light on their function. Since so many proteins exert their effects by interacting with

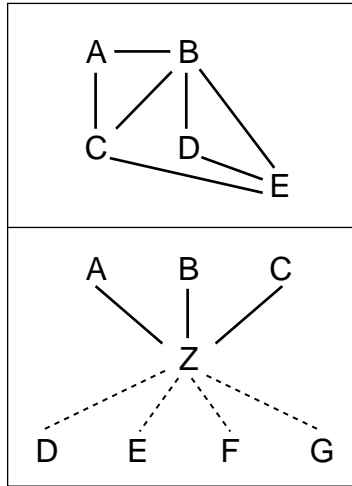


Figure 14 Functional inferences from complex patterns of binary protein interactions. In the *top panel*, only the A-B, A-C, B-D, D-E, E-C, and B-E interactions are sufficiently strong to be detected individually, but the data taken together clearly support the idea that all five proteins may form a complex. In the *bottom panel*, protein Z interacts tightly with the proteins above it and weakly with the proteins below it (indicated by dashed lines). This is the pattern of interaction expected from an enzyme that interacts tightly with proteins that regulate its activity, and weakly with its substrates, and might be expected if Z were, for example, a regulated protein kinase. In this case, D, E, F, and G may be identifiable as candidate substrates from interaction data alone.

other proteins, these methods are in principle able to provide useful information about a significant fraction of the genes of the genome. Given the significant polymorphism in human and other species, the existence of a relatively simple assay that can suggest functional differences between alleles suggests that interaction mating methods will augment current human genetic techniques in identifying disease-state genes (Figure 15). These will be particularly useful in identifying constellations of alleles that contribute to polygenic traits.

Testing Functional Inferences from Interaction Data

In the near term, it is easy to imagine that conventional two-hybrid technology will cease to be widely used to perform interactor hunts, as the need for such hunts diminishes due to the availability of genome-wide two-hybrid information from systematic application of interaction mating experiments. At that point, information needed to make functional inferences about proteins, and access to the plasmids encoding them, will be available to researchers at the cost of a few keystrokes. It is possible that even when such information is available, two-hybrid methods may still aid the assignment of function, by having enabled the selection of classes of molecules to aid function determination.

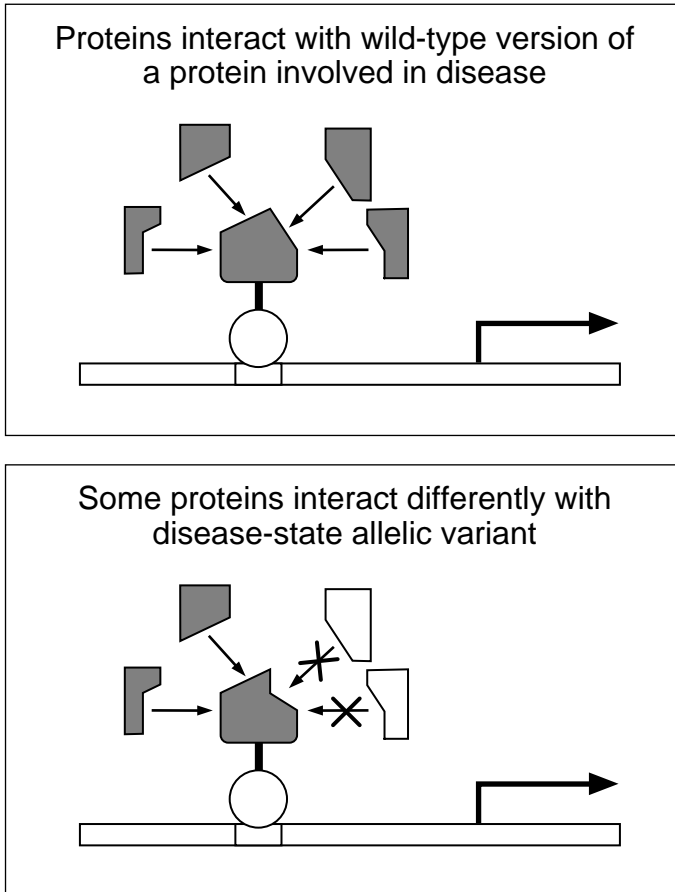


Figure 15 Identification of candidate proteins involved in disease states. The *top panel* shows proteins that interact with the wild-type allelic form of a protein involved in a disease. The *bottom panel* shows that some of these interactors interact differently with a disease-state allelic variant. These may be proteins that participate in the same biochemical function altered in the disease-state allelic form.

The idea that breaking interactions can be biologically informative is a natural corollary of the idea that detecting interactions can be informative. Viewed in the context of existing doctrine for systems with developed genetics (phage, bacteria, yeast, worms, flies), this idea is hardly new, since interruption of a specific protein interaction is an expected consequence of many missense loss-of-function mutations. However, for systems without well-developed genetics, any means to interrupt specific interactions, and thus to compensate for the

difficulty in isolating or manufacturing informative sets of missense mutations, would be useful. Moreover, if such means also allowed the investigator to bring about complete loss of gene function, they might constitute a supplemental means of genetic analysis, even in systems that now allow generation of gene knockouts.

Further, there are precedents for the idea of using artificial agents to break interactions to achieve heuristic or therapeutic effects. For example, injection of anti-Ras antibodies into cells transformed with mutant EGF receptors (EGFR) provided evidence that Ras lay downstream of them in a signaling cascade (71), and injection of peptides that mimic the phosphorylation site of EGFR and block the ability of that portion of EGFR to interact with Grb2 helped lay the groundwork for detection of this interaction (79). Perhaps more significantly, there are numerous examples of small organic molecules that interrupt protein interactions and cause significant biological effects. For example, interruption of beta-tubulin dimerization with benomyl is cytostatic for yeast (72), and interruption of the interaction of HIV-1 protease with its substrate blocks viral replication (67).

The above arguments and results suggested that molecules that break specific protein interactions might be useful genetic reagents in higher systems. One possible route to this goal, the intracellular synthesis of antibody derivatives against target molecules, has on occasion produced loss of function of the target, although loss of function in each case might be due to the complete inactivation and sequestration of the target, rather than abrogation of individual interactions (16).

GENERATION OF REAGENTS TO TEST FUNCTIONAL INFERENCES The advent of combinatorial technologies and ideologies has opened a way to rapid generation of smaller molecules that might break specific interactions. The idea here is that it might be possible, from libraries of nucleic acids or peptides, to select individual molecules that recognize one or the other surface of a target molecule, and block protein-protein interactions involving that surface. In recent experiments of this type, Ellington and coworkers selected aptameric RNAs that interact with the Rev protein of HIV-1, and demonstrated that some of these molecules interfere with the interactions between Rev-1 and Rip, a protein necessary for Rev function (40, 94). As expected, intracellular synthesis of these RNAs results in loss of the ability to propagate HIV, presumably due to loss of the Rev-1/Rip interaction.

Work from two groups has demonstrated that two-hybrid methods can be used to isolate peptides that bind specific proteins, and in one case such peptides have been shown to break interactions. Yang et al (109) made a library that expressed random 16-amino acid peptides fused to the Gal4 activation region

and screened it for peptides that interacted with a human retinoblastoma gene product, pRb, bait. They isolated 7 pRb interacting peptides out of 3×10^6 yeast transformed with the library, presumably representing 3×10^6 different peptides. Surprisingly, all 7 pRb interacting peptides contained a Leu-X-Cys-X-Glu sequence that is also conserved in natural pRb binding proteins like SV40 Large T, adenovirus E1A, human papilloma virus E7, and cyclin D (50). Evanescent wave binding experiments with a surface plasmon resonance instrument revealed that the proteins bound unexpectedly tightly, with K_{dS} on the order of $10 \mu\text{M}$ under the conditions used. These experiments demonstrated that, along with phage display and other combinatorial peptide methods (87), two-hybrid methods can be used to select unconstrained peptides that define consensus binding sequences found in nature.

Somewhat different experiments were performed by Colas et al (21). Here, these workers started from a different perspective: that antibodies could recognize most combinations of shape, charge, and hydrophobicity, and that they did so by displaying conformationally constrained peptide loops of variable sequence. They wished to isolate synthetic peptide agents that could bind to most faces of cellular proteins and specifically disrupt particular cellular protein-protein interactions. To this end, these workers used two-hybrid methods to isolate, from a library of conformationally constrained 20-mers displayed by *E. coli* thioredoxin, variable region sequences that recognized cyclin-dependent kinases. The workers referred to these thioredoxin-variable region chimeras as peptide aptamers. Some aptamers cross-reacted with cyclin-dependent kinases of related sequence, indicating that these proteins recognized conserved antigenic regions (epitopes) on the related proteins; moreover, aptamers that did cross-react interacted with different subsets of kinases, indicating that the epitopes recognized by these cross-reacting aptamers were distinct (85). As measured by evanescent wave experiments, binding was strong, with K_{dS} in the nanomolar range. These results showed that peptide aptamers could be constructed that were somewhat like antibodies, in that members of collections of them would recognize many different protein surfaces. In fact, it has proven possible to isolate peptide aptamers against all protein targets tested (HP Xu et al, unpublished; B Cohen et al, unpublished; M Kolonin & RL Finley Jr, unpublished). However, unlike antibodies, peptide aptamers are designed to work inside cells, providing reagents to probe protein function in vivo.

Some of the peptide aptamers isolated by Colas et al disrupted interactions with other proteins. Six peptide aptamers inhibited Cdk2 activity, blocking its ability to phosphorylate a model substrate, histone H1 (85). Inhibition was competitive, and was, in some cases, specific to the histone H1 substrate. This fact indicates that the aptamers inhibit by binding to the kinase in or near its active site, and blocking its interaction with the substrate. Moreover, it

demonstrates that some of the aptamers were selective in the protein interactions they blocked. Although not tested, other Cdk2 aptamers would be expected to inhibit interactions with other Cdk2 partner proteins.

As many workers have noted, the fact that two-hybrid selections themselves provide a genetic handle on individual interactions opens in principle the idea that small peptides could be directly selected from combinatorial libraries to break an individual interaction, turn off transcription of a counter-selectable reporter, and allow a yeast that bears them to form a colony on a selective medium (21, 68, 85, 100) (Figure 16a). As of this writing, the promise of this approach has not been realized, and counter-selectable reporters have not been used to directly identify aptamers that block particular protein interactions. Rather (Figure 16b), aptamers were first selected to bind a target, and then screened for their ability to block protein-protein interactions. This approach works well (P Colas, unpublished; HP Xu unpublished), and thus might seem to obviate the need for counter-selectable reporters. However, such reporters may be useful in another context; they may smooth the path to one-step identification of aptamers that interact specifically with one allelic form of a protein but not with another, rather than requiring the investigator to use the current two-step methods.

A number of pharmaceutical and biotechnology companies are employing two-hybrid methods to search for small molecules that interrupt particular protein interactions. Adoption of these techniques to screen for drugs required belief in the tractability of two problems. One is the fact that yeast is impermeable to most organic molecules, thus requiring researchers either to resign themselves to failing to detect otherwise active molecules or to resort to permeability enhancing agents, such as Polymyxin B nonapeptide (6) or yeast host mutations, such as *tmp* (10). The other is more significant: Experience in the pharmaceutical industry has suggested that most useful inhibitory molecules bind to the active sites of enzymes. This fact suggests that these screens would be most likely to work on protein-enzyme interactions that are weak enough to be disrupted by molecules with 10 nM^{-1} μM dissociation constants, but tight enough to be above the detection threshold of the reporters used. Within this narrow range, it is expected that inhibitors might be detected by two-hybrid methods, but, given the paucity of published data on the subject, it is difficult to make a fair comparison of the advantages of yeast-based two-hybrid methods compared with mammalian two-hybrid methods, in vitro two-hybrid transcription methods, and other in vitro screens such as fluorescent proximity assays and evanescent wave (surface plasmon resonance) techniques.

In the future, it is possible that advances in peptidomimetic chemistry will allow an alternative to screening for small molecules that break protein interactions. In this approach, quick determination of the structures of peptide aptamer

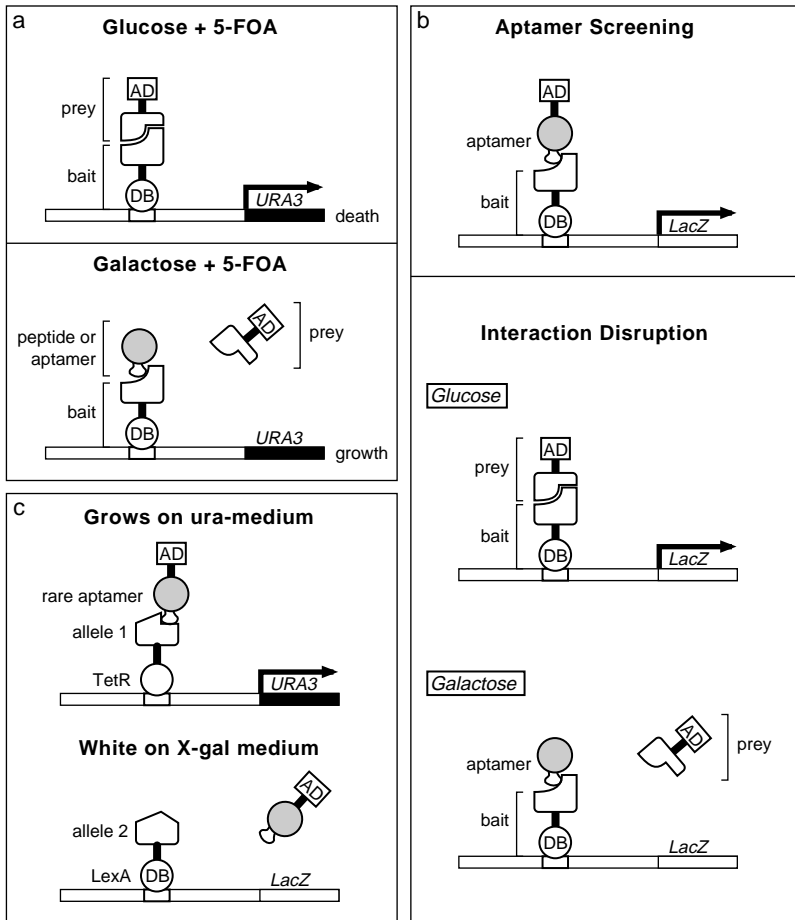


Figure 16 Isolation of aptamers or other peptides that disrupt protein-protein interactions. *a*. Still-hypothetical one-step isolation of peptides or aptamers that disrupt a protein-protein interaction. Here, a counterselectable *URA3* reporter allows direct selection of an aptamer that allows the cell to grow on 5-FOA; *b*. two-step isolation of disruptive aptamers. In this technique, peptide aptamers that bind a target are identified from an activation tagged library. The aptamers are then expressed conditionally on galactose medium, without activation domains, in a cell that contains a bait that interacts with an activation-tagged protein of interest. Some of the aptamers that bind the bait diminish the protein-protein interaction (P Colas, unpublished; HP Xu, unpublished). *c*. Selection and screen method to isolate rare discriminatory aptamers that interact with one allelic variant of a protein, but not another (HP Xu & R Brent, unpublished). In these two-bait cells, cells grow on ura-medium due to interaction of the activation tagged effector with allelic variant 1, but are white on Xgal due to the fact that this aptamer does not interact with allelic variant 2. Use of *URA3* as a counterselectable marker with a selectable marker other than *lacZ* might enable selection of such discriminatory aptamers in one step.

variable regions bound to their targets may provide sufficient information to allow synthesis of non-peptide interaction disrupting molecules.

USE OF REAGENTS TO TEST FUNCTIONAL INFERENCES As mentioned above, these aptamer-based approaches were not created to provide a way into the discovery of new drugs, but to provide a means of testing inferences about gene and allele function. At the time of writing, there have been no such published tests that use targeted disruption of interactions by aptamers to find new facts about gene function; rather, all relevant experiments are pilot experiments. For example, as mentioned above, Ellington and coworkers (40) isolated anti-Rev RNA aptamers, which they then expressed in cultured cells to inhibit HIV-1 gene expression. In another test of this doctrine, Kolonin et al (M Kolonin & RL Finley Jr, unpublished) have recently expressed in *Drosophila melanogaster* peptide aptamers that react with *Drosophila* Cdc2 and disrupt its interaction with its substrate. Cdc2 is necessary for cell divisions of the developing *Drosophila* eye, and expression of the aptamers, but not control peptides, resulted in improperly developed rough eyes. These results provide confirmation that ectopic expression of aptamers can disrupt protein interactions in a transgenic organism, and that the resulting phenotypes can directly demonstrate the function of the disrupted interaction.

Aptamers may also be able to perform functions beyond the interruption of protein interactions. For example, Colas et al have used aptamers as targeting domains to bring ubiquitin conjugating domains to an intracellular protein target (P Colas, R Brent, unpublished). This targeted ubiquitination could be used to inactivate specific proteins. Similarly, it may be possible to select peptide aptamers, or to construct bivalent aptamers, that bridge two proteins and increase their likelihood of interaction, and even to select peptide aptamers that heal the deleterious effects of hypomorphic mutations such as ts alleles. Because such reagents should allow more sophisticated perturbations of genetic networks than disruption of connections, they will broaden the number and kind of experiments that test functional inferences from the maps of connections generated by two-hybrid data.

Expected Future Developments in Functional Genomics

Application of interaction mating methods to the function of the wild-type and allelic variant genes of entire genomes will undoubtedly require mechanization. For this to occur, a number of issues may need to be resolved. First, it is not yet clear whether ordered mating approaches, “all-against-all” mate-and-sequence approaches, or mixed approaches will prove the most effective for large-scale data collection. Second, while it is apparent that any ordered or partially ordered mating approach will require large numbers of individual yeast matings, it is not yet clear whether these will be performed on plates, in wells

of microtiter dishes, in capillary tubes, or even in microdroplets. Third, given the limits to the protein interactions that can be detected by conventional two-hybrid methods, it is not clear how much supplemental data will be gathered by other genetic methods, and how much will eventually be gathered by scaled-up direct physical methods.

The nature of the computational work that will make use of the connection data is also unclear. As described above, we can anticipate some progress in automated detection of patterns of connections to identify potentially interesting proteins, and we can anticipate the development of sophisticated programs that take into account the degree of certainty of an interacting protein's function in making their predictions for the function of test proteins. Whether other possible lines of inquiry, from improved use of data visualization to correlate connection information with other kinds of genomic data, or algorithms to spot co-variation in connection patterns caused by co-variation in the states of more than one allele, will prove fruitful, remains to be seen.

CONCLUSION: A NEW CONTINENT OF GENETIC INFORMATION

The next decade will provide increasingly ramified maps of protein connections, perhaps eventually comprising a single map of all possible protein interactions. Such a map will depict a new continent of genetic information. Like the information derived from sequences of genomes, and from inventories of transcripts present in particular cell types, this information will clearly be useful. However, it is not yet clear how this continent will be explored, and how this map will be integrated with other sorts of information by working scientists.

Moreover, we can already recognize that neither inventories of cellular molecules, nor descriptions of their connections, give us anything other than a static picture of the relationship of cellular components. It will be necessary for us to determine relationships among cellular components, but our determination to catalog this cellular anatomy should not blind us to the need for the next step, which is an understanding of how the components work together, a cellular physiology. At the moment, intellectual and technical frameworks that would allow us to proceed from anatomical to physiological biology are utterly lacking, and the quest to create them will likely occupy the attention of biologists for decades to come.

ACKNOWLEDGMENTS

The authors thank Barak Cohen and Pierre Colas for helpful comments on the manuscript, and Pierre Colas, Andy Mendelsohn, Aleister Saunders, and Julia Khorana for help with the figures.

Visit the Annual Reviews home page at
<http://www.annurev.org>

Literature Cited

1. Ausubel FM, Brent R, Kingston RE, Moore D, Seidman JG, Struhl K, eds. 1997. *Current Protocols in Molecular Biology*, 1987–1997. New York: Greene & Wiley
2. Avery L, Wasserman S. 1992. Ordering gene function: the interpretation of epistasis in regulatory hierarchies. *Trends Genet.* 8(9):312–16
3. Bartel PL, Fields S. 1995. Analyzing protein-protein interactions using two-hybrid system. *Methods Enzymol.* 254:241–63
4. Bartel PL, Roeklein JA, SenGupta D, Fields S. 1996. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* 12(1):72–77
5. Bendixen C, Gangloff S, Rothstein R. 1994. A yeast mating-selection scheme for detection of protein-protein interactions. *Nucleic Acids Res.* 22(9):1778–79
6. Boguslawski G. 1985. Effects of polymyxin B sulfate and polymyxin B nonapeptide on growth and permeability of the yeast *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 199(3):401–5
7. Botstein D, Maurer R. 1982. Genetic approaches to the analysis of microbial development. *Annu. Rev. Genet.* 16:61–83
8. Bram RJ, Kornberg RD. 1985. Specific binding to far upstream activating sequences in polymerase II promoters. *Proc. Natl. Acad. Sci. USA* 82:43–47
9. Breitwieser W, Markussen FH, Horstmann H, Ephrussi A. 1996. Oskar protein interaction with Vasa represents an essential step in polar granule assembly. *Genes Dev.* 10(17):2179–88
10. Brendel M. 1976. A simple method for the isolation and characterization of thymidylate uptaking mutants in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* 147(2):209–15
11. Brent R. 1997. Looms to weave genomic nets. *Nat. Genet.* 16:216–17
12. Brent R, Mendelsohn A, Jessen T, Finley R, Cohen B, et al. 1997. <http://xanadu.mgh.harvard.edu/brentlabhomepage4.html>. Web site
13. Brent R, Ptashne M. 1984. A bacterial repressor protein or a yeast transcriptional terminator can block upstream activation of a yeast gene. *Nature* 312:612–15
14. Brent R, Ptashne M. 1985. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 43:729–36
15. Chang EC, Barr M, Wang Y, Jung V, Xu HP, Wigler MH. 1994. Cooperative interaction of *S. pombe* proteins required for mating and morphogenesis. *Cell* 79(1):131–41
16. Chen SY, Bagley J, Marasco WA. 1994. Intracellular antibodies as a new class of therapeutic molecules for gene therapy. *Hum. Gene Ther.* 5(5):595–601
17. Cherry JM, Ball C, Weng S, Juvik G, Schmidt R, et al. 1997. Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* 387(Suppl. 6632):67–73
18. Chien C-T, Bartel PL, Sternglanz R, Fields S. 1991. 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* 88:9578–82
19. Chiu MI, Katz H, Berlin V. 1994. RAPT1, a mammalian homolog of yeast Tor, interacts with the FKBP12/rapamycin complex. *Proc. Natl. Acad. Sci. USA* 91(26):12574–78
20. Choi KY, Satterberg B, Lyons DM, Elion E A. 1994. Ste5 tethers multiple protein kinases in the MAP kinase cascade required for mating in *S. cerevisiae*. *Cell* 78(3):499–512
21. Colas P, Cohen B, Jessen T, Grishina I, McCoy J, Brent R. 1996. Genetic selection of peptide aptamers that recognize and inhibit cyclin-dependent kinase 2. *Nature* 380:548–50
22. Dalton S, Treisman R. 1992. Characterization of Sap-1, a protein recruited by serum response factor to the c-fos serum response element. *Cell* 68:597–612
23. Dang CV, Barrett J, Villa-Garcia M, Resar LM, Kato GJ, Fearon E R. 1991. Intracellular leucine zipper interactions suggest c-Myc hetero-oligomerization. *Mol. Cell. Biol.* 11(2):954–62
24. Durfee T, Becherer K, Chen PL, Yeh SH, Yang Y, et al. 1993. The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. *Genes Dev.* 7(4):555–69
25. Estojak J, Brent R, Golemis EA. 1995.

- Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell Biol.* 15(10):5820–29
26. Fearon ER, Finkel T, Gillison ML, Kennedy SP, Casella JF, et al. 1992. Karyoplasmic interaction selection strategy: a general strategy to detect protein-protein interactions in mammalian cells. *Proc. Natl. Acad. Sci. USA* 89(17):7958–62
 27. Fields S, Song O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* 340:245–46
 28. Finley RL Jr, Brent R. 1994. Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci. USA* 91(26):12980–84
 29. Finley RL Jr, Brent R. 1995. Interaction trap cloning with yeast. In *DNA Cloning, Expression Systems: A Practical Approach*, ed. BD Hames, DM Glover, pp.169–203. Oxford: Oxford Univ. Press
 30. Finley RL Jr, Brent R. 1996. Two-hybrid analysis of genetic regulatory networks. In *The Yeast Two-Hybrid System*, ed. PL Bartel, S Fields, pp. 197–214. Oxford: Oxford Univ. Press
 31. Finley RL Jr, Kolonin M, Fonfara J, Bolin M. 1997. <http://cmmg.biosci.wayne.edu/rfinley/finlab/finlab-home.html>. Web site
 32. Fromont-Racine M, Rain J-C, Legrain P. 1997. Towards a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* 16:277–81
 33. Gaul U, Chang H, Choi T, Karim F, Rubin GM. 1993. Identification of ras targets using a genetic approach. *Ciba Found. Symp.* 176:85–92; discussion pp. 92–95
 34. Gill G, Ptashne M. 1988. Negative effect of the transcriptional activator GAL4. *Nature* 334:721–24
 - 34a. Gimeno RE, Espenshade P, Kaiser CA. 1996. COPII coat subunit interactions: Sec24p and Sec23p bind to adjacent sites of Sec16p. *Mol. Biol. Cell* 7:1815–23
 35. Giniger E, Varnum SM, Ptashne M. 1985. Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* 40:767–74
 36. Goffeau A, Barrell BG, Bussey H, Davis RW, Dujon B, et al. 1996. Life with 6000 genes. *Science* 274:546, 563–67
 37. Goffeau A, Barrell B, Bussey H, Davis R, Murakami Y. 1997. The complete DNA sequence of *S. cerevisiae*. (<http://genome.www.stanford.edu/Saccharomyces/sequence/done.html>) Web site
 38. Golemis E. 1997. Golemis Lab Web Site. <http://www.fccc.edu/research/labs/golemis/InteractionTrap1nWork.html>. Web site
 39. Golemis EA, Brent R. 1992. Fused protein domains inhibit DNA binding by LexA. *Mol. Cell Biol.* 12(7):3006–14
 40. Good PD, Krikos AJ, Li SX, Bertrand E, Lee NS, et al. 1997. Expression of small, therapeutic RNAs in human cell nuclei. *Gene Ther.* 4(1):45–54
 41. Gyuris J, Golemis E, Chertkov H, Brent R. 1993. Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. *Cell* 75(4):791–803
 42. Hanes SD, Brent R. 1989. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* 57(7):1275–83
 43. Hardy CF, Sussel L, Shore D. 1992. A RAP1-interacting protein involved in transcriptional silencing and telomere length regulation. *Genes Dev.* 6(5):801–14
 44. Harper JW, Adami GR, Wei N, Keyomarsi K, Elledge SJ. 1993. The p21 cdk-interacting protein cip1 is a potent inhibitor of g1 cyclin-dependent kinases. *Cell* 75(4):805–16
 45. Hartman PE, Roth JR. 1973. Mechanisms of suppression. *Adv. Genet.* 17:1–105
 46. Hartwell LH. 1974. *Saccharomyces cerevisiae* cell cycle. *Bacteriol. Rev.* 38(2):164–98
 47. Hartwell LH, Weinert TA. 1989. Checkpoints: controls that ensure the order of cell cycle events. *Science* 246:629–34
 48. Hengartner MO, Horvitz HR. 1994. Programmed cell death in *Caenorhabditis elegans*. *Curr. Opin. Genet. Dev.* 4(4):581–86
 49. Hereford LM, Hartwell LH. 1974. Sequential gene function in the initiation of *Saccharomyces cerevisiae* DNA synthesis. *J. Mol. Biol.* 84(3):445–61
 50. Hinds PW, Weinberg RA. 1994. Tumor suppressor genes. *Curr. Opin. Genet. Dev.* 4(1):135–41
 51. Hope IA, Struhl K. 1985. GCN4 protein, synthesized in vitro, binds HIS3 regulatory sequences: implications for general control of amino acid biosynthetic genes in yeast. *Cell* 43(1):177–88
 - 51a. Hope IA, Struhl K. 1986. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* 46(6):885–94
 52. Horvitz HR, Shaham S, Hengartner MO. 1994. The genetics of programmed cell death in the nematode *Caenorhabditis elegans* *Cold Spring Harbor Symp. Quant. Biol.* 59:377–85

53. Huffaker TC, Hoyt MA, Botstein D. 1987. Genetic analysis of the yeast cytoskeleton. *Annu. Rev. Genet.* 21:259–84
54. Inouye C, Remondelli P, Karin M, Elledge S. 1994. Isolation of a cDNA encoding a metal response element binding protein using a novel expression cloning procedure: the one hybrid system. *DNA Cell Biol.* 13(7):731–42
55. James P, Halladay J, Craig EA. 1996. Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* 144(4):1425–36
56. Jarvik J, Botstein D. 1973. A genetic method for determining the order of events in a biological pathway. *Proc. Natl. Acad. Sci. USA* 70(7):2046–50
- 56a. Jiang R, Carlson M. 1996. Glucose regulates protein interactions within the yeast SNF1 protein kinase complex. *Genes Dev.* 10:3105–15
57. Keegan L, Gill G, Ptashne M. 1986. Separation of the DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* 231:699–704
58. Lech K, Anderson K, Brent R. 1988. DNA-bound Fos proteins activate transcription in yeast. *Cell* 52(2):179–84
59. Le Douarin B, Pierrat B, Baur EV, Chambon P, Losson R. 1995. A new version of the two-hybrid assay for detection of protein-protein interactions. *Nucleic Acids Res.* 23(5):876–78
60. Lee JW, Ryan F, Swaffield JC, Johnston SA, Moore DD. 1995. Interaction of thyroid-hormone receptor with a conserved transcriptional mediator. *Nature* 374:91–94
61. Li JJ, Herskowitz I. 1993. Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. *Science* 262:1870–74
62. Licitra EJ, Liu JO. 1996. A three-hybrid system for detecting small ligand-protein receptor interactions. *Proc. Natl. Acad. Sci. USA* 93:12817–21
63. Ma J, Ptashne M. 1987. A new class of transcriptional activators. *Cell* 51:113–19
64. Ma J, Ptashne M. 1988. Converting a eukaryotic transcriptional inhibitor into an activator. *Cell* 55:443–46
65. Ma J, Ptashne M. 1987. Deletion analysis of GAL4 defines two transcriptional activating segments. *Cell* 48(5):847–53
66. Marcus S, Polverino A, Barr M, Wigler M. 1994. Complexes between STE5 and components of the pheromone-responsive mitogen-activated protein kinase module. *Proc. Natl. Acad. Sci. USA* 91:7762–66
67. McQuade TJ, Tomasselli AG, Liu L, Karacostas V, Moss B, et al. 1990. A synthetic HIV-1 protease inhibitor with antiviral activity arrests HIV-like particle maturation. *Science* 247:454–56
68. Mendelsohn AR, Brent R. 1994. Applications of interaction traps/two-hybrid systems to biotechnology research. *Curr. Opin. Biotechnol.* 5(5):482–86
69. Meroni G, Reymond A, Alcalay M, Borzani G, Tanigami A, et al. 1997. Rox, a novel bHLHZip protein expressed in quiescent cells that heterodimerizes with Max, binds a non-canonical E box and acts as a transcriptional repressor. *EMBO J.* 16(10):2892–906
70. Miklos GL, Rubin GM. 1996. The role of the genome project in determining gene function: insights from model organisms. *Cell* 86(4):521–29
71. Mulcahy LS, Smith MR, Stacey DW. 1985. Requirement for ras proto-oncogene function during serum-stimulated growth of NIH 3T3 cells. *Nature* 313:241–43
72. Neff NF, Thomas JH, Grisafi P, Botstein D. 1983. Isolation of the beta-tubulin gene from yeast and demonstration of its essential function in vivo. *Cell* 33(1):211–19
73. Osborne MA, Zenner G, Lubinus M, Zhang X, Songyang Z, et al. 1996. The inositol 5'-phosphatase SHIP binds to immunoreceptor signaling motifs and responds to high affinity IgE receptor aggregation. *J. Biol. Chem.* 271(46):29271–78
74. Ozenberger BA, Young KH. 1995. Functional interaction of ligands and receptors of the hematopoietic superfamily in yeast. *Mol. Endocrinol.* 9(10):1321–29
75. Printen JA, Sprague GF Jr. 1994. Protein-protein interactions in the yeast pheromone response pathway: Ste5p interacts with all members of the MAP kinase cascade. *Genetics* 138(3):609–19
76. Ptashne M, Gann A. 1997. Transcriptional activation by recruitment. *Nature* 386(6625):569–77
77. Reed JC. 1997. Double identity for proteins of the Bcl-2 family. *Nature* 387(6635):773–76
78. Reymond A, Brent R. 1995. p16 Proteins from melanoma-prone families are deficient in binding to Cdk4. *Oncogene* 11(6):1173–78
79. Rojas M, Yao S, Lin YZ. 1996. Controlling epidermal growth factor (EGF)-stimulated Ras activation in intact cells by a cell-permeable peptide mimicking phosphorylated EGF receptor. *J. Biol. Chem.* 271(44):27456–61

80. Ross-Macdonald P, Sheehan A, Roeder GS, Snyder M. 1997. A multipurpose transposon system for analyzing protein production, localization, and function in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 94(1):190-95
81. Rubin GM. 1991. Signal transduction and the fate of the R7 photoreceptor in *Drosophila*. *Trends Genet.* 7(11-12): 372-77
82. Schena M, Shalon D, Davis RW, Brown PO. 1995. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270(5235):467-70
83. Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis R W. 1996. Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc. Natl. Acad. Sci. USA* 93(20):10614-19
84. SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. 1996. A three-hybrid system to detect RNA-protein interactions in vivo. *Proc. Natl. Acad. Sci. USA* 93(16):8496-501
85. Shih HM, Goldman PS, DeMaggio AJ, Hollenberg SM, Goodman, RH, Hoekstra M F. 1996. A positive genetic selection for disrupting protein-protein interactions: identification of CREB mutations that prevent association with the coactivator CBP. *Proc. Natl. Acad. Sci. USA* 93(24):13896-901
86. Silver PA, Brent R, Ptashne M. 1986. DNA binding is not sufficient for nuclear localization of regulatory proteins in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 6(12):4763-66
87. Smith GP, Scott JK. 1993. Libraries of peptides and proteins displayed on filamentous phage. *Methods Enzymol.* 217:228-57
88. Stern S, Tanaka M, Herr W. 1989. The Oct-1 homeodomain directs formation of a multiprotein-DNA complex with the HSV transactivator VP16. *Nature* 341(6243):624-30
89. Sternberg PW, Horvitz HR. 1991. Signal transduction during *C. elegans* vulval induction. *Trends Genet.* 7(11-12):366-71
90. Sternberg PW, Horvitz HR. 1984. The genetic control of cell lineage during nematode development. *Annu. Rev. Genet.* 18:489-524
91. Strathern JN, Jones EW, Broach JR, eds. 1981. *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*. Cold Spring Harbor, NY: Cold Spring Harbor Press
92. Takacs AM, Das T, Banerjee AK. 1993. Mapping of interacting domains between the nucleocapsid protein and the phosphoprotein of vesicular stomatitis virus by using a two-hybrid system. *Proc. Natl. Acad. Sci. USA* 90(21):10375-79
93. Tam SW, Shay JW, Pagano M. 1994. Differential expression and cell cycle regulation of the cyclin-dependent kinase 4 inhibitor p16Ink4. *Cancer Res.* 54(22):5816-20
94. Tian Y, Adya N, Wagner S, Giam CZ, Green MR, Ellington, A D. 1995. Dissecting protein:protein interactions between transcription factors with an RNA aptamer. *RNA* 1(3):317-26
95. Triezenberg SJ, Kingsbury RC, McKnight SL. 1988. Functional dissection of VP16, the trans-activator of herpes simplex virus immediate early gene expression. *Genes Dev.* 2(6):718-29
96. Triezenberg SJ, LaMarco KL, McKnight SL. 1988. Evidence of DNA: protein interactions that mediate HSV-1 immediate early gene activation by VP16. *Genes Dev.* 2(6):730-42
97. Van Aelst L, Barr M, Marcus S, Polverino A, Wigler M. 1993. Complex formation between RAS and RAF and other protein kinases. *Proc. Natl. Acad. Sci. USA* 90:6213-17
98. Vasavada HA, Ganguly S, Germino FJ, Wang ZX, Weissman S M. 1991. A contingent replication assay for the detection of protein-protein interactions in animal cells. *Proc. Natl. Acad. Sci. USA* 88(23):10686-90
99. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW. 1995. Serial analysis of gene expression. *Science* 270(5235):484-87
100. Vidal M, Brachmann RK, Fattaey A, Harlow E, Boeke JD. 1996. Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* 93(19):10315-20
101. Vojtek AB, Hollenberg SM, Cooper JA. 1993. Mammalian Ras interacts directly with the serine/threonine kinase Raf. *Cell* 74(1):205-14
102. Wang T, Donahoe PK, Zervos AS. 1994. Specific interaction of type I receptors of the TGF-beta family with the immunophilin FKBP-12. *Science* 265(5172):674-76
103. Wang ZF, Whitfield ML, Ingledue TC, Dominski Z, Marzluff W F. 1996. The protein that binds the 3' end of histone mRNA: a novel RNA-binding protein required for histone pre-mRNA processing. *Genes Dev.* 10(23):3028-40

104. Weinstock KG, Kirkness EF, Lee NH, Earle-Hughes JA, Venter JC. 1994. cDNA sequencing: a means of understanding cellular physiology. *Curr. Opin. Biotechnol.* 5(6):599-603
105. West RW Jr, Yocum RR, Ptashne M. 1984. *Saccharomyces cerevisiae* GAL1-GAL10 divergent promoter region: Location and function of the upstream activating sequences UAS_G. *Mol. Cell. Biol.* 4(11):2467-78
106. White MA. 1996. The yeast two-hybrid system: forward and reverse. *Proc. Natl. Acad. Sci. USA* 93(19):10001-3
107. Wilson TE, Padgett KA, Johnston M, Milbrandt J. 1993. A genetic method for defining DNA-binding domains: application to the nuclear receptor NGFI-B. *Proc. Natl. Acad. Sci. USA* 90(19):9186-90
108. Xu CW, Mendelsohn A, Brent R. 1997. Cells that register logical relationships among proteins. Submitted
109. Yang M, Wu Z, Fields S. 1995. Protein-peptide interactions analyzed with the yeast two-hybrid system. *Nucleic Acids Res.* 23(7):1152-56
110. Yocum RR, Hanley S, West RJ, Ptashne M. 1984. Use of lacZ fusions to delimit regulatory elements of the inducible GAL1-GAL10 promoter in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 4:1985-88
111. Zervos AS, Gyuris J, Brent R. 1993. Mxi1, a protein that specifically interacts with Max to bind Myc-Max recognition sites. *Cell* 72(2):223-32