Genomic Libraries and a Host Strain Designed for Highly Efficient Two-Hybrid Selection in Yeast

Philip James, John Halladay and Elizabeth A. Craig

Department of Biomolecular Chemistry, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received August 5, 1996
Accepted for publication September 18, 1996

ABSTRACT

The two-hybrid system is a powerful technique for detecting protein-protein interactions that utilizes the well-developed molecular genetics of the yeast *Saccharomyces cerevisiae*. However, the full potential of this technique has not been realized due to limitations imposed by the components available for use in the system. These limitations include unwieldy plasmid vectors, incomplete or poorly designed two-hybrid libraries, and host strains that result in the selection of large numbers of false positives. We have used a novel multienzyme approach to generate a set of highly representative genomic libraries from *S. cerevisiae*. In addition, a unique host strain was created that contains three easily assayed reporter genes, each under the control of a different inducible promoter. This host strain is extremely sensitive to weak interactions and eliminates nearly all false positives using simple plate assays. Improved vectors were also constructed that simplify the construction of the gene fusions necessary for the two-hybrid system. Our analysis indicates that the libraries and host strain provide significant improvements in both the number of interacting clones identified and the efficiency of two-hybrid selections.

IN the few years since its introduction (FIELDS and SONG 1989; CHIEN *et al.* 1991), the two-hybrid system has proven invaluable for identifying interactions between proteins. It has been used to establish physical interactions between genetically identified partners (JACKSON *et al.* 1993), identify contacts among the subunits of multiprotein complexes (LALO *et al.* 1993; BROWN *et al.* 1994), and to map the specific domains within proteins responsible for an interaction (LEE 1987; KALPANA and GOFF 1993; TAKACS *et al.* 1993). Most importantly, the two-hybrid system has allowed the identification of new interacting proteins by screening expression libraries (HANNON *et al.* 1993; VOJTEK *et al.* 1993). The two-hybrid system utilizes two plasmid-borne gene fusions that are cotransformed into a host yeast strain containing inducible reporter genes. The protein of interest, or "bait", is encoded as a gene fusion to a DNA binding domain from either the *GAL4* or *LexA* protein. A second protein, or a library of proteins, is encoded as a gene fusion to a transcription activation domain. Interaction between the two proteins results in the localization of the transcription activation domain to the DNA of the host strain, activating transcription of the adjacent reporter genes and generating a phenotypic signal.

The recent completion of the sequence of the *Saccharomyces cerevisiae* genome has demonstrated that ~50% of yeast genes have no known function and bear no homology to known genes (JOHNSTON *et al.* 1994; KOONIN *et al.* 1994; GALIBERT *et al.* 1996). Identifying the functions of these genes is likely to enhance our understanding of many biological systems, as recent examples of interspecies similarities such as MEC1/ATM (SANCHEZ *et al.* 1996) have shown. The fact that these genes have not been previously identified despite intensive study of yeast biology suggests that many are likely to be resistant to standard genetic and molecular approaches. We expect the two-hybrid system to take on increasing importance in elucidating the roles of these proteins. However, a variety of problems have limited the usefulness of the two-hybrid system, particularly for yeast gene analysis. Primary among these are two-hybrid expression libraries that provide poor representation of the genome as a whole, and two-hybrid host strains that produce a very high background of false positives and often require the addition of drugs that compromise their sensitivity to weak interactions.

There are a number of characteristics that are required in a high quality, highly representative two-hybrid library. The domain responsible for an interaction may occur anywhere within a protein, and many fusions to protein fragments may be nonfunctional due to problems with folding or stability. Therefore a comprehensive two-hybrid library must be of sufficient complexity not only to represent each gene, but to provide as many different fusion endpoints within each gene as possible. The library must also include fusions that encode the amino terminus of as many proteins as possible; this is the most difficult region to represent adequately. In addition, inserts should be kept to the minimum size required to produce a complete library and allow the expression of sufficiently large protein frag-

Corresponding author: Philip James, Department of Biomolecular Chemistry, University of Wisconsin, 1300 University Ave., Madison, WI 53706. E-mail: pjames@gcad.doit.wisc.edu

ments. Larger insert sizes used do not increase the number of protein fusion endpoints in the library, but do significantly increase the number of false positives, due to overexpression of extraneous genes present on the fragment (BARTEL *et al.* 1993b).

Existing yeast genomic two-hybrid libraries have been constructed using partial digestion by *Sau*3AI (CHIEN *et al.* 1991; Clontech) and large insert sizes (up to 6 kb). There are a limited number of *Sau*3AI sites present in the genome, and the gaps between them can be large. As a result the amino termini of most proteins will not be represented in these libraries, and many proteins will be completely absent. Restriction site cutting biases and imperfect partial digestions will exacerbate these problems. A yeast cDNA two-hybrid library has also been constructed (S. ELLEDGE, personal communication). However, while necessary in organisms with large genomes, cDNA libraries suffer from the problem of low or absent representation of poorly or conditionally expressed proteins and may be no better at representing the amino termini of proteins, particularly large proteins or those with stop codons immediately upstream of their ATG. Also, fusions to complete proteins or large fragments may not identify interactions as effectively as fusions to smaller protein fragments.

Successful two hybrid selections require a host yeast strain that contains inducible promoter elements fused to reporter genes, providing a detectable response to protein-protein interactions. A good two-hybrid host strain should be extremely sensitive to small changes in reporter activity to detect weak or transient interactions, yet provide a very low background of false positives. Some existing strains such as Y153 (DURFEE *et al.* 1993) and Y190 (S. ELLEDGE, personal communication) utilize a *lacZ* reporter that is sensitive but does not provide a selection mechanism, and a *HIS3* reporter that is sensitive but very leaky. This leakiness can be reduced by the addition of the drug 3-aminotriazole to the media (DURFEE *et al.* 1993), but in the high concentrations required for most host strains 3-aminotriazole severely inhibits true positives, reducing the sensitivity of the system (our unpublished results). Other strains such as L40 (VOJTEK *et al.* 1993) and HF7c (FEILOTTER *et al.* 1994) use *HIS3* alleles that do not require 3-aminotriazole, while EGY48 (GYURIS *et al.* 1993) uses a *LEU2* reporter. However all of these reporters also produce a very large number of false positives. In addition, all existing strains use a single promoter element, either the *GAL1* UAS or *LexA* operator, to drive each reporter gene. This increases background, because many false positives are specific to the promoter element (BARTEL *et al.* 1993a,b).

We report here the construction of highly representative yeast genomic libraries and a host yeast strain that is extremely sensitive and yet eliminates nearly all false positives. To generate two-hybrid libraries with increased complexity we have developed a novel approach that utilizes multiple restriction enzymes to partially digest genomic DNA. An examination of restriction enzymes with 4-bp recognition sequences revealed five enzymes (*Aci*I, *Hin*P1I, *Mae*II, *Msp*I, and *Taq*I) that produce identical 5′-CG-3′ overhangs following cleavage. Fragments digested with these enzymes are compatible for ligation into vectors digested with *Cla*I, which produces the same overhang. We took advantage of these compatible enzymes to generate a new set of yeast genomic two-hybrid libraries, resulting in a 3.4-fold increase in the number of fusion endpoints available compared to *Sau*3AI libraries and a dramatic increase in the representation of amino terminal ends. In addition, the host strain PJ69-4A was created containing three different reporter genes, each driven by a different promoter, to reduce the incidence of false positives. One reporter in particular, *GAL2-ADE2,* displays excellent sensitivity and extremely low background, making it ideal for two-hybrid selection. These new libraries and host strain, as well as new expression vectors reported here, will greatly increase both the ease and success of two-hybrid studies for yeast as well as other organisms.

MATERIALS AND METHODS

Bacterial strains and media: *Escherichia coli* strains used were RR1 {F- Δ*(gpt-proA)62 leuB6 supE44 ara-14 galK2 lacY1* Δ*(mcrC-mrr) rpsL20(Str^r) xyl-5 mtl-1*}, DH10B {F- *mcrA* Δ*(mrr-hsdRMS-mcrBC) ϕ80lacZ*Δ*M15* Δ*lacX74 deoR recA1 endA1 araD139* Δ*(ara, leu)7697 galU galK* λ- *rpsL nupG*}(Gibco BRL), and JM110 {F′ *traD36 lacI^q*Δ*(lacZ)M15 proA+B+/rpsL(Str^r) thr leu thi lacY galK galT ara fhuA dam dcm supE44* Δ*(lac-proAB)*}. LB, TB, M9, and SOC media were prepared as described (AUSEBEL *et al.* 1989) and supplemented with 100 μg/ml ampicillin, except where noted. Rescue of the *leuB6* mutation in RR1 by the yeast *LEU2* gene was assayed on M9 media supplemented with 0.02% proline.

Yeast strains, transformations, and media: Yeast strains used are shown in Table 1. Media was prepared as described (ROSE *et al.* 1990). All yeast transformations were done using the high efficiency LiAc method of GIETZ and SCHIESTL (1995); for library transformations the protocol was scaled up five- to 30-fold.

DNA sequencing: All sequencing was done using the method of SANGER (1977). The primer 5′-TTCGATGATGAAGATACC-3′ was used to sequence the amino-terminal end of fusions in the pGAD series of activation domain vectors. The primer 5′-AAGAGAGTAGTAACAAAG-3′ was used to sequence the amino-terminal end of fusions in the pGBD and pGBDU series of DNA binding domain vectors. The 3′ ends of these primers are located 25 and 27 bp upstream of the *Eco*RI site, respectively. The carboxyl-terminal end of fusions in either pGAD, pGBD, or pGBDU vectors were all sequenced using the same primer, 5′-TGAAGTGAACTTGCGGGG-3′, which is located 22 bp from the *Bgl*II site.

Construction of two hybrid expression vectors: To generate two-hybrid libraries using a set of five different restriction enzymes, we needed transcription activation domain vectors with a unique *Cla*I site in the polylinker. The oligonucleotide primers (1) 5′-TGCTTTCGAAGCTCCCCHACAGGTGTCCC-3′ and (2) 5′-TTAAGAATTCCCCGGGGGATCCATCGATGT-

TABLE 1

Yeast strains used in this study

Strain	Genotype	Source
YM706	*MATα gal4-542 ura3-52 his3-200 ade2-101 lys2-801 trp1-901 tyr1-501*	M. JOHNSTON
DGY63::171	*MAT***a** *ade2 trp1-901 leu2-3,112 his3-200 gal4Δ gal80Δ URA3::lexA-lacZ*	D. GIETZ
HF7c	*MAT***a** *ura3-52 his3-200 ade2-101 lys2-801 trp1-901 leu2-3,112 gal4-542 gal80-538 LYS2::GAL1$_{UAS}$-GAL1$_{TATA}$-HIS3 URA3::GAL4$_{17mers(\times 3)}$-CYC1$_{TATA}$-lacZ*	FEILOTTER *et al.* (1994)
PJ69-4A	*MAT***a** *trp1-901 leu2-3,112 ura3-52 his3-200 gal4Δ gal80Δ LYS2::GAL1-HIS3 GAL2-ADE2 met2::GAL7-lacZ*	This study

CGACCTGCAGAGATC-3′ were used to amplify a 1666-bp fragment from the vector pGAD424 (BARTEL *et al.* 1993a). Primer 1 contains a *Bst*BI site (underlined) in place of the *Cla*I restriction site in the *LEU2* gene of pGAD424. Primer 2 introduces two single-bp insertions and a *Cla*I site into the pGAD424 polylinker (underlined). These changes were incorporated into pGAD424 in two steps. First the PCR product was digested with *Bst*BI/*Afl*II and the 194-bp fragment was ligated into pGAD424 digested with *Cla*I/*Afl*II. Next the resulting plasmid and the PCR product were both digested by *Afl*II/*Eco*RI, *Afl*II/*Sma*I, or *Afl*II/*Bam*HI and each ~1.5-kb PCR fragment was subcloned into the appropriately digested vector. The *Cla*I site in the *LEU2* gene of the resulting plasmids was destroyed and a new *Cla*I site was introduced into the polylinker. We verified that each plasmid is still able to rescue both a *leu2* mutation in yeast and the *leuB6* mutation in *E. coli*. Cloning with *Bam*HI, *Sma*I, or *Eco*RI also incorporated zero, one, or two single base insertions, respectively, into the polylinker. The resulting plasmids, pGAD-C1, pGAD-C2, and pGAD-C3 (Figure 1), contain unique *Cla*I sites in their polylinker regions, each in a different reading frame. The DNA sequence across each polylinker region was verified by sequencing.

To generate DNA binding domain vectors that contain the same polylinker regions as the activation domain plasmids just described, pGAD-C1, pGAD-C2, and pGAD-C3 were digested with *Eco*RI/*Sph*I and the 595-bp fragment from each was included in a triple ligation with the 2996-bp *Aat*II/*Sph*I and 2298-bp *Aat*II/*Eco*RI fragments of pGBT9 (BARTEL *et al.* 1993a). The polylinker and transcription termination regions of pGBT9 are thus replaced by the corresponding regions of the pGAD vectors, yielding pGBD-C1, pGBD-C2, and pGBD-C3 (Figure 1).

To generate DNA binding domain vectors marked by the *URA3* gene, we first destroyed the *Pst*I site in the promoter of *URA3* so that the polylinker *Pst*I site would remain unique. The *URA3* gene was digested with *Pst*I, blunted with T4 DNA polymerase, and religated. Vectors pGBD-C1, pGBD-C2, and pGBD-C3 were digested to completion with *Pvu*II and partially digested by *Sph*I, and the 4882-bp fragment of each was recovered. The *Pst*I-deleted *URA3* gene was recovered as a 1.1-kb *Sph*I/*Sma*I fragment and ligated together with each vector fragment. The resulting vectors are pGBDU-C1, pGBDU-C2, and pGBDU-C3 (Figure 1).

Preparation of yeast genomic DNA: Genomic DNA was prepared from the *Saccharomyces cerevisiae* strain YM706. A 2-liter culture grown in YPD was harvested by centrifugation at OD_{600} 1.6, washed once in 500 ml ddH_2O, and resuspended in 30 ml of 1 M sorbitol, 0.1 M EDTA. Zymolyase 20T (ICN Pharmaceuticals) was added to 5 mg/ml and incubated 1 hr at room temperature. The spheroplasts were collected by centrifugation, resuspended in 40 ml lysis buffer (HOFFMAN and WINSTON 1987), and split into two 50-ml Falcon tubes. Twenty milliliters PCI (phenol:chloroform:isoamyl alchohol, 25:24:1) and 6 ml acid-washed glass beads (0.45–0.55 mm) were added to each tube, and the mixtures were vortexed at top speed for 5 min. The supernatants were recovered and pooled, reextracted three times with 20 ml PCI, and precipitated with 2 volumes ethanol. The pellet was resuspended in 20 ml 1× TE (10 mM Tris pH 8.0, 1 mM EDTA pH 8.0) and RNase A was added to 50 μg/ml. After 1 hr at 37° the DNA was extracted three more times with 20 ml PCI to achieve a clear interface and was precipitated with 0.6 volumes of isopropanol. The pellet was resuspended in 5 ml H_2O, and 1 ml 5 M NaCl and 6 ml 13% PEG 8000 were added. The DNA was precipitated 4 hr on ice and spun 10 min at 10,000 rpm in a SS-34 rotor (Sorvall). The pellet was resuspended in 3 ml 1× TE and the final yield was determined by spectrophotometry to be 2.75 mg.

Construction of yeast genomic two-hybrid libraries: Vector DNA was prepared by digesting 15 μg each of pGAD-C1, pGAD-C2, and pGAD-C3 with 40 units of *Cla*I for 6 hr at 37°. Digests were treated with Wizard Cleanup resin (Promega) and recovered in 0.1 ml 1× TE. Each vector digest was treated with 2 units of calf intestinal alkaline phosphatase (CIP, Boerhinger Mannheim) for 30 min at 37°, 2 additional units of CIP were added, and the reactions were incubated an additional 30 min at 55°. The CIP was removed by treatment with Wizard Cleanup resin. Each vector preparation was then incubated with 2 units T4 DNA ligase (Boerhinger Mannheim) at room temperature overnight, placed at 65° for 15 min, and electrophoresed on preparative 0.7% agarose gels to separate linear monomers from circularized and multimeric forms. Each linear vector DNA was eluted onto DE81 paper (Whatman) and recovered into 0.1 ml 1× TE.

Insert DNA was prepared by partial digestion of YM706 genomic DNA with the enzymes *Aci*I, *Msp*I, *Hin*P1I (New England Biolabs), *Mae*II (Boerhinger Mannheim), and *Taq*I (Promega). For each enzyme the optimal concentration for partial digestion was determined in units/μg of DNA, and then six preparative reactions were carried out with enzyme concentrations that bracketed the optimum. Each preparative digest contained 40 μg of genomic DNA in a total volume of 0.2 ml and was incubated for 30 min at the recommended temperature. The reactions were stopped by the addition of 20 μl 0.25 M EDTA.

Partial digests were analyzed by Southern blotting. The 315-bp *Nci*I/*Xba*I fragment of the *ADE2* gene was radioactively labeled using the random priming method and used as a probe (AUSEBEL *et al.* 1989). For each enzyme the digest that produced the most even distribution of partial and complete digestion products was selected and 10 μg of each digest were size fractionated on 1% agarose gels. DNA from 500 to 2000 bp for *Mae*II and *Taq*I, 500–2500 bp for *Aci*I, and 500–3000 bp for *Msp*I and *Hin*P1I was eluted onto DE81 paper and recovered into 0.1 ml 1× TE.

Fifteen ligation reactions were carried out (three vectors × five enzyme digests). The vector:insert ratio that resulted in optimal ligation efficiency was empirically determined. Each ligation reaction contained 5 μl of the vector preparation, 40

μl of the insert preparation, and 6 units of T4 DNA ligase in a total reaction volume of 100 μl, and was incubated at room temperature for 9 hr. Ligation reactions were ethanol precipitated and resuspended in 25 μl ddH_2O. Ligations were transformed into ultracompetent *Escherichia coli* strain DH10B (Gibco-BRL) by electroporation, using a 1.8 kV pulse, 1 mm gap cuvettes (BTX, Inc.), 20 μl of cells, and 1 μl of ligation mix in each transformation reaction. For each of the 15 vector:insert combinations, three to seven transformation reactions were necessary to generate sufficient numbers of transformants. One milliliter of SOC was added to each transformation reaction and incubated 1 hr at 37°. The three to seven transformation reactions from each ligation were pooled and a small aliquot was plated to determine the total number of primary transformants. The remainder of the transformation mixes from each ligation were inoculated into 3 liters of T broth containing 200 μg/ml ampicillin and incubated overnight at 37°, resulting in a total of 15 cultures. At OD_{600} 1.5 (~12 hr) cultures were chilled in ice for 30 min. For each of the three different reading frame vectors, five cultures (corresponding to the five restriction enzyme digests) were pooled in appropriate quantities to produce an equal representation of all restriction sites in the final library. The resulting three pooled cultures, one representing each reading frame, contained ~11 liters each and were harvested by centrifugation. DNA was prepared from each pellet by large scale alkaline lysis and precipitated with ethanol. The pellets were resuspended in 120 ml 1× TE and treated 1 hr at 37° with 10 mg RNase A. Each was extracted five times with 60 ml PCI and ethanol precipitated in the presence of 300 mM NaAc pH 7.0. Pellets were resuspended in 33 ml ddH_2O, and 7 ml 5 M NaCl and 40 ml 13% PEG 8000 were added. The DNA was precipitated overnight on ice and spun down 10 min at 10,000 rpm in a GSA rotor (Sorvall). The DNA pellets for libraries Y2HL-C1, Y2HL-C2, and Y2HL-C3 were resuspended in 1× TE at a concentration of 1 mg/ml.

Construction of new two-hybrid reporter genes: The *ADE2* gene was isolated on a 2250-bp *Bgl*II fragment from the plasmid pAS11 (STOTZ and LINDER 1990) and subcloned into the *Bam*HI site of pRS316 (SIKORSKI and HIETER 1989) to give pRS316-ADE2. This was digested with *Eag*I/*Afl*II and religated to give pRS316-ade2Δ, which contains an amino-terminal fragment of the *ADE2* locus with a unique *Xba*I site at amino acid 2. A 1044-bp fragment of the *ADE2* locus was amplified from genomic DNA using the primers 5′-GCTCCCGGGTTAGCTATTTCGC-3′ and 5′-CAATAGGGACGTCTCACTGG-3′, and a 780-bp fragment of the *GAL2* promoter was amplified from genomic DNA using the primers 5′-GGGGAGCTCTGCAGAAGGCACATC-3′ and 5′-AGGTCTAGAGTTCTCCTCAACTGCC-3′. The 5-kb *Xma*I/*Xba*I fragment of pRS316-ade2Δ, the 226-bp *Xma*I/*Dra*I fragment of the *ADE2* PCR product, and the 774-bp *Ecl*136II/*Xba*I fragment of the *GAL2* PCR product were combined by triple ligation to yield a plasmid containing a fusion of the 5′ flanking region of *ADE2*, the promoter and amino acids 1–5 of *GAL2*, and the *ADE2* coding region from amino acid 2 to 233. This cassette was removed as a *Sma*I/*Aat*II fragment and subcloned into the same sites of pRS316-ADE2, restoring the carboxyl-terminal portion of the *ADE2* coding region and resulting in the plasmid pGAL2-ADE2.

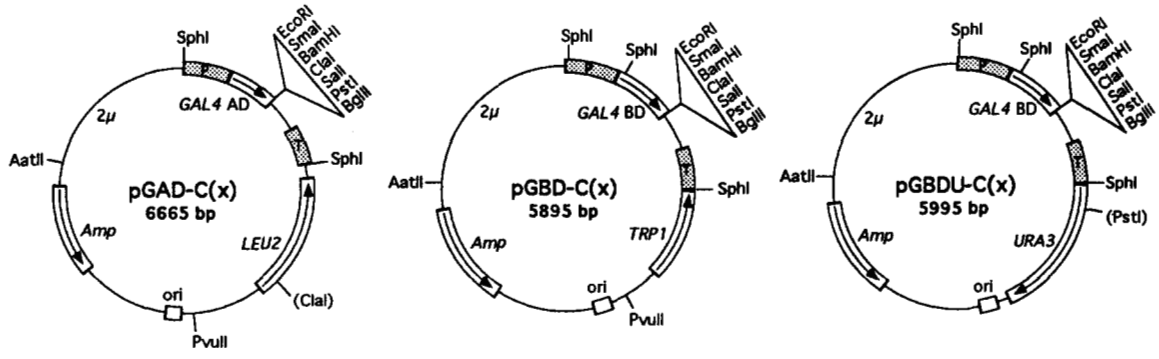
The *MET2* gene was recovered from the plasmid pPL4-2-1 (LANGIN *et al.* 1986) on a 2.1-kb *Sal*I/*Bam*HI fragment and subcloned into the same sites of pRS306 (SIKORSKI and HIETER 1989) to give pRS306-MET2SB. The *GAL7* promoter was recovered from the plasmid pBM756 (M. JOHNSTON, unpublished result) on a 1-kb *Eco*RI/*Bam*HI fragment and subcloned into the same sites in the vector pMC1790 (CASADABAN *et al.* 1983), creating an out-of-frame fusion of the *lacZ* gene to the promoter and amino acids 1–3 of *GAL7*. To correct the reading frame, the *Bam*HI site was made blunt with Klenow. We verified that the *Bam*HI site had been filled correctly by demonstrating that a Dam-sensitive *Cla*I site had been created in its place. The *GAL7-lacZ* cassette was subcloned as a 4-kb *Eco*RI/*Bsp*EI fragment into the *MET2* coding region using the same sites in pRS306-MET2SB. The resulting plasmid is pGAL7-lacZ.

Construction of a new yeast strain for two-hybrid selections: Yeast strain PJ69-4A was created using sequential rounds of plasmid integration and 5-FOA selected excision events. After each integration or excision step the correct strain phenotypes were verified in the presence and absence of the plasmid pCL1 (FIELDS and SONG 1989), which contains the wild-type *GAL4* gene. First, DGY63::171 was plated on 5-FOA media to select popouts of the existing *URA3*-marked *LexA-lacZ* construct. The resulting strain was transformed with the plasmid pGH1 (a gift from P. BARTEL) linearized with *Pvu*II. pGH1 is marked by *URA3* and contains the *HIS3* gene driven by the *GAL1* promoter. This construct is placed in the 3′ flanking region of the *LYS2* gene. Ura^+, His^+ transformants were plated on 5-FOA media to select for excision of the vector sequences. A transformant that was His^+ only in the presence of pCL1 was selected and transformed with a 2.6-kb *Sma*I/*Hae*II fragment of plasmid pGAL2-ADE2, resulting in a one-step gene replacement of the *ade2-101* locus. Transformants were selected on SC minus Ade media in the presence of plasmid pCL1. A transformant that displayed the expected phenotypes was transformed with the plasmid pGAL7-lacZ, which had been linearized by digestion with *Pac*I, resulting in integration of the plasmid at the *MET2* locus. Ura^+, $lacZ^+$ transformants were plated on 5-FOA media to select for excision of the pGAL7-lacZ vector sequences. The resulting strain, PJ69-4A, is Ura^-, Met^-, and Lys^+, and in the presence of Gal4 activity is His^+, Ade^+, and $lacZ^+$.

RESULTS

Construction of improved two-hybrid vectors: As a first step in creating improved genomic two-hybrid libraries, we needed a set of vectors with a unique *Cla*I cloning site that could be used to generate fusions to a transcription activation domain. The *GAL4* activation domain vector pGAD424 (BARTEL *et al.* 1993a) was chosen as a starting point because of its small size, convenient polylinker sites, and efficient rescue of *E. coli leuB* mutations. Site-directed PCR mutagenesis was used to make several changes in pGAD424. First, the endogenous *Cla*I site present in the *LEU2* marker gene was destroyed. This change did not alter the amino acid sequence of *LEU2*, and the plasmid retained the ability to rescue both *leu2* mutations in yeast and *leuB6* mutations in *E. coli* (data not shown). Second, a new *Cla*I site was introduced into the center of the polylinker and single base pairs were inserted between the *Eco*RI and *Sma*I sites and between the *Sma*I and *Bam*HI sites of the polylinker. By incorporating either zero, one, or two of the single base pair insertions into the final products, we generated a set of three new activation domain vectors, pGAD-C1, pGAD-C2, and pGAD-C3. Each contains a unique *Cla*I cloning site in the polylinker, but differs in the translational reading frame of the polylinker sites (Figure 1). In addition, we subcloned each of these polylinker regions into the *GAL4*

a.



b.

	EcoRI	SmaI	BamHI	ClaI	SalI	PstI	BglII
pGAD-C2	AAA GAG ATC	GAA TTC	CCG GGG	ATC CAT	CGA TGT	CGA CCT	GCA GAG ATC TAT GAA TCG TAG ATA CTG AAA
PGAD-C3	AAA GAG ATC	GAA TTC	CCG G*GG	GAT CCA	TCG ATG	TCG ACC	TGC AGA GAT CTA TGA ATC GTA GAT ACT GAA
PGAD-C1	AAA GAG ATC	GAA TTC	C*CC GG*G	GGA TCC	ATC GAT	GTC GAC	CTG CAG AGA TCT ATG AAT CGT AGA TAC TGA

FIGURE 1.—New expression vectors for use in two-hybrid analysis. (a) Structure of the pGAD-C(x), pGBD-C(x), and pGBDU-C(x) vectors. Each map represents a set of three different vectors that are identical except for the translational reading frame of the polylinker region. Stippled regions indicate the *ADH1* promoter (P) and transcription termination (T) elements. *GAL4* AD (activation domain) encodes amino acids 768–881. *GAL4* BD (DNA binding domain) encodes amino acids 1–147. The restriction sites destroyed by site-directed mutagenesis during construction are indicated in parentheses in pGBD-C(x) and pGBDU-C(x). (b) Sequences of the polylinker regions. Restriction sites are underlined; stop codons are boxed. *, single base pair insertions introduced during construction. *Eco*RI is present in one reading frame, *Sma*I in two, and all other sites in all three reading frames. Sequences shown are for the pGAD-C(x) vectors; sequences for pGBD-C(x) and pGBDU-C(x) vectors are identical except that the first nine nucleotides preceding the *Eco*RI site are GTA TCG CCG. The complete sequences of these vectors have been deposited in Genbank under the accession numbers U70018 (pGBD-C1), U70019 (pGBD-C2), U70020 (pGBD-C3), U70021 (pGBDU-C1), U70022 (pGBDU-C2), U70023 (pGBDU-C3), U70024 (pGAD-C1), U70025 (pGAD-C2), and U70026 (pGAD-C3).

DNA-binding domain vector pGBT9 (BARTEL *et al.* 1993a), resulting in the vectors pGBD-C1, pGBD-C2, and pGBD-C3. The multiple reading frames present in these vectors allow easier construction of fusion plasmids, and the identical polylinkers simplify the transfer of library inserts from the pGAD vectors into the DNA binding domain vectors.

Construction of yeast genomic two-hybrid libraries: To generate a highly complex set of libraries, we sought to increase the number of restriction enzymes used for the partial digestion of genomic DNA. There are five different commercially available enzymes with 4-bp recognition sequences that produce a 5′ overhang with the sequence 5′-CG-3′: *Aci*I, *Hin*P1I, *Mae*II, *Msp*I, and *Taq*I ("CG enzymes"). Each is compatible for ligation with the *Cla*I site introduced into the polylinkers of the vectors described above. ARNOLD *et al.* (1988) have predicted the frequency of all tetranucleotides in the yeast genome; we used this data to examine the potential of the CG enzymes to provide comprehensive genome coverage (Table 2). On average, a CG enzyme recognition site is predicted to occur every 97 bp, while *Sau*3AI alone, used in most library constructions, is predicted to cut genomic DNA once every 373 bp, a 3.8-fold difference.

Library inserts were prepared using genomic DNA from yeast strain YM706. This strain contains the mutant alleles *gal4-Δ542, his3-Δ200,* and *ade2-101,* eliminating the possibility that wild-type alleles of these genes are included in the libraries, which would lead to false positives. The genomic DNA was subjected to partial digestion by each of the CG enzymes. Because the inclusion of as many restriction sites as possible is critical to the quality of the library, partial digests were examined by Southern blotting (Figure 2). This examination was extremely important, as we found that it was impossible to predict the quality of the partial digest by ethidium bromide staining. Southern blotting results demonstrated that digests judged to be partial by ethidium bromide staining actually contained fragment mixes that ranged from nearly complete to nearly uncut.

After using Southern blots to select the best partial digest for each enzyme, we size-fractionated the selected digests by gel electrophoresis. The selection of small

TABLE 2

Library construction data

Enzyme	Site frequency (%)	Average fragment size (bp)	Predicted no. of 5′ ends	Library	Independent clones obtained	F_1	Expected no. of 5′ ends missed
*Aci*I	0.247	405	69,160	Y2HL-C1	1,645,700	4.63×10^{-11}	0
				C2	757,500	1.75×10^{-5}	1.2
				C3	2,373,400	1.25×10^{-15}	0
*Hin*PI	0.106	943	29,680	C1	834,000	6.26×10^{-13}	0
				C2	1,225,800	1.16×10^{-18}	0
				C3	1,138,300	2.21×10^{-17}	0
*Mae*II	0.250	400	70,000	C1	1,545,500	2.58×10^{-10}	0
				C2	654,700	8.67×10^{-5}	6.1
				C3	2,499,900	3.09×10^{-16}	0
*Msp*I	0.142	704	39,760	C1	1,461,600	1.08×10^{-16}	0
				C2	520,300	2.07×10^{-6}	0.1
				C3	1,901,200	1.71×10^{-21}	0
*Taq*I	0.281	356	78,680	C1	1,595,600	1.56×10^{-9}	0
				C2	686,800	1.62×10^{-4}	12.7
				C3	3,752,100	1.95×10^{-21}	0
Total CG sites	1.026	97	287,280				
*Sau*3AI	0.268	373	75,068				

All figures are based on a genome size of 14 Mb. Each restriction site in the genome generates two potential 5′ ends. Site frequency is the percentage of all tetranucleotides in the genome that match the enzyme recognition site (ARNOLD *et al.* 1988). F_1 is the probability that any individual 5′ end is not present in a library, and is defined by the formula $F_1 = (1 - p)^N$, where N is the number of independent clones and p is the inverse of the number of 5′ ends in the genome (FINKEL *et al.* 1994).

insert sizes was desirable to minimize the production of false positives, as discussed above. However, if the selected insert size is too small, many complete digestion products will not be represented in the library. To avoid excluding completely digested fragments from the library, we examined complete digestions of genomic DNA by ethidium bromide staining. The largest fragment size visible was chosen to be the maximum insert size selected for each enzyme. As a result, restriction fragments from 500 bp to a maximum of 2–3 kb were included in the libraries, with the maximum size depending on which enzyme was used to generate a partial digest.

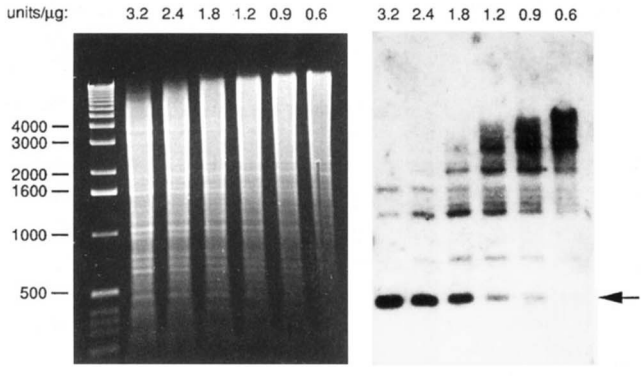


FIGURE 2.—Analysis of a partial digestion of genomic DNA using Southern blotting. The panel on the left shows a set of *Hin*PII partial digestion reactions separated by electrophoresis and stained with ethidium bromide. Enzyme concentrations are shown above each lane in units per microgram of DNA. Size markers are labeled on the left. The right panel shows a Southern blot of the same gel probed with a 315-bp fragment of the *ADE2* gene. The fragment size that results from complete digestion is indicated by an arrow.

Vector DNA was prepared by digesting pGAD-C1, pGAD-C2, and pGAD-C3 with *Cla*I and treating the cleaved ends with calf intestinal phosphatase to remove the 5′ phosphate groups. Dephosphorylated vectors were then religated, and those that remained as linear monomers were purified by electrophoresis. This step eliminated ~90% of the background caused by vector self-ligation (data not shown). After determining empirically that a 1:8 vector:insert ratio provided optimal ligation efficiency, 15 ligation reactions were performed and each was transformed into *E. coli*. Fifteen separate reactions were required to ligate each of the three different reading frame vectors to each of the five different partial digestions.

Production of comprehensive libraries required that a large number of independent clones be obtained from each of the 15 ligation reactions. Because of the practical limitations of recovering this number of colonies from plates, each reaction was inoculated into liquid T broth. To minimize the problem of clone loss due to competition, cultures were harvested while still in the logarithmic phase of growth. The five cultures that represent each reading frame were then mixed together in quantities that would produce equal representation of each individual restriction site in the genome. For instance, because the genome contains twice as many *Taq*I sites as *Msp*I sites (Table 2), one OD of cells will contain only one-half as many genome equiva-

lents of *TaqI* sites as *MspI*. Therefore, twice as many ODs of cells are required from the *TaqI* cultures to achieve equal representation of all sites. This approach is valid only if each of the 15 individual transformation reactions produce a highly significant number of independent clones; our results indicate that this requirement was met (Table 2). The mixing produced three large cultures, each representing a different reading frame and containing a comprehensive library of genomic DNA inserts. DNA was prepared from each of the three libraries, which we call Y2HL-C1, Y2HL-C2, and Y2HL-C3.

Assessment of two-hybrid library quality: The quality of the two-hybrid libraries has been examined by a variety of methods. First, as shown in Table 2, a large number of independent clones were obtained from each of the 15 transformation reactions, making the probability of failure (F_1) extremely low. F_1 is a statistical measure describing the probability that any particular 5′ end in the genome will be missing from one of the libraries (FINKEL *et al.* 1994). Each restriction site provides two potential 5′ ends for library fusions. This analysis predicts that for libraries C1 and C3 there is virtually no possibility that any 5′ ends are missing due to insufficient sampling. In the case of library C2, only 20 of the over 287,000 potential 5′ ends in the genome are predicted to be absent.

The quality of a library is also dependent on the percentage of library clones that contain an insert. DNA was prepared from 65–75 random clones from each of the three libraries and inserts were analyzed by digestion with the polylinker enzymes *SmaI* and *PstI* followed by gel electrophoresis. A total of 205 clones were analyzed; we found 10 without inserts and an additional 11 that contained small inserts of <400 bp. Thus 95% of the library plasmids contained inserts, and only 5% of those inserts were not within the size range we had selected. Also, as in any library of this type, a small number (5–10%) of library plasmids may contain two or more independent inserts.

The statistical analysis above makes the basic assumption that each restriction site in the genome will be digested and ligated with equal efficiency. However biases in restriction site cutting and ligation clearly exist, and can significantly affect the population of restriction sites actually present. We first assessed the representation of the five different restriction enzymes used in library construction. The fusion points of nearly 200 library clones have been sequenced by our lab and others. Analysis of these fusion points indicates that each of the five restriction enzymes are represented in the libraries at the expected frequency with the exception of *TaqI*, which is underrepresented by ~20% (data not shown). We then assessed the fidelity with which individual genomic restriction sites are represented in the libraries. Clones containing fragments from four different genes were examined to determine which of the restriction sites in those genes are actually present as fusion endpoints in the libraries. The sizes of library DNA fragments that result from using specific endpoints in *PTP2, PTC1,* and two unidentified ORFs on chromosomes *6* and *11* were predicted from the positions of CG enzyme restriction sites within their sequences. Using PCR and Southern blotting, DNA bands were identified that correspond to 38 of the 40 predicted fragment sizes (data not shown).

TABLE 3

Performance of the Y2HL libraries in two-hybrid selections

Group	No. of baits screened	Library	Known interactions identified	New interactions identified
A	3	Genomic[a]	N/A	2
		Y2HL	N/A	6
B	5	cDNA[b]	1 of 5	1
		Y2HL	3 of 5	5
C	3	Y2HL	2 of 2	9
Total	10	Y2HL	5 of 7	17

Column sums do not match the totals because one bait was screened using all three libraries and is included in both groups A and B.

[a] CHEIN *et al.* (1991).

[b] S. ELLEDGE, personal communication.

The libraries were functionally tested in two ways. A two-hybrid selection was carried out using the bait plasmid pSE1112 (DURFEE *et al.* 1993), which encodes the Snf1 protein fused to the Gal4 DNA binding domain. We chose 12 of the 30 positive clones, rescued the library plasmid from each, and sequenced them. Eight of the clones encoded in-frame fusions to either *GAL83, SIP1,* or *SIP3.* Each of these proteins has been previously reported to interact with *SNF1* (YANG *et al.* 1992, 1994). The remaining four clones encoded fusions to three different genes, none of which were in-frame. Interestingly, these do not appear to be false positives. Each is able to activate transcription in combination with pSE1112, but not with several other bait plasmids tested. Furthermore, all three genes encode polyglutamine repeat regions and were fused in the same reading frame in the antisense orientation. The predicted result in each case is the expression of a long in-frame polycysteine repeat.

In addition, six other laboratories have carried out complete two-hybrid screens using the Y2HL libraries. As outlined in Table 3, those groups have screened a total of 10 baits with the Y2HL libraries. Three of those same baits (Group A) were also used to screen another yeast genomic library (CHIEN *et al.* 1991), while five baits (Group B) were also used to screen a yeast cDNA library (S. ELLEDGE, personal communication). In each case, the Y2HL libraries were significantly better for identifying both previously described and novel interac-

tions. In total, five out of seven previously identified interacting proteins and 17 new interactions were identified from the Y2HL libraries using these 10 baits.

Construction of an improved two-hybrid yeast strain: A major difficulty in the use of the two-hybrid system is the elimination of false positives. These are clones that cause the activation of reporter genes, but not as a result of a specific interaction with the bait. False positives arise frequently in current host strains and many are promoter-specific (BARTEL *et al.* 1993a,b). Because all of the available host strains use a single promoter element to drive the transcription of each reporter gene, promoter-specific false positives are especially difficult to eliminate. We reasoned that the use of three reporter genes, each under the control of different inducible promoters, would improve the ability to discriminate between false positives and real interactions. The *GAL1, GAL2,* and *GAL7* promoters were selected for this purpose because all are induced to high levels (BRAM *et al.* 1986), all respond to the same inducer (Gal4), and they share a minimum of sequence identity.

The strain DGY63::171 was selected as a starting point because it is transformed with high efficiency and provides strong expression of β-galactosidase (D. GIETZ, personal communication). The endogenous *URA3::lexA-lacZ* allele was removed by selecting for the spontaneous excision event on 5-FOA media. Three separate reporter gene constructs were then introduced into the genome by sequential gene replacement (Figure 3). A *HIS3* gene under control of the *GAL1* promoter was introduced downstream of the *LYS2* gene. This construct exhibits more stringent regulation than other *GAL1*-driven *HIS3* alleles, reducing or eliminating the need for the drug 3-aminotriazole (P. BARTEL, personal communication). The *ade2-101* allele was replaced by a wild-type *ADE2* gene under the control of the *GAL2* promoter. *ADE2* was chosen as a reporter because *ade2* mutants form red colonies, while wild-type cells form white colonies. Intermediate expression of *ADE2* results in colonies in various shades of pink; thus colony color can be used as an initial indication of the strength of an interaction or to identify plasmid loss events. Finally, a *lacZ* gene under the control of the *GAL7* promoter was inserted into the coding region of the *MET2* gene to provide a quantitative measure of interaction strength. We have named the resulting two-hybrid host strain PJ69-4A.

Assessment of two-hybrid yeast strain quality: An analysis of the function of PJ69-4A in two-hybrid selections showed that the three reporter genes present provide both a high level of sensitivity and an extremely low background of false positives. The *GAL1-HIS3* reporter behaves as hoped; 0–2 mm 3-aminotriazole is sufficient to eliminate growth due to leakiness on SC minus His media. These low levels of 3-aminotriazole have no effect on the growth of true positives, making this a very sensitive reporter. However, as in other host strains, the *GAL1-HIS3* marker produces a very high background of false positives. This problem is overcome in PJ69-4A by the *GAL2-ADE2* reporter, which is equally sensitive, yet exhibits a background level that is two orders of magnitude lower.

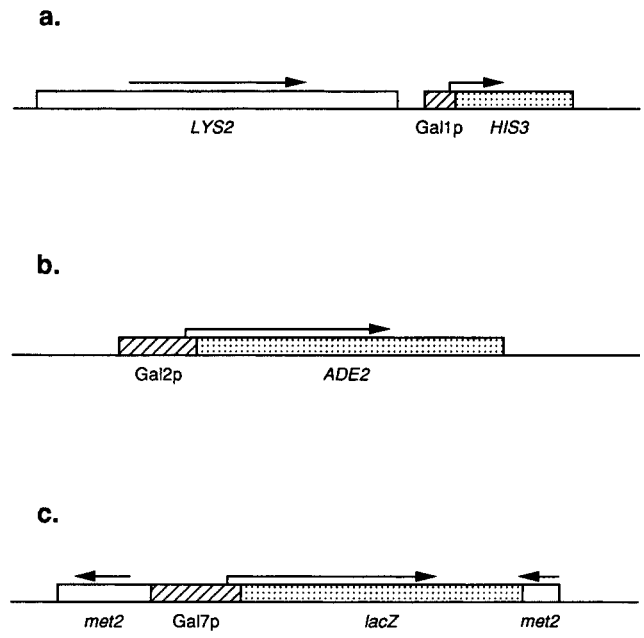


FIGURE 3.—Genomic structure at the reporter gene loci in PJ69-4A. Gal4-induced promoters are indicated by ▨, reporter genes by ▩, and adjacent coding regions by □. (a) Structure of the *LYS2::GAL1-HIS3* reporter locus. (b) Structure of the *GAL2-ADE2* reporter locus. (c) Structure of the *met2::GAL7-lacZ* reporter locus.

To compare the marker genes in PJ69-4A, we carried out two selections using the bait plasmid pSE1112, which carries a *SNF1* gene fusion. One selection was plated on SC minus Trp, Leu, His media with 1 mM 3-aminotriazole and the other on SC minus Trp, Leu, Ade media; in both cases $3–4 \times 10^6$ library transformants were plated. The first 350 His^+ colonies to arise on SC minus Trp, Leu, His media were picked; new colonies, totaling over 1000, continued to arise until the plates became dry. Only 18 of the 350 His^+ colonies selected were also Ade^+ and had β-galactosidase activity. We observed similar results in screens with three other baits; despite utilizing the more stringent *HIS3* reporter construct, 95–99.5% of colonies selected as His^+ appear to be false positives. We also found that the proportion of true positive clones did not diminish among colonies that arose late in the incubation period, so selecting only the earliest colonies to appear would exclude many true positives. In contrast, the second selection on SC minus Trp, Leu, Ade media identified only 36 Ade^+ colonies; all appeared within the first 5 days of growth. Thirty of those colonies were also His^+ and had β-galactosidase activity (these are the same 30 positives used to assess library quality, above). In both

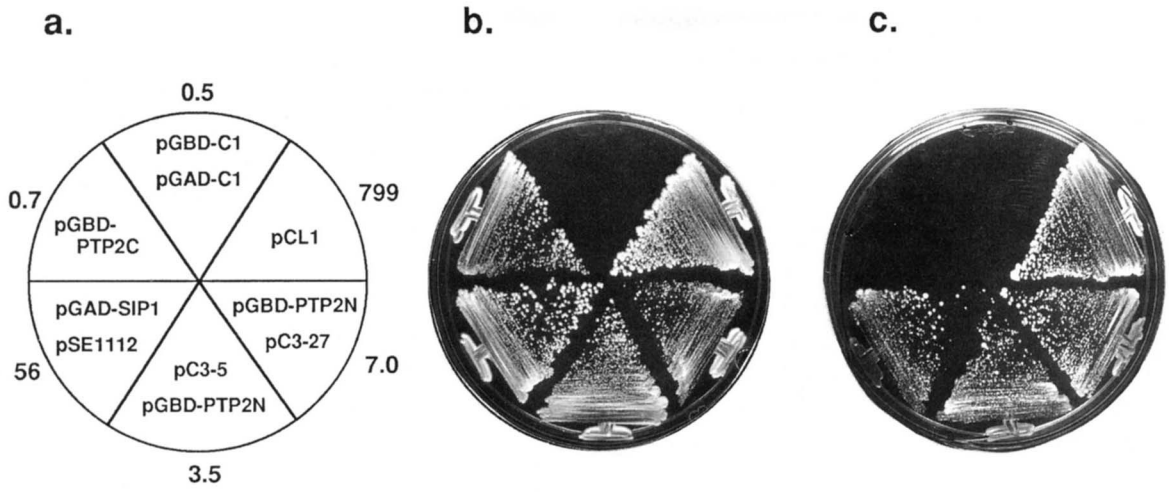


FIGURE 4.—Phenotypes of the two hybrid host strain PJ69-4A. (a) Plasmids transformed into PJ69-4A are indicated in each sector. The level of β-galactosidase activity produced by each plasmid combination is indicated around the outside of the circle. (b) Growth produced by different plasmid combinations on SC minus Trp, Leu, His media containing 1 mM 3-aminotriazole. (c) Growth produced by different plasmid combinations on SC minus Trp, Leu, Ade media. Plasmid pCL1 expresses the wild-type Gal4 protein. pGBD-C1 and pGAD-C1 are vectors without inserts. pSE1112 expresses a DNA binding domain fusion to the Snf1 protein. pGAD-SIP1 (YANG *et al.* 1992) expresses a transcription activation domain fusion to the Sip1 protein. pGBD-PTP2N and pGBD-PTP2C express DNA binding domain fusions to the amino- and carboxyl-terminal regions of Ptp2, respectively. pC3-5 and pC3-27 are library plasmids whose fusion proteins interact specifically with the pGBD-PTP2N product. β-galactosidase assays were performed as previously described (MILLER 1972).

screens, every colony that was positive in all three reporter assays resulted from specific interactions with *SNF1*, as judged by plasmid loss assays and tests with alternate bait plasmids.

We were concerned that the *GAL2-ADE2* reporter might in fact be too stringent, and might fail to detect weak interactions. This is not the case, since protein-protein interactions that generate as little as 3.5 units of β-galactosidase activity are able to grow without difficulty on SC minus Ade media (Figure 4). This activity level is only seven times the background level of 0.5 units, and only 0.5% of the activity produced by the wild-type Gal4 protein. In addition, we used plasmid loss assays or sequencing to examine 20 library clones that arose as weak His^+ colonies and had borderline β-galactosidase activity (1–1.5 units), but were Ade^-. All 20 were false positives. Given these results we feel the best selection method using PJ69-4A is to carry out the primary selection step using SC minus Trp, Leu, Ade media.

The rationale for using a different galactose-inducible promoter for each reporter in PJ69-4A was to eliminate promoter-specific false positives. One example of the effectiveness of this strategy is provided by the bait plasmid pGBD-PTP2C, which contains a fusion to the carboxyl-terminal half of the Ptp2 protein (Figure 4). This plasmid activates transcription of *GAL1* promoter-driven reporters in a variety of host strains, including SFY526 (BARTEL *et al.* 1993b), Y153 (DURFEE *et al.* 1993), HF7c (FEILOTTER *et al.* 1994), and the *GAL1-HIS3* allele in PJ69-4A. As a result it has not been possible to screen libraries with this bait using other Gal4-induced host strains, because all reporters in these strains are under the control of *GAL1* promoter elements. However PJ69-4A remains Ade^- and $lacZ^-$ in the presence of pGBD-PTP2C, and thus should allow library screening using this bait. In addition, we examined a collection of positive clones from a different genomic library (CHIEN *et al.* 1991) that had been selected using the bait pGBD-PTP2N, containing a fusion to the amino-terminal half of Ptp2. These clones were selected in the host strain HF7c (FEILOTTER *et al.* 1994), which contains *GAL1-HIS3* and $GAL4_{17mers(x3)}$-$CYC1_{TATA}$-*lacZ* reporters; both promoters are derived from *GAL1*. The clones had passed every test available in HF7c, including plasmid loss assays and retransformation with alternate baits. However, when part of this collection was sequenced, many clones contained no in-frame fusion but did carry nonfused downstream genes encoding transcription factors. When transformed into PJ69-4A, all were able to activate the *GAL1-HIS3* reporter; however, about half, including those noted above, were unable to activate *GAL2-ADE2* and *GAL7-lacZ*.

An additional benefit provided by PJ69-4A is that the strain remains $Ura3^-$, because gene replacement, rather than plasmid integration, was used to introduce reporter constructs into the genome. We have taken advantage of this fact by constructing an additional set of DNA binding domain vectors that are marked by the *URA3* gene rather than *TRP1* (Figure 1). When these vectors are used to construct bait plasmids, plasmid loss assays can be carried out quickly and simply by plating candidate clones on SC minus Ade media containing 5-FOA. Clones that are able to grow on this media are false positives and are discarded.

DISCUSSION

One factor influencing the quality of a two-hybrid library is the number of potential fusion sites that are represented. Because some fusions will be nonfunctional due to misfolding, instability, or lack of the correct domain for interaction, additional fusion sites increase the chance that a protein will be represented as a productive fusion. We analyzed the recently completed yeast genome sequence to compare the effectiveness of the CG enzymes to *Sau*3AI, the enzyme used most often for genomic library construction. The yeast genome (exclusive of the rDNA repeat region) contains 123,230 CG enzyme recognition sites, or an average of one every 98 bp, almost precisely the predicted rate (ARNOLD *et al.* 1988) (Figure 2). *Sau*3AI sites occur an average of every 337 bp, slightly more frequently than predicted. A more detailed analysis of four randomly chosen chromosomes, *I, III, VI,* and *X* (OLIVER *et al.* 1992; BUSSEY *et al.* 1995; MURAKAMI *et al.* 1995; GALIBERT *et al.* 1996), which constitute ~13% of the genome, shows that digestion by *Sau*3AI would generate 227 restriction fragments that are over 1 kb in length. Because 1 kb is roughly the size of an average gene in yeast, fragment sizes this large make it likely that many genes will contain no *Sau*3AI sites and will be completely excluded from a *Sau*3AI library. In contrast, digestion by the CG enzymes would generate many more fragments than *Sau*3AI, but only four are larger than 1 kb and <1% are larger than 500 bp.

The 3.4-fold increase in the number of potential fusion sites, and the corresponding decrease in fragment sizes, predict a significant improvement in genome coverage in the Y2HL libraries. However it also increases the number of clones to be screened from each library, since more fusion endpoints must be sampled for each genome equivalent. To screen the Y2HL libraries at a 95% confidence level, ~860,000 clones must be examined from each of the three libraries; a 99% confidence level will require ~1,325,000 clones from each library (FINKEL *et al.* 1994). Generating this number of primary transformants is routine using current yeast transformation protocols (GIETZ and SCHIESTL 1995).

A second factor determining library quality is the ability to generate fusions close to the amino-terminus of proteins, since this is the most poorly represented region in all two-hybrid libraries. We examined the first 60 coding regions in the sequence of chromosome *VI* (MURAKAMI *et al.* 1995) to determine how well the amino-termini of proteins might be represented. We identified the first two sites that could provide an in-frame fusion point for each gene. Examining more than one site is important because any individual site might not be represented due to enzyme cutting biases or might produce a nonfunctional fusion protein. In the case of *Sau*3AI, seven of the 60 genes, greater than 10%, contained no sites at all and thus are excluded from *Sau*3AI libraries. Eight other genes contained no second site, and in total 25 of the 60 genes had no *Sau*3AI site within the first 300 bp of the coding region. In contrast, every gene had at least two CG sites. In 50% of the genes both CG sites were within the first 100 bp of the coding regions; 74% had at least one CG site within that region. In addition, for 19 of the 60 genes there were in-frame CG sites located upstream of the ATG codon; thus one-third of the proteins could be represented in their entirety. Only one gene had a similarly placed *Sau*3AI site. This analysis suggests that our approach to library construction should result in a much more complete two-hybrid library; the functional assessments of the libraries presented in this paper indicate this is in fact true.

We developed the two-hybrid host strain PJ69-4A based on the hypothesis that utilizing different inducible promoters would eliminate many false positives. The *GAL1, GAL2,* and *GAL7* promoter elements share homology only in their 17-bp Gal4 binding sites, and among the three promoters the most similar elements match in only 11 of 17 positions. The effectiveness of this approach is demonstrated by the fact that we have carried out screens using four different baits and have not yet identified any false positives that activate more than one reporter gene, nor have we found any real interactions that are unable to activate all three reporters.

As a result, PJ69-4A allows for extremely efficient two-hybrid selections. The *GAL2-ADE2* reporter is especially useful, because it is very sensitive to weak interactions and eliminates nearly all false positives in the first step of the selection. Most host strains produce a large number of false positives, which are screened out using labor intensive plasmid loss assays followed by either mating or retransformation to introduce alternate bait plasmids. Using PJ69-4A, the few false positives that remain after a primary selection on SC minus Ade media are easily screened out using the *GAL1-HIS3* and *GAL7-lacZ* reporters. PJ69-4A does contain a mutant *ura3* allele, allowing the use of 5-FOA selection for increased efficiency in plasmid loss assays. However, as noted above, in four separate screens we have not yet identified any false positive that can activate all three reporters but is subsequently screened out, either by plasmid loss assays or through nonspecific interactions with other baits. In fact, we have screened libraries with two baits for which there appear to be no detectable interactions. It took over 2 months to eliminate the many false positives and arrive at this conclusion using the host strain HF7c. However using PJ69-4A, no Ade^+ colonies arose, and the same result was clear within 10 days.

Many researchers are accustomed to using β-galactosidase filter assays as a secondary screen in other host strains. In PJ69-4A, the *GAL7* promoter appears to be activated in response to treatments commonly used for cell permeabilization, such as flash freezing. This causes background β-galactosidase activity that makes the use

of filter assays unfeasible. However this is not a serious problem, since two simple and effective plate assays already exist using the *GAL2-ADE2* and *GAL1-HIS3* reporters. Also, because the *GAL2-ADE2* reporter in PJ69-4A is so much more efficient than other host strains, only true positives remain by the final stage of the screening process. Therefore there is no need to deal with large numbers of clones in β-galactosidase assays. Finally, the function of the *GAL7-lacZ* reporter in quantitative liquid assays is excellent (Figure 4); the reporter can also be assayed using X-gal plates, although this is less sensitive than filter assays.

We anticipate that these new tools will be extremely valuable to researchers studying a wide variety of organisms. The vectors and approach to library construction presented here can be used to produce highly representative libraries from other organisms. The host strain PJ69-4A can be used to screen any two-hybrid library containing fusions to a transcription activation domain, regardless of the organism from which the fusion proteins are derived. The only requirement is that the bait plasmid contain a fusion to the Gal4 DNA binding domain. In addition, the Y2HL libraries reported here can be used in combination with any bait plasmid and host strain. It is increasingly clear that yeast can serve as an excellent model for the biology of higher organisms, as many proteins from other organisms have been shown to function in yeast. We believe that the Y2HL libraries are likely to be the most complete two-hybrid libraries available from any organism. With the sequencing of the yeast genome now complete, it will be even more productive to screen yeast libraries such as these for proteins that can interact with those from other organisms.

The authors thank MARK JOHNSTON, DAN GIETZ, PAUL BARTEL, and MARIAN CARLSON for gifts of strains and plasmids, RODOLFO ARAMAYO for extensive discussions, and the laboratories of ROLF STERNGLANZ, DAVID SHORE, JEREMY THORNER, DAN GIETZ, JAN PALACEK, and JON ROSAMOND for sharing results before publication. This work was supported by grant GM-27870 (E.A.C.) from the National Institutes of Health.

LITERATURE CITED

ARNOLD, J., A. J. CUTICCHIA, D. A. NEWSOME, W. W. JENNINGS and R. IVARIE, 1988 Mono- through hexanucleotide composition of the sense strand of yeast DNA: a Markov chain analysis. Nucleic Acids Res. **16:** 7145–7158.

AUSEBEL, F., R. BRENT, R. KINGSTON, D. MOORE, J. G. SEIDMAN *et al.*, 1989 *Current Protocols in Molecular Biology*. John Wiley and Sons, New York.

BARTEL, P. L., C.-T. CHIEN, R. STERNGLANZ and S. FIELDS, 1993a Using the two-hybrid system to detect protein-protein interactions, pp. 153–179 in *Cellular Interactions in Development: A Practical Approach*, edited by D. A. HARTLEY. Oxford University Press, Oxford.

BARTEL, P. L., C.-T. CHIEN, R. STERNGLANZ and S. FIELDS, 1993b Elimination of false positives that arise in using the two-hybrid system. Biotechniques **14:** 920–924.

BRAM, R. J., N. F. LUE and R. D. KORNBERG, 1986 A GAL family of upstream activating sequences in yeast: roles in both induction and repression of transcription. EMBO J. **5:** 603–608.

BROWN, N. G., M. C. COSTANZO and T. D. FOX, 1994 Interactions among three proteins that specifically activate translation of the mitochondrial COX3 mRNA in *Saccharomyces cerevisiae*. Mol. Cell. Biol. **14:** 1045–1053.

BUSSEY, H., D. B. KABACK, W. ZHONG, D. T. VO, M. W. CLARK *et al.*, 1995 The nucleotide sequence of chromosome I from *Saccharomyces cerevisiae*. Proc. Natl. Acad. Sci. USA **92:** 3809–3813.

CASADABAN, M. J., A. MARTINEZ-ARIAS, S. K. SHAPIRA and J. CHOU, 1983 β-galactosidase gene fusions for analyzing gene expression in *Escherichia coli* and yeast, pp. 293–308 in *Methods in Enzymology*, edited by R. WU, L. GROSSMAN and K. MOLDAVE. Academic Press, New York.

CHIEN, C. T., P. L. BARTEL, R. STERNGLANZ and S. FIELDS, 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:** 9578–9582.

DURFEE, T., K. BECHERER, P. L. CHEN, S. H. YEH, Y. YANG *et al.*, 1993 The retinoblastoma protein associates with the protein phosphatase type 1 catalytic subunit. Genes Dev. **7:** 555–569.

FEILOTTER, H. E., G. J. HANNON, C. J. RUDDELL and D. BEACH, 1994 Construction of an improved host strain for two hybrid screening. Nucleic Acids Res. **22:** 1502–1503.

FIELDS, S., and O. SONG, 1989 A novel genetic system to detect protein:protein interactions. Nature **340:** 245–246.

FINKEL, R. A., S. BENT and C. L. SCHARDL, 1994 Determining the probability of obtaining a desired clone in an amplified or shuttle library. BioTechniques **16:** 580–582.

GALIBERT, F., D. ALEXANDRAKI, A. BAUR, E. BOLES, N. CHALWATZIS *et al.*, 1996 Complete nucleotide sequence of Saccharomyces cerevisiae chromosome X. EMBO J. **15:** 2031–2049.

GIETZ, R. D., and R. H. SCHIESTL, 1995 Transforming yeast with DNA. Methods Mol. Cell. Biol. **5:** 255–269.

GYURIS, J., E. GOLEMIS, H. CHERTKOV and R. BRENT, 1993 Cdi1, a human G1 and S phase protein phosphatase that associates with Cdk2. Cell **75:** 791–803.

HANNON, G. J., D. DEMETRICK and D. BEACH, 1993 Isolation of the Rb-related p130 through its interaction with CDK2 and cyclins. Genes Dev. **7:** 2378–2391.

HOFFMAN, C. S., and F. WINSTON, 1987 A ten-minute DNA preparation from yeast efficiently releases autonomous plasmids for transformation of *Escherichia coli*. Gene **57:** 267–272.

JACKSON, A. L., P. M. PAHL, K. HARRISON, J. ROSAMOND and R. A. SCLAFANI, 1993 Cell cycle regulation of the yeast Cdc7 protein kinase by association with the Dbf4 protein. Mol. Cell. Biol. **13:** 2899–2908.

JOHNSTON, M., S. ANDREWS, R. BRINKMAN, J. COOPER, H. DING *et al.*, 1994 Complete nucleotide sequence of *Saccharomyces cerevisiae* chromosome VIII. Science **265:** 2077–2082.

KALPANA, G. V., and S. P. GOFF, 1993 Genetic analysis of homomeric interactions of human immunodeficiency virus type 1 integrase using the yeast two-hybrid system. Proc. Natl. Acad. Sci. USA **90:** 10593–10597.

KOONIN, E. V., P. BORK and C. SANDER, 1994 Yeast chromosome III: new gene functions. EMBO J. **13:** 493–503.

LALO, D., C. CARLES, A. SENTENAC and P. THURIAUX, 1993 Interactions between three common subunits of yeast RNA polymerases I and III. Proc. Natl. Acad. Sci. USA **90:** 5524–5528.

LANGIN, T., G. FAUGERON, C. GOYON, A. NICOLAS and J. L. ROSSIGNOL, 1986 The *MET2* gene of *Saccharomyces cerevisiae:* molecular cloning and nucleotide sequence. Gene **49:** 283–293.

LEE, A. S., 1987 Coordinated regulation of a set of genes by glucose and calcium ionophore in mammalian cells. Trends Biochem. Sci. **12:** 20–23.

MILLER, J. H., 1972 *Experiments in Molecular Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

MURAKAMI, Y., M. NAITOU, H. HAGIWARA, T. SHIBATA, M. OZAWA *et al.*, 1995 Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*. Nature Genet. **10:** 261–268.

OLIVER, S. G., Q. J. M. VAN DER AART, M. L. AGOSTONI-CARBONE, M. AIGLE, L. ALBERGHINA *et al.*, 1992 The complete DNA sequence of yeast chromosome III. Nature **357:** 38–46.

ROSE, M. D., F. WINSTON and P. HIETER, 1990 *Methods in Yeast Genetics*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

SANCHEZ, Y., B. A. DESANY, W. J. JONES, Q. LIU, B. WANG *et al.*, 1996 Regulation of RAD53 by the ATM-like kinases MEC1 and TEL1 in yeast cell cycle checkpoint pathways. Science **271:** 357–360.

SANGER, F., S. NICKLEN and A. R. COULSON, 1977 DNA sequence

analysis with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74:** 5463–5467.

SIKORSKI, R. S., and P. HIETER, 1989 A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae.* Genetics **122:** 19–27.

STOTZ, A., and P. LINDER, 1990 The ADE2 gene from *Saccharomyces cerevisiae:* sequence and new vectors. Gene **95:** 91–98.

TAKACS, A. M., T. DAS and A. K. BANERJEE, 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:** 10375–10379.

VOJTEK, A. B., S. M. HOLLENBERG and J. A. COOPER, 1993 Mammalian Ras interacts directly with the serine/threonine kinase Raf. Cell **74:** 205–214.

YANG, X., E. J. HUBBARD and M. CARLSON, 1992 A protein kinase substrate identified by the two-hybrid system. Science **257:** 680–682.

YANG, X., R. JIANG and M. CARLSON, 1994 A family of proteins containing a conserved domain that mediates interaction with the yeast *SNF1* protein kinase complex. EMBO J. **13:** 5878–5886.

Communicating editor: M. JOHNSTON