

A Strategy for Constructing Large Protein Interaction Maps Using the Yeast Two-Hybrid System: Regulated Expression Arrays and Two-Phase Mating

Jinhui Zhong,¹ Huamei Zhang,¹ Clement A. Stanyon,¹ Gerard Tromp,¹ and Russell L. Finley Jr.^{1,2,3}

¹Center for Molecular Medicine and Genetics and ²Department of Biochemistry and Molecular Biology, Wayne State University School of Medicine, Detroit, Michigan 48201, USA

Maps representing the binary interactions among proteins have become valuable tools for understanding how proteins work together to mediate biological processes. One of the most effective methods for detecting biologically important protein interactions has been the yeast two-hybrid system. Here we present an efficient two-hybrid strategy to facilitate construction of protein interaction maps on a genome-wide scale. The strategy begins with two arrays of yeast expressing known proteins fused to either a DNA binding domain (BD), or a transcription activation domain (AD). The fusion proteins are conditionally expressed using regulated promoters that can be repressed during construction and amplification of the yeast arrays. Interaction assays are conducted in two phases. In the first phase, small pools of AD strains are mated with the array of BD strains. In the second phase, individual BD strains are mated with appropriate subsets of the AD array corresponding to positive pools in the first phase. This strategy has several advantages over previously described approaches, including the ability to detect interactions with proteins that inhibit yeast growth or that activate transcription as BD fusions. Moreover, by minimizing the number of mating operations and sequencing reactions needed to test large sets of binary interactions, this strategy is more efficient than either matrix or library screening approaches. We also present a three-dimensional pooling scheme to further increase the efficiency of large-scale two-hybrid analyses.

[Supplemental material is available online at www.genome.org and proteome.wayne.edu.]

Methods to detect biologically relevant protein interactions are essential for understanding biological systems and the functions of individual proteins. Protein-protein interactions participate in almost all biological processes. Many important regulatory pathways, for example, are controlled by large networks of interacting proteins. A full understanding of any pathway or cellular process will require a map of the binary interactions among the proteins involved. Moreover, maps depicting specific protein interactions can indicate functions for novel proteins by placing them into partially characterized networks. This has become a particularly important task, because many of the proteins identified or predicted from genome sequences have not been assigned a role in any particular biological pathway or cellular process. Several methods have recently been developed to detect interactions among large sets of proteins, yet no single method has been proven to detect all interactions or to be free from generating false positives (Edwards et al. 2002; von Mering et al. 2002). Thus, the task of mapping protein interactions will continue to benefit from the development of new approaches and efficient variations of existing approaches.

One of the most widely used methods to detect biologically important protein-protein interactions is the yeast two-hybrid system (Fields and Song 1989; for reviews, see Fields and Sternglanz 1994; Brent and Finley Jr. 1997). In a two-hybrid assay, the two proteins are expressed in yeast, one fused to a DNA-binding

domain (BD) and the other fused to a transcription activation domain (AD). If the two proteins interact, they activate transcription of one or more reporter genes that contain binding sites for the BD. The assay is often conducted by mating a strain expressing the BD protein with a strain expressing the AD protein, and detecting reporter activity in the resulting diploid (Finley Jr. and Brent 1994). Several large-scale two-hybrid screens have been conducted to detect interactions using defined sets of proteins (Fromont-Racine et al. 2000; Walhout et al. 2000; Drees et al. 2001), and a few studies have set out to map the interactions among proteins encoded by a genome, including genome-wide screens for T7 bacteriophage, vaccinia virus, *Helicobacter pylori*, and the yeast *Saccharomyces cerevisiae* (Bartel et al. 1996; McCraith et al. 2000; Uetz et al. 2000; Ito et al. 2001; Rain et al. 2001). Although a tremendous amount of useful data has been obtained from these studies, comparison of the different data sets indicates that many protein interactions have been missed, even in attempts at comprehensive screens. For example, among the 1813 interactions that were identified in three different large-scale screens with yeast proteins, only six interactions were found in common among all three data sets (Ito et al. 2001). The lack of overlap among the data sets indicates that none of the screening approaches was truly comprehensive, and further indicates that alternative yeast two-hybrid screening methods might be used to detect more interactions.

Most of the approaches that have been used for high-throughput two-hybrid studies can be classified as either matrix screens or library screens. In a matrix screen, a single yeast strain expressing a BD-fused protein is mated with an array of strains

³Corresponding author.

E-MAIL rfinley@genetics.wayne.edu; FAX (313) 577-5218.

Article and publication are at <http://www.genome.org/cgi/doi/10.1101/gr.1134603>. Article published online before print in November 2003.

expressing different AD-fused proteins, and this is repeated for each new BD fusion (Finley and Brent 1994; Bartel et al. 1996; Uetz et al. 2000; Walhout et al. 2000). The result is a matrix in which each BD strain has been directly tested against every strain in the AD array. This approach is likely to be more comprehensive than library screening approaches (see below), which generally fail to test all possible binary combinations. However, for large sets of proteins, the matrix approach requires a large number of time-consuming and labor-intensive manipulations. Each BD strain must be mated with all AD strains, which are usually distributed over a number of 96-well or 384-well plates; thus, each BD strain must be mated with many AD plates, which amounts to thousands of mating operations for large arrays (e.g., see Supplemental Table 1). In a library screen, one or a small pool of BD strains is mated with a library or pool of strains expressing AD fusions, and the diploids are selected based on expression of the reporter gene(s). Colonies are picked, and the particular interacting AD protein in each is identified by sequencing the insert in the AD vector (Buckholz et al. 1999; Ito et al. 2000; Uetz et al. 2000). In cases in which small pools of BD strains are mated with the AD library, the particular BD-fused protein in each positive must also be determined by sequencing (Ito et al. 2000, 2001). Compared with the matrix approach, far fewer mating operations are required in a library screen. The entire AD library can be mated with 96 different BD strains in a single assay plate; for example, 6000 different BD strains could be mated with one library on just 63 plates. Library screening approaches, however, require a substantial amount of sequencing, because multiple positives from each library screen must be sequenced to attempt to detect all of the unique interactions. Thus, practical limits on the number of positives that can be sequenced in a given screen will contribute to missed interactions, or false negatives.

In addition to being inefficient, the library and matrix screening approaches work poorly with proteins that may be toxic or reduce yeast viability, and with BD fusions that activate the reporters on their own. Proteins that are transactivators are a particular problem for library screens because the positives that

result from an actual interaction cannot be distinguished from those due to activation of the reporters by the BD alone. Thus, activator BDs must be removed from any collection of BD strains before they can be used in a typical library approach (Walhout and Vidal 1999). Here we describe a large-scale two-hybrid screening strategy that, like the matrix approach, avoids the large number of sequencing reactions and the possible false negatives of library screening approaches. By using a novel two-phase mating scheme, our strategy requires only a fraction of the mating operations of previously described matrix approaches. We also show that our strategy allows us to detect interactions with BD proteins that activate transcription and with proteins that are detrimental to yeast growth.

RESULTS

A Two-Phase Pooling Scheme for Mating Large Two-Hybrid Arrays

An efficient approach to conducting two-hybrid assays has been to collect yeast clones expressing different BD-fused proteins and systematically mate them with collections of strains expressing AD-fused proteins (Finley Jr. and Brent 1994). Genome-wide two-hybrid studies have become feasible as high-throughput methods have been developed to construct very large arrays of yeast expressing fusion proteins (Hudson Jr. et al. 1997; Hua et al. 1998; Reboul et al. 2001). Here we propose a scheme to systematically mate large AD and BD yeast arrays to sample all possible binary combinations between them. The mating scheme involves two distinct steps or phases as illustrated in Figure 1. In the first phase, 96 AD clones are pooled, and the pool is mated with an array of 96 BD clones. This identifies individual members of the BD array that interact with one or more members of the AD pool. In the second phase, the specific BD clone(s) that interacted with the AD pool are individually mated with the part of the AD array from which the pool was derived. This reveals which AD(s) interact with that particular BD. If there is only one interaction

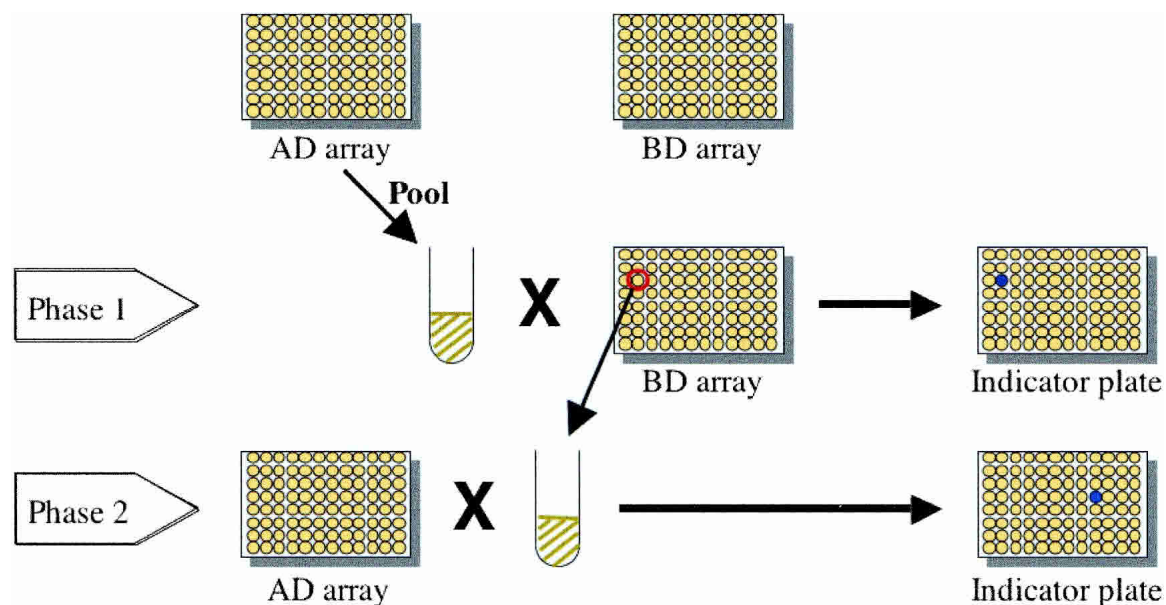


Figure 1 A two-phase scheme for sampling all possible interactions between two arrays. The strategy is illustrated for an array of 96 BD strains and an array of 96 AD strains. In the first step (Phase 1), the 96 AD clones are pooled, and the pool is mated with the BD array. This will reveal which of the BD-fused proteins interacts with one or more of the 96 AD-fused proteins. In the second step (Phase 2), positive BD clones are individually mated with the AD array. This will reveal which AD(s) in the array interact(s) with the BD. If there is only one interaction between the two arrays, this pooling scheme will require only two mating procedures to sample 9216 possible combinations.

between two 96-member arrays, this pooling scheme will require only two plate matings to identify the interaction. In contrast, the matrix approach would require 96 plate mating operations to find the one interaction, as each of the 96 BD strains would be mated individually with the AD array.

Several features of this two-phase strategy are notably different from related approaches. First, the two-phase strategy does not require prescreening and removal of strains containing BD-fused proteins that activate the reporters on their own. Such BD strains can be identified in a single experiment by mating the BD array with a strain not expressing an AD-fused protein; if the reporters are active for a particular BD strain, it indicates that the BD is a transactivator. In subsequent interaction experiments with the BD array, the position of the transactivating BD will be known. Interactions can be detected with such BD proteins by screening for AD pools or AD strains that cause an increase in reporter activation over the level generated by the BD fusion alone, as we show below. Second, the final interaction assay in the two-phase approach is equivalent to a one-on-one test of a binary interaction. This enables quantification of the level of reporter activation. Reporter activity has been shown to correlate with interaction affinity (Estojak et al. 1995), and may provide an additional metric for interpreting protein interaction maps. Third, the two-phase strategy dramatically reduces the number of mating operations needed to sample all possible binary interactions among the members of two large arrays.

The efficiency of a two-phase mating approach can be illustrated by comparing it with other approaches for a given set of proteins. Arrays of strains expressing the 6000 yeast proteins, for example, would occupy about 65 plates if configured in a standard 96-well format. To test all possible combinations between 6000 AD and BD strains using a standard matrix approach would require that each of the 6000 BD strains be individually mated with the 65 AD plates, for a total of 390,000 mating operations. In the two-phase approach, each of the 65 AD plates would be pooled and mated with each of the 65 BD plates, which would require 4225 mating operations. In the second phase, the number of matings to be performed depends on the number of interactions detected. Estimates from large-scale screens using the most sensitive matrix approach indicate an overall average of approximately two AD interactors per BD; roughly 50% of the BDs tested show no interactions, whereas the other 50% each interacted with an average of four ADs (Uetz et al. 2000). To detect 12,000 interactions among the 6000 yeast proteins in the second phase of a two-phase approach, the 6000 BD strains would each be mated with just two of the AD plates. Thus, the two-phase approach would require a total of 16,225 mating operations or only 1/24 the number required in the previously described matrix approaches. If a 384-well format were used in the matrix approach, our two-phase approach would still require only 1/8 of the mating operations. For larger sets of proteins, the difference becomes more dramatic. For example, to detect interactions among the ~14,000 predicted *Drosophila* proteins, the two-phase strategy would require 1/40 of the mating operations of the matrix approach (Supplemental Table 1).

Detection of a Single Weak Interactor in a Pool of Noninteractors

The two-phase pooling scheme outlined above would require the ability to detect a single strain expressing an interacting AD protein in a pool of 96 or more strains expressing noninteracting AD proteins. To test this, we selected a series of interacting proteins representing very weak to strong interactions, as defined by their levels of reporter activation in two-hybrid assays. Previous results

have shown that a broad range of detection sensitivities can be achieved by separately assaying the *LEU2* and *lacZ* reporters (Estojak et al. 1995). Very weak interactions, for example, can be detected by the sensitive assay for *LEU2* activity, growth on $-leu$ media (see Cdk2-Cdi12, Fig. 2A). The *lacZ* reporter, on the other hand, provides a range of phenotypes, from light blue to dark blue colonies on X-Gal plates, which distinguishes between weak and strong reporter activation (Fig. 2A). Interestingly, use of *LEU2* reporter selection in combination with X-Gal increases the sensitivity of *lacZ* detection, as seen on $-leu$ X-Gal plates. For example, weak interactions that lead to no blue or very light blue on X-Gal plates result in significantly blue colonies on $-leu$ X-Gal plates (e.g., see Cdk2-Cdi12 and Cdk2-Cdi5, Fig 2A; Supplemental Fig. 1). This is likely because of the selection of yeast containing more copies of the multicopy AD and BD plasmids on the $-leu$ media. Although this narrows the dynamic range of *lacZ* reporter detection, the increased sensitivity is useful for detecting interactions in pools (see below).

To test our ability to detect interactions using pools of AD clones, as would be necessary in the two-phase mating scheme, we made a series of pools by combining one strain expressing an interacting AD fusion with 95 strains expressing noninteracting AD fusions. We mated the AD pools with individual BD strains on a 96-well formatted plate and replicated the diploids onto three indicator plates; X-Gal, $-leu$, and $-leu$ X-Gal. The results shown in Figure 2B indicate that all but the weakest interaction could be detected when the interacting AD strain was diluted to 1/96. For example, when the AD-Cdi5 strain was diluted to 1/96, the interaction with BD-Cdk2 could be detected on both the $-leu$ and $-leu$ X-Gal plates. Most interactions could be detected when the AD strain was further diluted to over 1/192 (Supplemental Figs. 1 and 2). Interactions that strongly activated the reporters, however, were detectable in more diluted pools than those with weak reporter activation. For example, the strong interaction between BD-Cdk2 and AD-Cdi2 can be detected when AD-Cdi2 is diluted up to 1/3072, whereas the weaker BD-Cdk2-AD-Cdi5 interaction is not detectable when the AD strain is diluted beyond 1/384 (Supplemental Figs. 1 and 2). Combined, these results indicate that the two-phase pooling scheme outlined in Figure 1 could detect a broad range of interactions. Additionally, pooling schemes with pool sizes larger than 96 could be used, although weaker interactions will be missed as the pool size increases.

Use of Arrayed BD Strains Allows Detection of Interactions With Proteins That Activate Transcription on Their Own

Some BD-fused proteins activate the two-hybrid reporters on their own. This is common for eukaryotic transcription factors, but is also observed for other proteins that contain acidic domains or other features that fortuitously activate transcription in yeast. For example, BD-Cdp1 activates reporters at a low level, even when mated with a strain expressing no AD-fusion protein (Fig. 2A). Despite this transactivation, we can test for interactions with BD-Cdp1 by looking for increases in reporter activity in the presence of specific AD fusions. The interaction between BD-Cdp1 and AD-Cdi3, for example, results in a significant increase in *lacZ* activity over that from BD-Cdp1 alone (Fig. 2A). Furthermore, the interaction between BD-Cdp1 and AD-Cdi3 could be detected even when the AD-Cdi3 strain was diluted to 1/96 with noninteracting strains (Fig. 2B). Some interactions with transactivating BD fusions would be missed with this approach. For example, interactions involving very strong BD transactivators, or very weakly interacting AD proteins, would not be detected

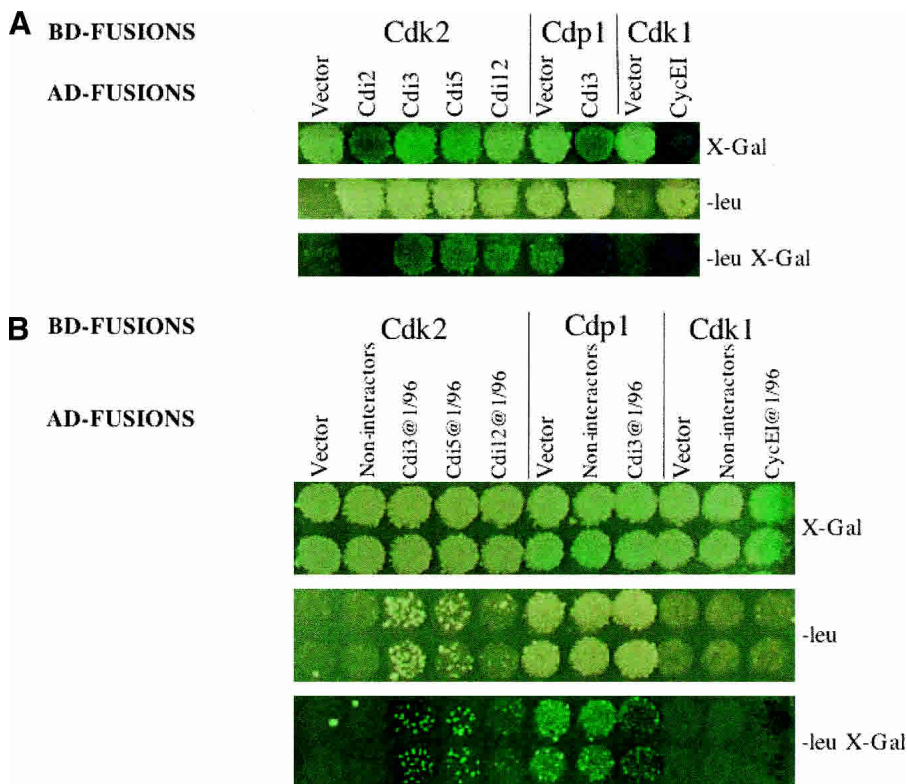


Figure 2 Detecting interactions in pools of AD strains. (A) Strains expressing LexA (BD) fused to one of three *Drosophila* proteins (Cdk2, Cdp1, Cdk1) were mated with individual strains expressing the AD fused to *Drosophila* proteins Cdi2, Cdi3, Cdi5, Cdi12, or CycEI, or no fusion (vector). Diploids were replicated to three kinds of indicator plates: X-Gal, $-leu$, and $-leu$ X-Gal. Growth on the $-leu$ plate or blue color on the X-Gal plate indicates activation of the *LEU2* or *lacZ* reporters, respectively. The interactions of Cdk2 with Cdi2, Cdi3, Cdi5, and Cdi12 result in strong to weak reporter activation. Cdp1 activates the reporter on its own. The interaction of Cdp1 with Cdi3 is evident from the increase in reporter expression in the presence of AD-Cdi3 (cf. Cdp1-vector with Cdp1-Cdi3). CycEI is toxic to yeast and leads to poor growth on all media. The strong interaction between Cdk1 and CycEI can be seen from the high level of *lacZ* expression, even though the yeast do not grow well. (B) The interactions shown in A in which the strain expressing the interacting AD has been diluted 1/96 by strains expressing noninteractors. Diploids were replicated onto the same three kinds of indicator plates as in A. The matings were performed in duplicate rows. The diluted interactions could not be detected on the X-Gal plates, but were detected on the $-leu$ and $-leu$ X-Gal plates. All plates in A and B were also $-his -ura -trp$ Gal/Raf to select for diploids and to induce expression of the AD protein. Use of $-leu$ X-Gal plates for detecting interactions in pools distinguishes true positives, which are blue, from the rare false positive leu^+ colonies, which are not blue.

because the increase in reporter expression would be small. Nevertheless, the results shown here indicate that the two-phase pooling scheme can detect some interactions with transactivating BD proteins.

Regulated Expression of AD and BD Proteins Allows Detection of Interactions Involving Toxic Proteins

Some proteins inhibit yeast growth or viability when they are expressed at levels typical for two-hybrid studies. Yeast expressing such proteins either fail to grow, or more typically, down-regulate the expression of the protein over several generations. This results in the gradual loss of phenotypes associated with the protein, such as the ability to interact in two-hybrid assays, as the cells are passaged. We reasoned that this problem could be avoided by constructing yeast arrays in which the AD- and BD-fused proteins are conditionally expressed. The yeast could then be grown and maintained under repressing conditions, and the

two-hybrid assays could be performed under conditions that induce expression of the fusion proteins.

In the LexA-based yeast two-hybrid system (Gyuris et al. 1993), expression of the AD-fused protein is controlled by the *GAL1* promoter (*GAL1p*), which can be regulated by adding glucose or galactose to the medium to repress or induce transcription, respectively. Strains containing plasmids that encode a toxic AD fusion can be grown and maintained in the array under repressing conditions (glucose). To detect interactions with such toxic proteins, expression can be briefly induced during the assay by replicating the diploids onto indicator plates containing galactose. *Drosophila* CycEI is an example of a protein that is toxic to yeast, as is evident by the light growth on galactose medium (Fig. 2A). Nevertheless, an interaction can be detected with AD-CycEI, as the poorly growing yeast turn dark blue on X-Gal indicator plates (Fig. 2A,B).

In most two-hybrid systems, the BD-fusion protein is expressed under control of the constitutive *ADHI* promoter (*ADHIp*). As a result, yeast expressing toxic BD fusions cannot be maintained in the array. This can be illustrated by following the loss of phenotypes associated with a toxic BD fusion. BD-CycEI, for example, is toxic to yeast and produces two phenotypes that can be monitored. It activates the reporters on its own, as seen when a strain expressing it is mated with a strain expressing no AD fusion, and it interacts with both AD-Rux and AD-Cdi4, which is evident from the increased reporter expression in the presence of these two proteins (Fig. 3). When BD-CycEI expression is under *ADHIp* control, its ability to transactivate and to interact are lost as the yeast are passaged (Fig. 3). After just 6 passages following plasmid transfor-

mation, the BD-CycEI interactions and transactivation were barely detectable (Fig. 3). To overcome this problem, we used a plasmid for regulated expression of BD fusions using the yeast *MAL62* promoter (*MAL62p*; Finley et al. 2002). *MAL62p* is repressed by glucose and induced by maltose. Yeast bearing the *MAL62p*-BD-CycEI plasmid grown in repressing conditions do not lose their ability to express the BD-CycEI fusion when induced. Thus, the transactivation ability of BD-CycEI, and its interactions with Rux and Cdi4, can be detected regardless of the number of passages (Fig. 3).

Detection of Protein Interactions Using the Two-Phase Approach

To further test the two-phase mating strategy, we used it to detect interactions between two members of two test arrays. We constructed an array of 47 BD strains, expressing 47 different known proteins, including several cell cycle regulators and unrelated proteins (see Methods). We also constructed three arrays of 96

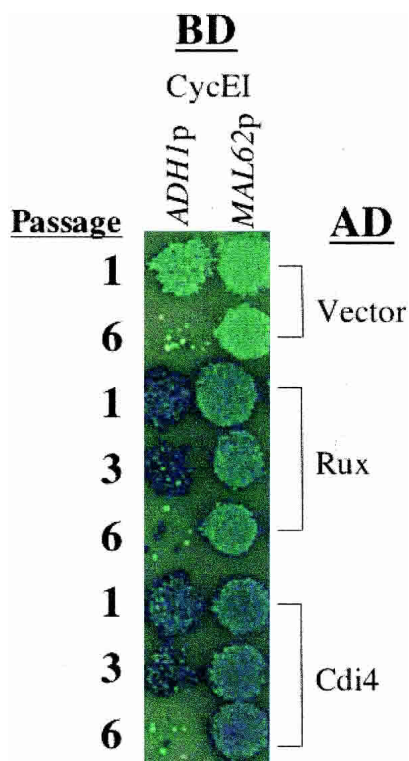


Figure 3 Regulated expression of toxic BD fusions. Constitutive expression of a BD fused to *Drosophila* cyclin E1 (CycEI) leads to loss of a functional BD fusion as yeast are passaged. Yeast were transformed with plasmids expressing BD-CycEI from the constitutive *ADHIp* promoter (*ADHIp*) or from the regulated *MAL62p* promoter (*MAL62p*), and then grown for the indicated number of passages on glucose media, which represses expression from *MAL62p* but not *ADHIp*. Yeast from each passage were mated with strains containing the empty AD vector or expressing AD fused to Rux or Cdi4, and the diploids were replicated onto $-leu$ X-Gal indicator plates containing both galactose and maltose to induce expression of the AD fusion and the *MAL62p*-driven BD-CycEI. The strain in which BD-CycEI expression is controlled by *ADHIp* gradually loses its interaction phenotype with Rux and Cdi4 as cells are passaged. Correspondingly, the level of background activation by BD-CycEI is also lost. In contrast, the strain in which BD-CycEI expression is controlled by *MAL62p* maintains its interaction phenotype and background activation from passage to passage.

AD strains. One expressed AD-fused proteins known not to interact strongly with any of the 47 BD fusions in the BD array. The other two AD arrays were identical to the first except that in each, one clone was replaced with a strain expressing an AD fusion known to interact with at least one of the BD fusions. To apply the two-phase strategy, we first mated the BD array with a strain containing just the AD vector to detect any background activation by the 47 BD fusions. As shown in Figure 4A, 18 of the BD fusions activated the reporters on their own, from a very low level (e.g., position E1) to a moderate level (e.g., position A6). To perform the first phase of the two-phase mating scheme, we mated the BD array with three pools of AD strains derived from each AD array (Fig. 4A). The pool of strains from AD Array I interacted with at least seven different BD strains: at positions A6, B1, B2, F5, F6, G1, and G5, and at the duplicated positions on the right half of the BD array. For example, the BD clone at G1 and G7 (Cdp1) shows darker blue when mated with the pool from AD Array I than when it was mated with the strain containing just the AD vector or with the pool of noninteractors. The pool of strains from AD Array II interacted with at least five different BDs, including BD-Cdk2 at positions B1 and B7.

Next, we performed the second phase of the two-phase approach by mating individual positive BD strains with the arrayed AD strains (Fig. 4B). This revealed that BD-Cdp1 interacted with the clone at position C2 of AD Array I, which contained AD-Cdi3. Also, BD-Cdk2 interacted with AD-Cdi5 at position E7 of AD Array II (Fig. 4B). Additional second-phase matings with the other positive BD strains also revealed that they resulted from specific interactions with either Cdi3 in AD Array I or Cdi5 in AD Array II, as expected (data not shown; Supplemental Table 4). Note that the second-phase mating produces semiquantitative information on reporter activity by separately assaying growth on $-leu$ and the level of blue on X-Gal media. These results demonstrate that the two-phase mating strategy is an effective approach to detecting a range of interactions, including interactions that are relatively weak and those involving BD fusions that activate transcription, and that the strategy works well with arrays containing a diverse set of proteins.

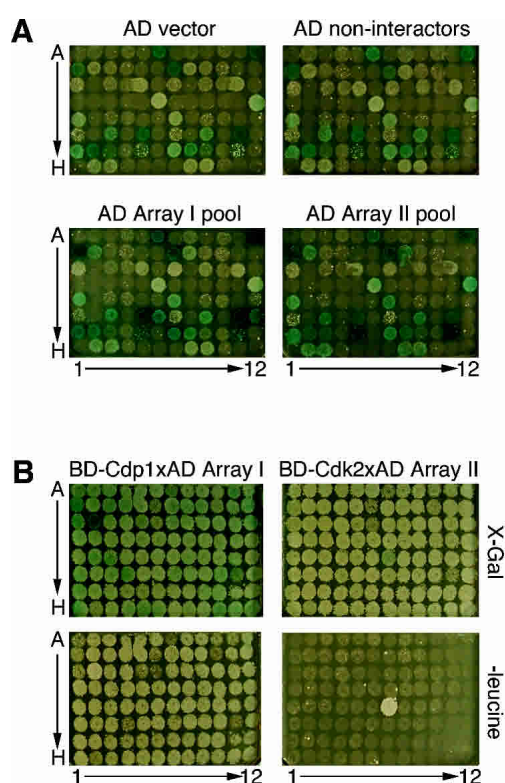


Figure 4 Detection of interactions using the two-phase strategy. (A) Phase 1 mating results. An array of 47 BD strains, duplicated on the right and left halves of a 96-well plate, was mated with four different AD cultures: the strain with the empty AD vector, a pool of strains from the array of 96 noninteractors (AD noninteractors), a pool of 96 strains from AD Array I, and a pool of 96 AD strains from AD Array II. Positions H6 and H12 have no BD strain. The diploids from each mating were replicated to $-leu$ X-Gal plates. The results show that one or more strains in the AD Array I pool interact with BD fusions at positions A6, B1, B2, F5, F6, G1, and G5, and the duplicate versions on the right half of the plate. The AD Array II pool interacts with BD fusions at positions B1, B2, D6, G2, and G5, and the duplicates. (B) Phase 2 mating results. A strain expressing BD-Cdp1 was mated with AD Array I, which contains AD-Cdi3 at position C2. A strain expressing BD-Cdk2 was mated with AD Array II, which contains AD-Cdi5 at position E7. Diploids were replicated onto two kinds of indicator plates: X-Gal and $-leu$. The Cdk2-Cdi5 interaction is evident from growth on the $-leu$ plate and the light blue yeast on the X-Gal plate. Because BD-Cdp1 activates the *LEU2* reporter on its own, the Cdp1-Cdi3 interaction is only evident on the X-Gal plate, where the yeast are blue.

A Three-Dimensional Pooling Scheme to Further Improve the Efficiency of Mating Large Arrays

For very large arrays, the two-phase pooling scheme we described here would still require a substantial amount of labor and materials. To further reduce the work needed to sample all possible combinations between two large arrays, we devised a three-dimensional (3D) pooling scheme. In this scheme, each pool of AD strains is constructed from the strains in one of three mutually orthogonal planes in a stack of 96-well plates (Fig. 5). One plane (*XY*) consists of the strains from individual plates, and the other two planes (*XZ* and *YZ*) consist of strains from rows and columns, respectively, from several different plates. To ensure equal representation of all AD strains, each pool is constructed to have the same number of strains. For example, pools of 192 strains can be constructed from a stack of 48 plates as illustrated in Table 1. As a result, each strain is found in only three of the 72 pools: one *XY* plate pool, one *XZ* row pool, and one *YZ* column pool. To test for interactions, each AD pool is mated with an array of BD strains, as in the two-phase scheme. If a BD strain interacts with one *XY* pool, one *XZ* pool, and one *YZ* pool, the specific interacting AD can be determined by the position in the 3D array where all three pools intersect (see Methods). This is possible because planes *XY* (plate pools), *XZ* (row pools), and *YZ* (column pools) intersect at only one point, corresponding to one AD strain. A second mating can be done to confirm interactions and to derive quantitative information on reporter activity, as in the two-phase scheme. In this second phase, individual BD strains are mated with individual AD strains, so that 96 interactions can be verified and quantified on one plate. The 3D pooling scheme has all the advantages of the two-phase mating scheme, but would require fewer matings to sample all possible interactions between two large arrays.

A comparison of the different array approaches illustrates the potential value of two-phase and 3D approaches (Supplemental Table 1). Consider BD and AD arrays each with 4608 strains distributed over 48 96-well plates, and in which each BD protein will interact with an average of two AD proteins. To sample all possible combinations using the matrix approach, each of the 4608 BD strains would be individually mated with each of the 48 plates, for a total of 221,184 mating operations. The two-phase

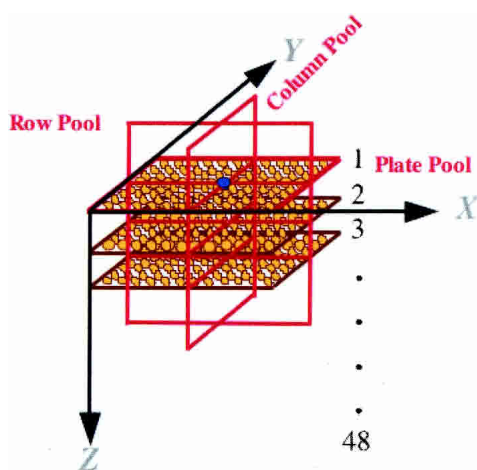


Figure 5 A 3D pooling scheme. A stack of 96-well plates represents a three-dimensional space defined by *X*, *Y*, and *Z* axes. AD pools are constructed from plates, which are in the *XY* plane, rows in the *XZ* plane, and columns in the *YZ* plane. If a BD interacts with three orthogonal AD pools, one from each plane, the interacting AD is indicated by the only common AD among all three pools. Pools can be constructed to contain an equal number of clones as shown in Table 1.

Table 1. Construction of 3D Pools

Plane	No. of pools	No. of clones per pool	Example of pools
<i>XY</i> (plate)	24	192	Plate #1 plus #25
<i>XZ</i> (row)	24	192	Row A from Plate #1 to #16
<i>YZ</i> (column)	24	192	Column 1 from Plate #1 to #24

Pools are constructed from a stack of 48 plates, each with 96 wells as illustrated in Figure 5.

approach outlined in Figure 1 would require 11,520 plate mating operations, 19-fold fewer than the matrix approach. Unlike in the matrix approach, the number of matings in the two-phase approach is influenced by the total number of interactions that will be detected; for example, if the average number of interactions for each BD strain were 20 rather than two, the two-phase approach would provide only a 2.3-fold reduction in matings over the matrix approach (Supplemental Table 1). In the 3D approach, the 48 BD array plates would each be mated with the 72 AD pools, which would require 3456 plate mating operations. The resulting 9216 interactions could be confirmed and quantified in 96 additional matings, with 96 interactions per plate. Thus, the total mating operations for the 3D approach would be 3552, a further threefold reduction from the two-phase approach, and 62-fold fewer than the matrix approach.

DISCUSSION

Maps depicting binary protein interactions that are likely to take place *in vivo* provide a good starting point for understanding biological pathways. Several techniques have been used to generate such maps for increasingly large sets of proteins. The techniques include computational predictions based on protein properties (Eisenberg et al. 2000) and experimental techniques that detect specific physical interactions between proteins. Two experimental methods that have been used for large sets of proteins are the yeast two-hybrid system and mass-spectrometry-based determination of the proteins in purified complexes (Gavin et al. 2002; Ho et al. 2002). Data from these two approaches are likely to be complementary. Two-hybrid assays detect binary connections between proteins but reveal little about possible multiprotein complexes or changes that may occur to complexes *in vivo*. Determination of protein complexes, on the other hand, may be used to detect the aggregate members of complexes in cells under a variety of conditions, but provides very little information about specific protein-protein contacts. Data from different approaches are not only complementary in the sense that they may generate subtly different data types (e.g., binary vs. complex interactions), but also in that they can be combined to create more comprehensive and reliable protein interaction maps. For example, comparison of yeast protein interaction data from several computational and experimental approaches has shown that combining multiple approaches increases the coverage and reliability of the data because different approaches produce different sets of false positives and false negatives (Bader and Hogue 2002; Deane et al. 2002; Edwards et al. 2002; von Mering et al. 2002).

Many of the benefits obtained by combining data sets from unrelated technologies could similarly be obtained by combining data from different two-hybrid screening strategies. For example, interactions found by two different methods are enriched for biologically relevant, true-positive interactions (Edwards et al. 2002; von Mering et al. 2002). This is also true for protein inter-

actions detected by different two-hybrid screens, which have used different expression constructs, reporter systems, and screening strategies (Schwikowski et al. 2000; Ito et al. 2001; Deane et al. 2002). Although it is possible that certain false positives will be common to all two-hybrid screens, the fact that the overlapping data are enriched for true positives lends support to the validity of interactions detected in more than one screen. Such cross-validation would be most useful if the screening approaches were comprehensive and detected many overlapping interactions. Unfortunately, use of different screening strategies on a similar set of proteins has thus far led to data sets with little overlap. The minimal level of overlap could be explained by two factors. One possible explanation is that the screens generated mostly false positives. However, even after filtering out the least reliable data, the proportion of overlap is still very small (Ito et al. 2001; Deane et al. 2002). Moreover, analyses based on comparisons with known interactions, on correlations with expression profiles, and on clustering of functionally related proteins within the interaction networks have indicated that the filtered data sets contain more true positives than false positives (Schwikowski et al. 2000; Deane et al. 2002). Thus, a more likely explanation for the lack of overlap is that the screens were not comprehensive in their ability to identify all of the interactions that could be detected with the yeast two-hybrid system.

We set out to design a two-hybrid strategy that would be more comprehensive than library screening approaches and more efficient than the matrix approach. Our strategy uses arrays of BD and AD strains, similar to the matrix approach. Arrays are relatively easy to construct, for example, by using recombination in yeast (Hudson et al. 1997; Hua et al. 1998), and each clone can be verified by sequencing. Arrays can also be used to construct pools with defined sets of clones in equal abundance. Our screening strategy uses a two-phase mating scheme to systematically mate the large yeast arrays. The increased efficiency should make it possible to screen much larger sets of proteins than would be practical by a matrix approach. During the screening, the BD and AD strains remain in an array format so that their identities are known from their positions. This is in contrast to library screening approaches in which each AD clone must be sequenced at least once, but usually many times, for every interaction. In one large library screen, for example, >6000 sequences had to be determined to identify just 692 unique interactions involving 817 proteins (Uetz et al. 2000); in another large screen, >13,000 sequences were determined to identify interactions among 3278 proteins (Ito et al. 2001). In the two-phase mating strategy, as in the matrix approach, each clone can be sequence-verified once, regardless of the number of interactions in which it is involved.

The two-phase strategy does not require pooling of BD strains, which remain in an array format. This allows the detection of interactions involving BD-fused proteins that activate transcription of the reporters on their own. Because the BD strains are in defined positions in the array, their inherent abilities to transactivate the reporters can be determined, and this level of activation can be subtracted from the level obtained in the presence of a particular AD-fused protein. Another element of our strategy is to construct the arrays so that the fusion proteins are conditionally expressed using regulated promoters. We have shown that this enables detection of interactions with proteins that are toxic to yeast.

The yeast two-hybrid system is a well-established technique for detecting biologically relevant binary protein interactions. It is robust, working with many different protein types from a variety of organisms, and it is simple and inexpensive. Moreover the significant amount of experience gained with the assay over the past 10 yr has provided a good understanding of its advantages and pitfalls, resulting in a sound framework for interpreting

two-hybrid data. Thus, it is likely that the ability to conduct large two-hybrid screens to map interactions on a genome-wide scale will continue to be an important part of functional genomics. The high-throughput two-hybrid strategy we described here should contribute to these efforts.

METHODS

Plasmids, Strains, and Yeast Two-Hybrid Assays

The yeast two-hybrid system used in this study (Gyuris et al. 1993; Kolonin et al. 2000) uses *Escherichia coli* LexA amino acids 1–202 as the DNA-binding domain (BD), *E. coli* B42 protein as the activation domain (AD), and two reporters, *lacZ* and *LEU2*, each containing upstream LexA binding elements. The plasmids used in this study are listed in Supplemental Table 1. Sequences and maps are available at proteome.wayne.edu. All vectors for expressing LexA fused to the N terminus of proteins (BD vectors) contain *HIS3* and the 2 μ origin, and are derived from pLEX(202 + PL) (Ruden et al. 1991). In the BD vectors pEG202 (Estojak et al. 1995) and pNLexAattR (Finley Jr. et al. 2002), the BD fusion is expressed from the constitutive *ADHI1* promoter (*ADHI1p*). In pHZ5attR (Finley et al. 2002), the BD fusion is expressed from the yeast *MAL62* promoter (*MAL62p*). pNLexAattR and pHZ5attR are destination vectors for use in the Gateway cloning system (Invitrogen). Both encode an SV40 nuclear localization signal (NLS) immediately downstream of LexA and upstream of the multiple cloning site. The BD plasmids used in Figure 2 for *ADHI1p*-driven expression, LexA-Cdk1 and LexA-Cdk2, were described previously (Finley Jr. and Brent 1994; Finley Jr. et al. 1996). *Drosophila* Cdp1 cDNA was isolated in a yeast two-hybrid screen using a *Drosophila* Cyclin D cDNA, Cdi3, as the bait (R.L. Finley, unpubl.). Cdp1 encodes the C-terminal 353 amino acids of the 665-amino-acid protein (GenBank accession no. AAM50984) encoded by predicted gene CG15610. The Cdp1 cDNA was subcloned from the AD library vector into pEG202 as an *EcoRI/XhoI* fragment. The *Drosophila* Cyclin EI cDNA from codon 1 to the end of the 602-amino-acid protein (GenBank accession no. NP_476960) was subcloned from a sequence-verified Gateway entry clone into the two destination vectors, pNLexAattR and pHZ5attR, to create pNLexAattB-CycEI and pHZ5attB-CycEI. All vectors for expressing an AD fused to the N terminus of proteins contain *TRP1* and the 2 μ origin, and are derived from pJG4-5 (Gyuris et al. 1993); the AD is expressed from the yeast *GAL1* promoter. pJG4-5 derivatives expressing AD fusions to *Drosophila* proteins, Cdi2, Cdi3, Cdi4, Cdi5, Cdi12, CycEI, and Rux, and their interactions with the BD fusions in this study were previously described (Finley Jr. and Brent 1994; Finley Jr. et al. 1996; Thomas et al. 1997).

S. cerevisiae haploid yeast strains, RFY231 (MAT α *his3*, *trp1* Δ ::*hisG*, *ura3*, 3*LexAop-LEU2*::*leu2* MAL⁺) and RFY206 (MAT α *trp1* Δ ::*hisG* *his3* Δ 200 *leu2-3 lys 2* Δ 201 *ura3-52 mal*⁻) were described previously (Finley Jr. and Brent 1994; Kolonin and Finley Jr. 2000). Yeast YPD and minimal media lacking the indicated nutrients were prepared as described (Golemis et al. 2002); –his lacks histidine, –ura lacks uracil, –trp lacks tryptophan, and –leu lacks leucine. The concentrations of sugars in media were as follows: glucose (Glu) at 2%, galactose (Gal) at 2%, raffinose (Raf) at 0.2%, and maltose (Mal) at 0.5%. The X-Gal concentration of indicator plates was 0.32 mg/mL in –leu X-Gal plates and 0.16 mg/mL (1 \times) in X-Gal plates, unless otherwise noted in the figure. Strains expressing AD fusions were RFY231 with the *TRP1* AD vector. Strains expressing BD fusions were RFY206 with the *HIS3* BD vector, plus the *lacZ* reporter plasmid, pSH18-34. Interaction mating assays were performed as described (Finley Jr. and Brent 1994) with the following modifications: 6 μ L of saturated BD strain liquid culture and 6 μ L of saturated AD or AD pool liquid culture were added onto the same spot in a 96-well configuration on YPD agar plates pored in an OmniTray (Corning) using a multichannel pipettor or a robot. The plates were incubated at 30°C for 24–48 h, and then yeast were replicated onto indicator plates by velvet. Growth on the –leu plate is scored in whole

numbers from 0 for no growth to 3 for heavy growth; the level of blue for diploids on the X-Gal plates is scored in whole numbers from 0 for white to 5 for dark blue. These measurements are highly reproducible for a given interacting pair (e.g., cf. Figs. 2A, 4B, and Supplemental Figs. 1 and 2).

Toxicity Assays

For the toxicity assays in Figure 3, yeast haploid strain RFY206/pSH18-34 was transformed with pNlexattBCycEI (*ADH1p*) and pHZSattBCycEI (*MAL62p*), and single colonies were picked into 2.0 mL of $-ura$, $-his$ Glu medium. Cultures were grown at 30°C with shaking (passage 0). Subsequent passages were generated by inoculating 50 μ L of saturated culture into 3.0 mL of medium ($OD_{600} \sim 0.07$) and growing to saturation ($OD_{600} = 3.4$), which corresponded to five to six generations. After each passage, cultures were stored at 4°C, which previous results had shown has no effect on toxicity or two-hybrid assays. For two-hybrid assays, 6 μ L of BD culture and 6 μ L of AD culture were pipetted onto the same spot of a YPD agar plate to mate.

Construction of Arrays and Pools

The array of 96 BD yeast clones expressing *Drosophila* BD fusions was constructed by duplicating 47 strains expressing different LexA fusions under *ADH1p* control from plasmids pEG202 or pLexA(1–202) + PL (Ruden et al. 1991; Estojak et al. 1995; for the list of strains, see Supplemental Table 3). Two wells were blank as negative controls. The array included LexA-Cdk1, LexA-Cdk2, and LexA-Cdp1, described above.

To construct an array of 96 yeast clones expressing AD-fused proteins that do not interact with any of the BD proteins used in this study, we transformed yeast strain RFY231 with AD-cDNA libraries derived from *Drosophila* embryos or imaginal discs (Finley Jr. et al. 1996), plated transformants on $-trp$ Glu agar plates, and randomly picked ~400 yeast colonies. We sequenced the cDNA from the 5'-end to determine whether or not it was fused in-frame with upstream sequences encoding the AD and whether or not the cDNA is unique. We then rearrayed 95 unique in-frame AD-cDNA yeast clones, and one strain containing the AD vector, into a single 96-well plate. This array was called AD-cDNAs-random (for the list of strains, see Supplemental Table 4). We mated the AD-cDNAs-random array with 85 known BD strains, including all the BD strains used in this study. Out of the 95 clones, 11 interacted with two or more BD strains and were removed from the array and replaced with the strain containing the empty AD vector (Supplemental Table 4). This new array of noninteracting AD clones was called AD-noninteractors. Two additional AD arrays were made by replacing specific strains in the AD-noninteractors array. For AD Array I, the clone at position C2 was replaced by a clone expressing AD-Cdi3; for AD Array II, the clone at position E7 was replaced by a strain expressing AD-Cdi5. The lists of strains in each array are in Supplemental Table 4.

AD pools were made as follows. We made a pool of noninteracting AD clones by growing the AD-noninteractors array in 96-well cluster tubes (maximum volume 1.2 mL per tube) to saturation in $-trp$ Glu 15% glycerol medium, and combining 40 μ L from each well into a one-well reservoir filled with 45 mL of $-trp$ Glu 15% glycerol liquid medium. We distributed 500 μ L per well into 96-well cluster tubes and grew them to saturation. AD pools containing one interactor and 95 noninteractors were made similarly except that one well of the array of 96 noninteractors was replaced with the AD interactor, which had been cultured to saturation before pooling. For more diluted pools, the interactor was further diluted with the pool of noninteractors.

ACKNOWLEDGMENTS

We thank members of the Finley laboratory for many helpful discussions, Guozhen Liu for help with bioinformatics, Bernie Mangiola and the Michigan Proteome Consortium for technical expertise with robotics, and Jodi Parrish for Figure 1. This work was supported by a grant from the National Institutes of Health R01HG01536.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

- Bader, G.D. and Hogue, C.W. 2002. Analyzing yeast protein-protein interaction data obtained from different sources. *Nat. Biotechnol.* **20**: 991–997.
- Bartel, P.L., Roecklein, J.A., SenGupta, D., and Fields, S. 1996. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat. Genet.* **12**: 72–77.
- Brent, R. and Finley Jr., R.L. 1997. Understanding gene and allele function with two-hybrid methods. *Annu. Rev. Genet.* **31**: 663–704.
- Buckholz, R.G., Simmons, C.A., Stuart, J.M., and Weiner, M.P. 1999. Automation of yeast two-hybrid screening. *J. Mol. Microbiol. Biotechnol.* **1**: 135–140.
- Deane, C.M., Salwinski, L., Xenarios, I., and Eisenberg, D. 2002. Protein interactions: Two methods for assessment of the reliability of high throughput observations. *Mol. Cell Proteomics* **1**: 349–356.
- Drees, B.L., Sundin, B., Brazeau, E., Caviston, J.P., Chen, G.C., Guo, W., Kozminski, K.G., Lau, M.W., Moskow, J.J., Tong, A., et al. 2001. A protein interaction map for cell polarity development. *J. Cell Biol.* **154**: 549–571.
- Edwards, A.M., Kus, B., Jansen, R., Greenbaum, D., Greenblatt, J., and Gerstein, M. 2002. Bridging structural biology and genomics: Assessing protein interaction data with known complexes. *Trends Genet.* **18**: 529–536.
- Eisenberg, D., Marcotte, E.M., Xenarios, I., and Yeates, T.O. 2000. Protein function in the post-genomic era. *Nature* **405**: 823–826.
- Estojak, J., Brent, R., and Golemis, E.A. 1995. Correlation of two-hybrid affinity data with in vitro measurements. *Mol. Cell. Biol.* **15**: 5820–5829.
- Fields, S. and Song, O. 1989. A novel genetic system to detect protein-protein interactions. *Nature* **340**: 245–246.
- Fields, S. and Sternglanz, R. 1994. The two-hybrid system: An assay for protein-protein interactions. *Trends Genet.* **10**: 286–292.
- Finley Jr., R.L. and Brent, R. 1994. Interaction mating reveals binary and ternary connections between *Drosophila* cell cycle regulators. *Proc. Natl. Acad. Sci.* **91**: 12980–12984.
- Finley Jr., R.L., Thomas, B.J., Zipursky, S.L., and Brent, R. 1996. Isolation of *Drosophila* cyclin D, a protein expressed in the morphogenetic furrow before entry into S phase. *Proc. Natl. Acad. Sci.* **93**: 3011–3015.
- Finley Jr., R.L., Zhang, H., Zhong, J., and Stanyon, C.A. 2002. Regulated expression of proteins in yeast using the MAL61–62 promoter and a mating scheme to increase dynamic range. *Gene* **285**: 49–57.
- Fromont-Racine, M., Mayes, A.E., Brunet-Simon, A., Rain, J.C., Colley, A., Dix, I., Decourty, L., Joly, N., Ricard, F., Beggs, J.D., et al. 2000. Genome-wide protein interaction screens reveal functional networks involving Sm-like proteins. *Yeast* **17**: 95–110.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415**: 141–147.
- Golemis, E.A., Serebriiskii, I., Finley Jr., R.L., Kolonin, M.G., Gyuris, J., and Brent, R. 2002. Interaction trap/two-hybrid system to identify interacting proteins. In *Current protocols in molecular biology* (eds. F.M. Ausubel et al.), Unit 20. John Wiley & Sons, Inc., New York.
- Gyuris, J., Golemis, E., Chertkov, H., and Brent, R. 1993. Cdi1, a human G₁ and S phase protein phosphatase that associates with Cdk2. *Cell* **75**: 791–803.
- Ho, Y., Gruhler, A., Heilbut, A., Bader, G.D., Moore, L., Adams, S.L., Millar, A., Taylor, P., Bennett, K., Boutilier, K., et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* **415**: 180–183.
- Hua, S.B., Luo, Y., Qiu, M., Chan, E., Zhou, H., and Zhu, L. 1998. Construction of a modular yeast two-hybrid cDNA library from human EST clones for the human genome protein linkage map. *Gene* **215**: 143–152.
- Hudson Jr., J.R., Dawson, E.P., Rushing, K.L., Jackson, C.H., Lockshon, D., Conover, D., Lanciault, C., Harris, J.R., Simmons, S.J., Rothstein, R., et al. 1997. The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res.* **7**: 1169–1173.
- Ito, T., Tashiro, K., Muta, S., Ozawa, R., Chiba, T., Nishizawa, M., Yamamoto, K., Kuhara, S., and Sakaki, Y. 2000. Toward a protein-protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl. Acad.*

- Sci.* **97**: 1143–1147.
- Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. 2001. A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci.* **98**: 4569–4574.
- Kolonin, M.G. and Finley Jr., R.L. 2000. A role for cyclin J in the rapid nuclear division cycles of early *Drosophila* embryogenesis. *Dev. Biol.* **227**: 661–672.
- Kolonin, M.G., Zhong, J., and Finley, R.L. 2000. Interaction mating methods in two-hybrid systems. *Methods Enzymol.* **328**: 26–46.
- McCraith, S., Holtzman, T., Moss, B., and Fields, S. 2000. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc. Natl. Acad. Sci.* **97**: 4879–4884.
- Rain, J.C., Selig, L., De Reuse, H., Battaglia, V., Reverdy, C., Simon, S., Lenzen, G., Petel, F., Wojcik, J., Schachter, V., et al. 2001. The protein–protein interaction map of *Helicobacter pylori*. *Nature* **409**: 211–215.
- Reboul, J., Vaglio, P., Tzellas, N., Thierry-Mieg, N., Moore, T., Jackson, C., Shin-i, T., Kohara, Y., Thierry-Mieg, D., Thierry-Mieg, J., et al. 2001. Open-reading-frame sequence tags (OSTs) support the existence of at least 17,300 genes in *C. elegans*. *Nat. Genet.* **27**: 332–336.
- Ruden, D.M., Ma, J., Li, Y., Wood, K., and Ptashne, M. 1991. Generating yeast transcriptional activators containing no yeast protein sequences. *Nature* **350**: 250–252.
- Schwikowski, B., Uetz, P., and Fields, S. 2000. A network of protein–protein interactions in yeast. *Nat. Biotechnol.* **18**: 1257–1261.
- Thomas, B.J., Zavitz, K.H., Dong, X., Lane, M.E., Weigmann, K., Finley Jr., R.L., Brent, R., Lehner, C.F., and Zipursky, S.L. 1997. roughex down-regulates G₂ cyclins in G₁. *Genes & Dev.* **11**: 1289–1298.
- Uetz, P., Giot, L., Cagney, G., Mansfield, T.A., Judson, R.S., Knight, J.R., Lockshon, D., Narayan, V., Srinivasan, M., Pochart, P., et al. 2000. A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403**: 623–627.
- von Mering, C., Krause, R., Snel, B., Cornell, M., Oliver, S.G., Fields, S., and Bork, P. 2002. Comparative assessment of large-scale data sets of protein–protein interactions. *Nature* **417**: 399–403.
- Walhout, A.J. and Vidal, M. 1999. A genetic strategy to eliminate self-activator baits prior to high-throughput yeast two-hybrid screens. *Genome Res.* **9**: 1128–1134.
- Walhout, A.J., Sordella, R., Lu, X., Hartley, J.L., Temple, G.F., Brasch, M.A., Thierry-Mieg, N., and Vidal, M. 2000. Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287**: 116–122.

WEB SITE REFERENCES

proteome.wayne.edu; Site with access to two-hybrid and other protocols, supplemental data, and protein interaction data.

Received December 23, 2002; accepted in revised form September 11, 2003.