Affinity Selection of DNA-Binding Proteins from Yeast Genomic DNA Libraries by Improved λ Phage Display Vector[1]

Hiroko Hagiwara,* Sumiko Kunihiro,* Keiichi Nakajima,* Motoaki Sano,* Haruhiko Masaki,† Midori Yamamoto,* Jeong Won Pak,* Yan Zhang,* Kumiko Takase,* Ichiro Kuwabara,‡,2 Ichiro N. Maruyama,‡ and Masayuki Machida*,3

Research Center for Glycoscience, Advanced Institute of Industrial Science and Technology (AIST), Central 6, 1-1-1, Higashi, Tsukuba, Ibaraki 305-8566; †Department of Biotechnology, Graduate School of Agriculture and Life Sciences, University of the Tokyo, Yayoi 1-1-1, Bunkyo-ku, Tokyo 113-8657; and ‡The Scripps Research Institute, 10550 North Torrey Pines Road, La Jolla, CA 92037, USA

Received August 19, 2002; accepted October 3, 2002

Phage display is a useful means of identifying and selecting proteins of interest that bind specific targets. In order to examine the potential of phage display for the genome-wide screening of DNA-binding proteins, we constructed yeast genomic libraries using λfoo-based vectors devised in this work. After affinity selection using GAL4 UAS_G as a probe, phages expressing GAL4 were enriched approximately 5×10^5-fold from the library. Approximately 90% of polypeptides encoded in correct translation reading frames by the selected phages were known or putative polynucleotide-binding proteins. This result clearly indicates that the modified lambda phage display vector in combination with our enrichment technique has great potential for the enrichment of DNA-binding proteins in a sequence-specific manner.

Key words: colicin E3, DNA-binding protein, GAL4, phage display, sequence-specific selection.

The genome sequencing projects have resulted in the accumulation of a wealth of protein sequences, many of which remain to be analyzed. The identification of interacting partners, such as other proteins and DNA, is crucial for the elucidation of protein function, and high-throughput methods are desired for such studies. The yeast one-hybrid (*1–3*) and two-hybrid (*4, 5*) methods have been used extensively, and may also be applied to the genome-wide characterization of DNA–protein and protein–protein networks. Because of significant false signals, however, these approaches may require laborious, time-consuming effort.

Phage display is a method for producing foreign proteins fused to the surface of a virus particle, and allows the expression, selection and subsequent amplification of the proteins encoded. The phenotypes of the displayed proteins can be directly linked to the gene sequence encoded by the phage, and enables effective screening of particular genes that encode proteins of interest (*6–8*). Several phage display systems have been described, including filamentous phage display through pIII (*6*), pVI (*9*), and pVIII (*10*).

Phage display has been extensively applied to the analysis of peptide-antibody interaction by using chemically synthesized oligonucleotides and fragmented DNA fused to the coat protein gene. Tailed bacteriophages such as lambda, T4 or T7 phage have also been used for the display of foreign proteins (*11–13*). Unlike the filamentous bacteriophages, these phages do not require protein secretion across the bacterial cell membrane. This is the great advantage of tailed bacteriophages for the effective display of both intracellular and extracellular proteins encoded on cDNA or genomic DNA fragments.

Among the surface display systems, λfoo is particularly attractive because the vector can express homo-multimeric proteins and because it has a conditional fusion scheme (*14, 15*). Using λfoo and other λ phage vectors, many functional prokaryotic and eukaryotic proteins have been produced on the surface (*16–18*). λfoo has also been applied to linear and conformational epitope mapping using antibodies immobilized in microtiter wells (*19, 20*), to protein-folding domain mapping using ligand-immobilized agarose beads (*21*), and to cDNA library screening of autoantigens using patient sera immobilized as ligands (*22*).

Phage display has also been utilized for the expression of DNA-binding proteins. To modify the sequence specificity of a DNA-binding domain with a zinc finger motif, the filamentous display system has been used (*23–25*). An artificially designed short oligopeptide consisting of 88 amino acid residues was expressed and served as an origin of a DNA-binding protein for *in vitro* evolution (*11, 23–25*). We have also successfully expressed functional transcription factors on the surface of λfoo (*26*). To our knowledge, however, there is no report describing the enrichment of DNA-binding proteins from a genomic or cDNA library in a

[1]This work was supported by the New Energy and Industrial Technology Development Organization (NEDO) of Japan.

[2]Present address: La Jolla Institute for Allergy and Immunology (LIAI), 10355 Science Center Drive, San Diego, CA 92121, USA.

[3]To whom correspondence should be addressed. Tel: +81-298-61-6164, Fax: +81-298-61-6174, E-mail: m.machida@aist.go.jp

Abbreviations: nt, nucleotide(s); aa, amino acid(s); pfu, plaque formation unit; MOI, multiplicity of infection; EMSA, electrophoretic mobility shift assay; MCS, multiple cloning site; CRD; COOH-terminal domain; PEG, polyethylene glycol; GST, glutathione S-transferase; QC-PCR, Quantitative, competitive PCR.

© 2002 by The Japanese Biochemical Society.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

sequence-specific manner. Therefore, the identification of DNA-binding proteins from genomic or cDNA libraries has remained elusive. Since many *trans*-acting DNA-binding factors, including GAL4, have homo-dimeric structures (*27, 28*), λfoo may be particularly suited for genome-wide screening of DNA binding proteins.

GAL4 is a yeast transcriptional activator of the *GAL* genes involved in the metabolism of galactose and related sugars (*29*). Like most activators, GAL4 contains a sequence-specific DNA binding domain, an N-terminal dimerization domain, and C-terminal activation domain (*30*). The DNA binding domain binds the 17-bp UAS_G (CGGAc/gGACa/tGTCg/cTCCG) of the genes involved in galactose metabolism. Previous studies revealed that only the N-terminal 74 amino acid residues of GAL4 specifically bind UAS_G (*31*). The N-terminal region contains the C_6 zinc binuclear cluster $[Zn(II)_2Cys_6]$, which is also known as the DNA-binding motif in fungi (*32, 33*). Our previous work shows that λfoo displaying the N-terminal 147 amino acid residues of GAL4 binds UAS_G in a sequence-specific manner (*26*).

In the present work, we describe the construction of shared yeast genomic DNA libraries as well as the successful selection of DNA binding proteins, including GAL4, using UAS_G as a probe. We also describe a modification of λfoo for efficient library construction, and of an affinity selection protocol for efficient enrichment of clones specifically interacting with the DNA probe.

MATERIALS AND METHODS

Bacteria and Yeast Strains—Bacterial strains, *Escherichia coli* JM105 (*27*), XL1-Blue (*34*), and Q447 (*22*), were used as hosts for plasmid transformation and phage transfection. A yeast strain, *Saccharomyces cerevisiae* S288c (*35*), was used for the construction of genomic DNA libraries.

Vector Construction—Lambda phage display vector, λfoo (*23*), was modified to accommodate two *Sfi*I recognition sites in the multiple cloning site (MCS) for efficient library construction (*36*) (Fig.1). The *Sfi*I–*Bam*HI segment of the λfoo MCS was amplified by PCR with a pair of oligonucleotide primers (5′-GCATGGTCTCGACCCCGACCACTCCCA and 5′-GCATGGATCCTCTAGAGTCGACC). The forward primer was designed to contain the recognition sequence (underlined) of the type IIS restriction enzyme *Bsa*I that cleaves DNA 8-nt away from its recognition site, and the reverse primer encodes a *Bam*HI site (underlined). The amplified fragment was digested with *Bsa*I and *Bam*HI, and ligated with two double-stranded synthetic oligonucleotides:

adaptor 1,
5′-GGTCCCACTCCG-3
3′-CAACCAGGGTGAGGCTGGG-5′;
and adaptor 2,
5′-GATCCCCGGGTACCGAGGCCGCCTCGGCCGAGCTCGAATTCGGCCGGCCATAGCGGCCGC
3′-GGGCCCATGGCTCCGGCGGAGCCGGCTCGAGCTTAAGCCGGCCGGTATCGCCGGCGTTAA

(Underlines indicate the *Sfi*I recognition sequences).

The *Sfi*I site at 9484 nt of the λfoo vector (*14*) was disrupted by the introduction of adaptor 1. The resulting DNA fragments were subsequently ligated with λfoo DNA digested with *Sfi*I and *Eco*RI, yielding λfooVSfi.

The colicin E3 COOH-terminal domain (CRD) (1363 to 1656 nt of the colicin E3 structural gene) (*37*) has a strong activity to kill *E. coli*. We have confirmed that the DNA fragment encoding colicin E3 CRD introduced with an amber stop codon at the 5′ terminus of the fragment (TAG-E3-CRD) can kill *E. coli* with suppressor activity (unpublished data). λfooVcoli was constructed by inserting the colicin DNA fragment between the two *Sfi*I sites of the λfooVSfi MCS. The λfooVcoli formed plaques when grown with a sup^o host, Q447, but failed to form plaques with XL1-Blue, a suppressor-positive strain. When the TAG-E3-CRD was replaced with a foreign DNA fragment, the λfooVcoli grew with a sup^+ host.

Construction of Yeast Genomic Libraries—Yeast genomic DNA was prepared from S288c cells by the standard procedure. Approximately 1 mg of genomic DNA was partially digested with 40–50 units of restriction enzyme, *Aci*I, *Msp*I, *Hin*PI, or *Taq*I, for 5–10 min at the manufacturer's recommended temperatures. The reaction was stopped by the addition of 50 μl of 0.5 M EDTA. Digested genomic DNA was electrophoresed in preparative 1.0% agarose gels for size fractionation. DNA fragments ranging from 0.5 to 2 kb in length were excised and purified by Ultrafree-MC (Millipore, Bedford, MA). Approximately 2 μg of the purified genomic DNA fragments were ligated with two *Sfi*I adaptors, which are complementary to the λfooVSfi arms generated by *Sfi*I digestion. Three 5′-end adaptors,

Sfi5AD0 5′-CGGCCAGGATCTGGTGG
3′-GGAGCCGGTCCTAGACCACCGC,
Sfi5AD1 5′-CGGCCAGGATCTGGTGGG
3′-GGAGCCGGTCCTAGACCACCCGC, or
and Sfi5AD2 5′-CGGCCAGGATCTGGTGGGG
3′-GGAGCCGGTCCTAGACCACCCCGC,
and 3′-end adaptor,
Sfi3AD 5′-CGCTGACTGACTGAGGCCATAG
3′-GACTGACTGACTCCGGT,

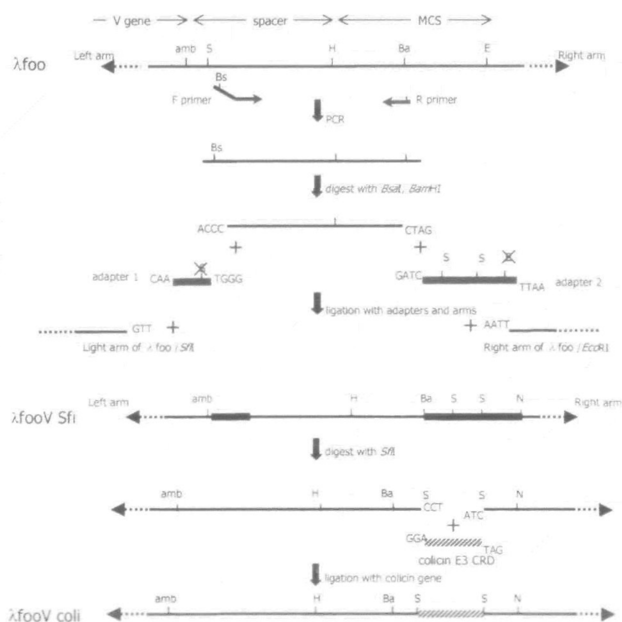

Fig. 1. **Construction of λfooVSfi and λfooVcoli vectors.** Restriction sits are S, *Sfi*I; H, *Hin*dIII; Ba, *Bam*HI; E, *Eco*RI; Bs, *Bsa*I; N, *Not*I.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

were also prepared. The three 5′-end adaptors were designed to differ in length from each other by one nucleotide, so that one of the three adaptors would match the translation reading frame of each of the genomic DNA inserts. The 3′ adaptor Sfi3AD was designed to have translation termination codons in all three frames.

The three different 5′ adaptors were ligated with genomic DNA digested by the four enzymes for twelve ligation reactions in total. Unligated adaptors were removed from the genomic DNA fragments by passing the ligation mixture through an S-400 (Amersham Biosciences, Uppsala, Sweden) spin column, and then through a cDNA Size Fractionation Column (Invitrogen, Carlsbad, CA). The genomic DNA was also fractionated by 1.0% agarose electrophoresis. Purified DNA fragments were ligated with 0.5 μg of λfooVcoli and λfooVSfi DNA digested with *Sfi*I. The λfooVcoli vector was used for library construction of all four genomic DNA preparations digested with *Aci*I, *Hin*PI, *Msp*I or *Taq*I; the λfooVSfi vector was only used for library construction with DNA digested with *Taq*I. After incubating 2 h at 25°C, the ligated DNAs were packaged using Giga Pack III Gold (Stratagene, La Jolla, CA), and used to infect *E. coli* Q447. Restriction digestion, adaptor ligation and phage DNA packaging were performed by standard procedures as previously described (*34*).

Preparation of Probe DNA—A DNA fragment was prepared as a probe for library affinity selection by the following procedure. PCR primers, 5′-CGCCAGGGTTTTCCCAGTCACGAC biotinylated at the 5′ end and 5′-AGCGGATAACAATTTCACACAGGAAAC, were to amplify a 550-bp fragment containing eight repeats of the GAL4 binding consensus sequence (*22*). The amplified fragments were purified with an S-400 spin column and by successive ethanol precipitations. Seventy micrograms of the PCR product and 2.3 mg of streptavidin-coated paramagnetic beads (GenoPrep streptavidin beads, Genovision, Oslo, Norway) were mixed in BW buffer [1 M NaCl, 10 mM Tris-HCl (pH 8.0), 1 mM EDTA] and incubated for 15 minutes at room temperature. Unbound probe DNA was removed by washing the beads twice with BW buffer.

Affinity Selection of Genomic Libraries—Affinity selection procedures are illustrated in Fig. 2. Phage libraries were cultured using XL1-Blue as a host. After complete lysis of the host bacteria, DNase I was added to the lysate at a final concentration of 10 units per 100 ml, and the lysate was incubated at 37°C for 20 min. The phages were collected by centrifugation at 10,000 ×*g* for 10 min after the addition of 6% (w/v) polyethylene glycol (PEG 8000; SIGMA, St. Louis, MO) and 0.6 M NaCl. The precipitated phages were resuspended in SM buffer, and precipitated again in the presence of 3.3% PEG 8000 and 0.4 M NaCl. Finally, the phages were resuspended in binding buffer [20 mM HEPES (pH 7.5), 0.5 mM $MgCl_2$, 0.1 mM EDTA, 0.1 μM $ZnCl_2$, 100 mM NaCl, 1 mM DTT, 2.5% glycerol, and 0.1% Tween-20] supplemented with 1× protease inhibitor cocktail (SIGMA). Enrichment with the immobilized probe DNA was performed in a volume of 50–200 μl containing ~10^{12} plaque-forming units (pfu) of phages, 500 μg of paramagnetic beads coated with probe DNA, 0.1 mg/ml of poly(dI-dC) and 1× protease inhibitor cocktail in binding buffer. After a 1-h incubation with agitation at room temperature, the beads were separated from the supernatant with a magnet and washed three times with binding buffer

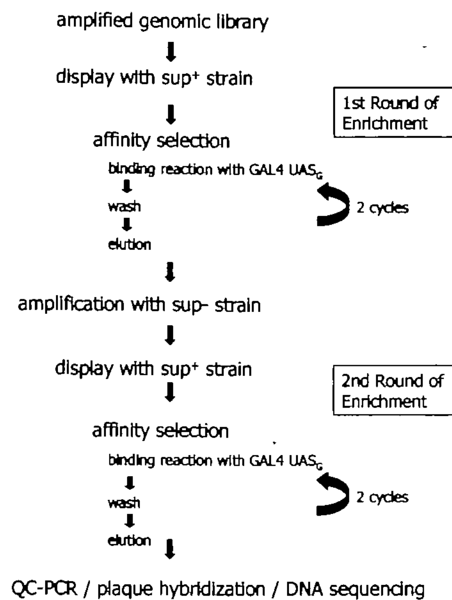

Fig. 2. **Affinity selection scheme for the enrichment of GAL4 clones using UAS$_G$ as a probe.** UAS$_G$ fragments were immobilized on paramagnetic beads as described in "MATERIALS AND METHODS." The library was selected twice without amplification between the two enrichment steps. The enriched library was amplified by infecting a bacterial host, and then further enriched two more times.

containing 200 mM NaCl. Bound phages were eluted from the beads by incubating in 100 μl of elution buffer (SM buffer containing 1 M NaCl and 0.1% Tween-20) at room temperature for 1 h. The eluted phages (~10^{7-8} pfu) were precipitated by the addition of 5% PEG 8000 and 0.5 M NaCl (final concentrations), and suspended in 5 μl of binding buffer. The second binding reaction was carried out with 50 μg of paramagnetic beads coated with probe DNA in a volume of 10 μl. The second enrichment was done using the same protocol as in the first enrichment, except that the phages were eluted with 50 μl of elution buffer. A small portion of the eluted phages was used for competitive PCR to estimate the amount of the *GAL4* gene encoded by the enriched library as well as for plate lysate preparation, which was subjected to the second round of enrichment. The second round of affinity selection was carried out using 1/10 the scale of the first round. GAL4-encoding phages in the eluate were detected by plaque hybridization with a DNA probe encompassing 21–620 nt from the translation initiation codon of the *GAL4* gene.

Quantitative, Competitive PCR (QC-PCR)—A pair of oligonucleotide primers, 5′-GCCGACTTAAAAAGCTCAAG (forward) and 5′-TTAACAATGCTTTTATATCC (reverse), were designed to amplify the 240-bp fragment of GAL4 (41–280 nt from the translation initiation codon), which encodes the core region of the GAL4 DNA-binding domain, and was used for competitive PCR. DNA competitive to the GAL4 target was designed to have a 20-bp deletion (241–260 nt from the translation initiation codon), prepared by PCR, and cloned to the pGEM-T plasmid vector (Promega, Madison, MI). Genomic phage DNA was prepared by phenol/chloroform extraction and successive ethanol precipitation. A QC-PCR mixture (20 μl) consisted of 1 μM primers, 0.2 mM dNTPs, genomic DNA derived from 10^2–10^7 pfu of

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

phages, 0.0001–1 amol of the competitor plasmid and 5 units of ExTaq DNA polymerase (Takara, Kyoto). Thermocycling consisted of 94°C for 2 min, followed by 35 cycles of denaturation (94°C; 20 s), annealing (55°C; 30 s), and extension (72°C; 60 s). Two products with the expected sizes (220 and 240 bp) were analyzed by polyacrylamide gel electrophoresis (PAGE), and stained with ethidium bromide.

Plaque Hybridization—Plaque hybridization was carried out using a standard protocol (*34*). Phage DNA cross-linked to nylon membranes was hybridized overnight at 42°C with a DIG-labeled probe that encodes the 5′-terminal end (21–620 nt from the initiation codon) of the *GAL4* structural gene.

Expression of GAL4 Protein and Electrophoretic Mobility Shift Assay—After affinity selection, genomic DNA fragments encoded by phages were amplified by PCR using primers, 5′-CGATTGAATTCAATCGAACAAGCATGCGATATTTGCCGACT and 5′-ATAGTTTAGCGGCCGCTTCGAAGAATACATAAGAGAGAACCGTCG, and cloned into a pGEM-4T-3 module (Amersham Biosciences) that can produce proteins encoded by the genomic insert in a fused form with glutathione S-transferase (GST). The fusion protein was expressed in *E. coli* strain BL21 cells and purified by a MicroSpin GST purification module (Amersham Biosciences). Double-stranded oligonucleotides were chemically synthesized and annealed for electrophoretic mobility shift assay (EMSA). IRDye800-labeled 60 bp in length probes, 5′-GTACGGATTAGAAGCCGCCGAGCGGGTGACAGCCCTCCGAAGGAAGACTCTCCTCCGTGC (GalI-IIIu) and 5′-GCACGGAGGAGAGTCTTCCTTCGGAGGGCTGTCACCCGCTCGGCGGCTTCTAATCCGTAC (GalI-IIIb) encoded three GAL4-binding consensus sequences (UAS_G, underlined) of the *GAL1* promoter. The coding region of the *A. oryzae pgmA* gene (191–240 nt) was used as a non-specific competitor. The GST-GAL4 fusion protein and probes were incubated in binding buffer for 20 min at room temperature. The competitor DNA was mixed with the probe DNA prior to incubation with the fused proteins for competition assay. EMSAs for the analysis of DNA-protein complexes were carried out as described by Sano *et al.* (*38, 39*).

Phage Culture, Plaque Staining and DNA Sequencing—The general manipulation of phages has been described previously (*34*). For phage DNA sequencing, plaque PCR was carried out using a pair of oligonucleotide primers, 5′-CGACTCCCACCCCGACC (TsrifooF) and 5′-GAAAGGGGGATGTGCTG (TsrifooR). The PCR product was treated with a pre-sequencing kit (USB, Cleaveland, OH) and subjected to sequencing reactions using an IRDye-labeled primer (u-70; IRDye800-5′-ACCCCGACCCCGACTCCGACCGTT) and a Thermo sequenase cycle sequence kit (USB). Sequence analysis was done using a LI-COR Model 4200L DNA sequencer (LI-COR, Lincoln, NE).

RESULTS

Construction of Yeast Genomic Libraries—Shared yeast genomic DNA was prepared according to the method of James *et al.* (*40*). There are four commercially available enzymes, *Aci*I, *Hin*P1I, *Msp*I, and *Taq*I, with a 4-bp recognition sequence that produces a common 5′ overhang with 5′-CG-3′. As previously predicted (*41*), the CG enzyme recognition sites exist every 128 bp, on average, while *Sau*3AI alone, used in most library constructions, cuts genomic DNA once every 373 bp. The shared yeast genomic DNA digested with the four CG enzymes was ligated with one of the three 5′ adaptors (refer to "MATERIALS AND METHODS") and the 3′ adaptor, and cloned into λfooVcoli digested with *Sfi*I to produce fooVcoli-A (A1–A3), fooVcoli-H (H1–H3), fooVcoli-M (M1–M3), and fooVcoli-T (T1–T3) libraries (Table I). The overall complexity of the libraries constructed was calculated to be 6.6×10^5; namely, 14×10^6-bp yeast genome in length/128-bp restriction fragments on average ×2 directions ×3 reading frames. *Taq*I-digested genomic DNA fragments were ligated with the three 5′ adaptors, Sfi5AD1, Sfi5AD2, and Sfi5AD3, and were used to construct λfooVSfi libraries, resulting in three libraries designated as λfooVSfi-T1, -T2, and -T3 , respectively.

The libraries constructed with λfooVcoli and λfooVSfi consisted of independent recombinants, the numbers of which range from 3.40×10^5 to 1.75×10^6 pfu, or from 1.25×10^6 to 3.75×10^6 pfu, respectively (Table I). The libraries constructed using λfooVSfi contained 30–50% insert-harboring phages. The twelve libraries were amplified separately by infecting a sup^0 *E. coli* host, Q447. Roughly the same pfu of the three libraries derived from the different adaptors for each restriction enzyme were mixed together, yielding a restriction enzyme-specific library. The resulted four libraries, fooVcoli-A, fooVcoli-H, fooVcoli-M, and fooV-

TABLE I. **Number of independent recombinants in genomic DNA libraries constructed.**

Enzyme	Vector	Library		No. of independent recombinants[a]	White plaque (%)
*Aci*I	λfooVcoli	fooVcoli-A	fooVcoli-A1	3.40×10^5	
			fooVcoli-A2	5.50×10^5	
			fooVcoli-A3	8.80×10^5	
*Hin*PI	λfooVcoli	fooVcoli-H	fooVcoli-H1	4.90×10^5	
			fooVcoli-H2	1.20×10^6	
			fooVcoli-H2	1.17×10^6	
*Msp*I	λfooVcoli	fooVcoli-M	fooVcoli-M1	5.45×10^5	
			fooVcoli-M2	1.75×10^6	
			fooVcoli-M3	1.12×10^6	
*Taq*I	λfooVcoli	fooVcoli-T	fooVcoli-T1	1.90×10^6	
			fooVcoli-T2	9.60×10^5	
			fooVcoli-T3	7.00×10^5	
	λfooVSfi	fooVSfi-T	fooVSfi-T1	3.75×10^6 (7.40×10^6)	33.6
			fooVSfi-T2	1.25×10^6 (1.35×10^6)	48.1
			fooVSfi-T3	2.25×10^6 (2.30×10^6)	49.5

[a]The numbers of independent recombinants were estimated from the white plaques of titration of packaged libraries described in "MATERIALS AND METHODS." The numbers of vector phages forming blue plaques on plates containing IPTG are indicated in parentheses.

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

TABLE II. **Affinity selection of yeast genomic DNA libraries.**
(a) *Taq*I library

Library		Phage titer (pfu)		GAL4 DNA (fmol) or clone (%)[a]	Enrichment factor[b]
			White plaque (%)		
fooVcoli-T	1st round input	7.00×10^{11}		0.0086 fmol	
	Recovered	2.70×10^{5}		11.1 fmol	1.3×10^{3}
	2nd round input	1.63×10^{10}		0.9%	
	Recovered	1.08×10^{7}		31.5%	4.9
fooVSfi-T	1st round input	1.40×10^{12}	6.8	0.00043 fmol	
	Recovered	5.50×10^{4}	18.2	18.2 fmol	4.2×10^{4}
	2nd round input	3.40×10^{10}	10.1	2.0%	
	Recovered	1.31×10^{7}	28.7	50.8%	25.4

(b) CG library

λfooVcoli	Phage titer (pfu)	GAL4 DNA (fmol) or clone (%)[a]	Enrichment factor[b]
1st round input	2.10×10^{12}	0.00561 fmol	
Recovered	3.14×10^{6}	95.5 fmol	1.7 × 104
2nd round input	1.44×10^{10}	0.1%	
Recovered	1.02×10^{7}	5.7%	55.7

[a]In the first round of enrichment, the amount of GAL4 DNA in the libraries was estimated by competitive PCR as described in "MATERIALS AND METHODS" and is expressed as the amount (fmol) of DNA per 10^{10} library phages (pfu). In the second round of selection, the proportion (%) of GAL4 phage in the libraries was determined by plaque hybridization. [b]Enrichment factors were calculated from the ratios between recovered and input GAL4 phages.

TABLE III. **Number of GAL4 clones recovered from the libraries.**

Library	fooVcoli-T		fooVSfi-T	
GAL4 peptide	7–146 aa	7–206 aa	7–146 aa	7–206 aa
No. of clones	17 (85%)	3 (15%)	24 (89%)	3 (11%)

coli-T, were separately cultivated in XL1-Blue cells. Translation termination at the amber stop codon between the phage *V* gene and yeast genomic DNA insert prevents foreign proteins from being displayed on the phage surface in this suppressor-negative host. Therefore, the library amplification with Q447 minimizes biased growth of the recombinant phages in the library. Nonetheless, the proportion of phages with inserts decreased to 10–20% of the library after amplification, when the host was infected with the library phages at a multiplicity of infection (MOI) of 0.01. When a suppressor-positive strain, XL1-Blue, was infected with libraries at MOI of 10–100, the ratios of recombinants to non-recombinants further decreased to 10% or less as shown in Table IIa.

Comparison of λfooVcoli and λfooVSfi Vectors—To evaluate the λfooVcoli vector, which has a positive selection scheme for recombinants based upon the colicin gene toxic to *E. coli*, for the construction and affinity selection of libraries, two libraries, fooVcoli-T and fooVSfi-T, were constructed using the λfooVcoli and λfooVSfi vectors, respectively. *S. cerevisiae* genomic DNA was partially digested with *Taq*I, and 0.3–1.0-kb fragments were cloned into the vectors. *Taq*I digestion can produce a DNA fragment containing the entire GAL4 DNA-binding domain. The *E. coli* strain XL1-Blue was used as a host cell having a strong amber suppressor activity for the display of yeast proteins on the phage surface. The library phages, 5×10^{11} pfu, cultivated in XL1-Blue were incubated with 500 μg of paramagnetic beads coated with the GAL4-binding consensus sequence. Bound phages were eluted, precipitated with PEG, and subjected to a second affinity selection without amplification as described in "MATERIALS AND METHODS." In the second selection, the eluted phages, ~10^{7}–10^{8} pfu, were incubated with 50 μg of probe DNA-immobilized paramagnetic beads. After the two successive affinity selections, 10^{4}–10^{6} pfu of phages were recovered. These results are shown in Table IIa.

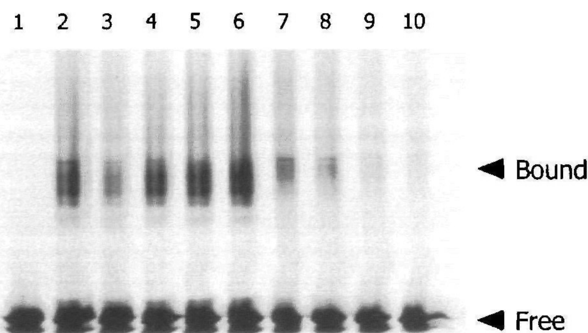

Fig. 3. **Electrophoretic mobility shift assay of isolated polypeptides.** Approximately 0.1 μg of the GST-fused GAL4BD-L expressed from the DNA fragment of isolated phages was mixed with 120 fmol of the DNA probe, and incubated for 20 min at room temperature (lane 1). Non-specific competitor (lane 2–5) or specific competitor (lane 6–9) DNA was added to the binding reaction. Either 500 fmol (lanes 2 and 6), 1 pmol (lanes 3 and 7), 5 pmol (lanes 4 and 8), or 10 pmol (lanes 5 and 9) of the competitor DNA was added. Construction of the competitor DNA fragments is described in "MATERIALS AND METHODS."

Before and after the affinity selection, the amount of phage DNA encoding GAL4 in the library was quantified by QC-PCR using primers specific for the GAL4 DNA-binding domain. *GAL4* DNA, 8.6 amol, in 10^{10} pfu of library phages was detected in the library amplified with XL1-Blue. After the first round of the selection, the amount of GAL4 DNA was 11.1 fmol in 10^{10} pfu of the selected library phage. This result indicates that the GAL4 phages were enriched 10^{3}–10^{4}-fold by the two-step affinity selection.

The GAL4 phage clone was more efficiently enriched for from the fooVSfi-T library than from the fooVcoli-T library,

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

TABLE IV. **Gene products encoded by phage clones isolated from the CG library.**

Gene product	Frequency	Fusion site	Biochemical function
TBS1	3	*Msp*I	Putative transcription factor of the $Zn(II)_2Cys_6$ domain
GAL4	2	*Taq*I	Zinc finger transcription factor of the $Zn(II)_2Cys_6$ binuclear cluster domain
SNM1	2	*Msp*I	Ribonuclease MRP, RNA-binding protein with a zinc-cluster domain
RLF2	2	*Taq*I	Chromatin Assembly Complex
SMB1	1	*Aci*I	mRNA splicing
BOI2	1	*Aci*I	RHO protein signal transduction
SOK1	1	*Hin*PI	Transcription factor (putative)

as shown in Table IIa. This may be due to the initial low content of the GAL4 clone in the fooVSfi library, which has an approximately 10-fold larger number of non-recombinant clones than the fooVcoli library. Therefore, the efficiency of the affinity selection is similar in both libraries.

Phages, ~10^4 pfu, from the first round of selection were amplified and subjected to a second affinity selection. Plaque hybridization analysis showed that the second round enriched the GAL4 phage only 10- to 100-fold, probably due to a high proportion of the GAL4 clone in the eluate population after the first round (Table IIa). The GAL4-phage content in the selected population was ~50% after the two successive selections, and did not significantly depend on the presence of the colicin gene in the vector. DNA sequence analysis of 32 randomly selected clones in the eluate of the second selection revealed that 63% (20 clones) of the fooVcoli population and 84% (27 clones) of fooVSfi population encoded genomic DNA fragments corresponding to the N-terminal regions, 7–146 aa (GAL4BD-S) or 7–206 aa (GAL4BD-L), of the GAL4 protein (Table III). Both of the polypeptides contained the N-terminal 74 residues required for GAL4 dimerization and for GAL4 binding to UAS_G. EMSA using the polypeptide corresponding to residues 7–206 of GAL4 confirmed the sequence-specific binding of the enriched polypeptide to UAS_G (Fig. 3).

We also tried the affinity selection of GAL4 clones after diluting the T1 library 10- or 100-fold with the T2 and T3 libraries, which did not contain clones displaying functional GAL4. The GAL4 phage clones could be effectively enriched for from both of the diluted libraries, demonstrating that less abundant DNA-binding proteins than GAL4 can be specifically enriched by the method described here (data not shown).

Affinity Selection of Yeast Genomic Libraries—Using the λfooVcoli vector, 12 different libraries were prepared from genomic DNA fragments digested with four restriction enzymes, *Aci*I, *Msp*I, *Hin*PI, or *Taq*I, each of which was ligated with the three different adaptors as described in "MATERIALS AND METHODS." The four libraries (fooVcoli-A, fooVcoli-H, fooVcoli-M, fooVcoli-T) were cultivated separately in XL1-Blue cells, and mixed together to give roughly the same pfu of phages from each library. When ~10^{12} pfu library phages were subjected to the two successive affinity selections without amplification in the first round, the GAL4-phage concentration reached 5.61 amol of GAL4 DNA in 10^{10} pfu library phages (Table IIb). The recovered phages from the first round of selection were amplified and subjected to the second round of selection. Plaque hybridization using a DNA fragment encoding the $Zn(II)_2Cys_6$ domain as a probe showed that GAL4-encoding phages occupied 5.7% of the library. When 39 clones were randomly analyzed by DNA sequencing, we found that two clones encoded the GAL4BD-S domain of GAL4, and that all other clones except one encoded proteins that have or are predicted to have polynucleotide-interacting activities (Table IV).

DISCUSSION

In the present study, using the λfoo display system, we have successfully enriched and isolated GAL4-encoding clones from yeast genomic DNA libraries through affinity selection with the GAL4-binding consensus sequence, UAS_G. A number of factors, such as protein instability and fusion through incorrect reading frames, influence the expression of functional proteins on the vector surface. In particular, the availability of DNA fragments encoding proteins of interest is most critical. We adopted four CG enzymes that were used to construct the yeast genomic library for the two-hybrid selection (*40*). The CG enzymes provide 2.9-fold more potential fusion sites than a single enzyme such as *Sau*3AI (*41*). Large DNA fragments are also less likely to express functional proteins than small fragments. Therefore, we purified genomic DNA 0.2–1.0 kb in size after the enzyme digestion.

Affinity selection procedures are also critical for successful enrichment. We have designed a protocol that minimizes the number of library amplification steps. This may decrease the possibility of losing phage clones of interest during the selection steps. An *E. coli* strain with no suppressor activity was used for library amplification, and this reduces the loss of clones from the library that expresses proteins unfavorable for phage assembly. Nonetheless, the proportion of recombinant phages with inserts decreased to 10% or less in the amplified λfooVSfi libraries (Table II).

The concentration of GAL4 clones in the fooVcoli library was more than 10-fold higher than that in the fooVSfi library cultivated with a suppressor-positive host before the affinity selection (Table IIa). This is because the λfooVcoli vector without an insert cannot grow in sup^+ hosts, while recombinant clones with the insertion of a DNA fragment, which disrupts the colicin gene, can grow. However, no significant difference was observed in the enrichment of GAL4 clones in the two libraries constructed with λfooVcoli and λfooVSfi (Table IIa).

A repertoire of approximately 10^6 genomic DNA fragments are estimated to be required for the complete coverage of all protein domains encoded in the yeast genome, which is 14 Mb in size. In the case of higher eukaryotes with a larger genome size than yeast, an even larger repertoire is required. Moreover, the analysis of cDNA may also require a large number of cDNA fragments, particularly for cDNA encoding low abundant mRNA. In this work, phage lysates consisting of ~10^{11} infectious phage particles were used for the first screening. In the phage lysate, non-recombinant phages, namely vectors without inserts, occupied

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

~90% of the population of the fooVSfi library. When libraries were made with λfooVcoli, such non-recombinant vector clones could not grow in a suppressor-positive host during library amplification. Therefore, 10-fold more library phage particles can be subjected to screening, and this may be more efficient for successful enrichment than libraries made with λfooVSfi. Although we did not examine this possibility in this work, this favorable character of λfooVcoli may be particularly useful for the analysis of genomes of organisms higher than yeast.

DNA sequence analysis of isolated clones from the libraries made by CG enzyme digestion revealed that 11 out of 42 clones chosen at random had inserts fused in a correct translation reading frame (Table IV). The most frequently enriched polypeptides contained a region responsible for polynucleotide binding, such as $Zn(II)_2Cys_6$ in TBS1 and GAL4, or the zinc cluster domain in SMN1. TBS1, a putative DNA-binding transcription factor, was also isolated from the *Taq*I library (data not shown). The TBS1 clones contained the entire $Zn(II)_2Cys_6$ motif that is known to recognize symmetrically disposed CGG triplets with 6–10 or 18 nt spacing between them. This suggests the similarity of the recognition sequence of TBS1 to that of GAL4. All the proteins except for BIO2 are known or are predicted to have affinity for polynucleotides. Furthermore, the GAL4 polypeptides enriched in the present experiment were confirmed to have UAS_G-binding activity (Fig. 3). These results indicate that phage clones encoding polypeptides having affinity for DNA are specifically enriched from the yeast library.

The results presented here indicate that the enrichment of phage clones is achieved through the DNA-binding activity of the displayed polypeptide with a correct folding domain. Recently, an alternative vehicle using yeast cells has been developed for the display of proteins with eukaryotic posttranslational modifications (*42*). However, the yeast surface display is not suitable for the display of intracellular proteins (*43*). The λfooVcoli phage display, although it lacks eukaryotic posttranslational modification systems, is able to enrich many different polypeptides, regardless of the cellular localization of the proteins. Thus, the phage display system described in this work could be used effectively in the genome-wide analysis of DNA-binding proteins, possibly, of proteins with an affinity for other ligands, in eukaryotes.

REFERENCES

1. Luo, Y., Vijaychander, S., Stile, J., and Zhu, L. (1996) Cloning and analysis of DNA-binding proteins by yeast one-hybrid and one-two-hybrid systems. *Biotechniques* **20**, 564–568
2. Vidal, M., Brachmann, R.K., Fattaey, A., Harlow, E., and Boeke, J.D. (1996) Reverse two-hybrid and one-hybrid systems to detect dissociation of protein-protein and DNA-protein interactions. *Proc. Natl. Acad. Sci. USA* **93**, 10315–10320
3. Lemercier, C., To, R.Q., Swanson, B.J., Lyons, G.E., and Konieczny, S.F. (1997) Mist1: a novel basic helix-loop-helix transcription factor exhibits a developmentally regulated expression pattern. *Dev. Biol.* **182**, 101–113
4. Fromont-Racine, M., Rain, J.C., and Legrain, P. (1997) Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat. Genet.* **16**, 277–282
5. Legrain, P., and Selig, L. (2000) Genome-wide protein interaction maps using two-hybrid systems. *FEBS Lett.* **480**, 32–36
6. Crameri, R. and Suter, M. (1993) Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* **137**, 69–75
7. Lang, I.M., Chuang, T.L., Barbas, C.F., III, and Schleef, R.R. (1996) Purification of storage granule protein-23. A novel protein identified by phage display technology and interaction with type I plasminogen activator inhibitor. *J. Biol. Chem.* **271**, 30126–30135
8. Light, J., Maki, R., and Assa-Munt, N. (1996) Expression cloning of cDNA by phage display selection. *Nucleic Acids Res.* **24**, 4367–4368
9. Jespers, L.S., Messens, J.H., De Keyser, A., Eeckhout, D., Van, d.B., I, Gansemans, Y.G., Lauwereys, M.J., Vlasuk, G.P., and Stanssens, P.E. (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology (NY)* **13**, 378–382
10. Duncan, P.M., and Robins, S.P. (2001) External surface display of proteins linked to DNA-binding domains. *Analytical Biochemistry* **294**, 108–117
11. Jespers, L.S., De Keyser, A., and Stanssens, P.E. (1996) LambdaZLG6: a phage lambda vector for high-efficiency cloning and surface expression of cDNA libraries on filamentous phage. *Gene* **173**, 179–181
12. Sche, P.P., McKenzie, K.M., White, J.D., and Austin, D.J. (1999) Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol.* **6**, 707–716
13. Yamamoto, M., Kominato, Y., and Yamamoto, F. (1999) Phage display cDNA cloning of protein with carbohydrate affinity. *Biochem. Biophys. Res. Commun.* **255**, 194–199
14. Maruyama, I.N., Maruyama, H.I., and Brenner, S. (1994) Lambda foo: a lambda phage vector for the expression of foreign proteins. *Proc. Natl. Acad. Sci. USA* **91**, 8273–8277
15. Mikawa, Y.G., Maruyama, I.N., and Brenner, S. (1996) Surface display of proteins on bacteriophage lambda heads. *J. Mol. Biol.* **262**, 21–30
16. Dunn, I.S. (1995) Assembly of functional bacteriophage lambda virions incorporating C-terminal peptide or protein fusions with the major tail protein. *J. Mol. Biol.* **248**, 497–506
17. Santini, C., Brennan, D., Mennuni, C., Hoess, R.H., Nicosia, A., Cortese, R., and Luzzago, A. (1998) Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. *J. Mol. Biol.* **282**, 125–135
18. Sternberg, N. and Hoess, R.H. (1995) Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. USA* **92**, 1609–1613
19. Kuwabara, I., Maruyama, H., Mikawa, Y.G., Zuberi, R.I., Liu, F.T., and Maruyama, I.N. (1997) Efficient epitope mapping by bacteriophage lambda surface display. *Nat. Biotechnol.* **15**, 74–78
20. Kuwabara, I., Maruyama, H., Kamisue, S., Shima, M., Yoshioka, A., and Maruyama, I.N. (1999) Mapping of the minimal domain encoding a conformational epitope by lambda phage surface display: factor VIII inhibitor antibodies from haemophilia A patients. *J. Immunol. Methods* **224**, 89–99
21. Moriki, T., Kuwabara, I., Liu, F.T., and Maruyama, I.N. (1999) Protein domain mapping by lambda phage display: the minimal lactose-binding domain of galectin-3. *Biochem. Biophys. Res. Commun.* **265**, 291–296
22. Niwa, M., Maruyama, H., Fujimoto, T., Dohi, K., and Maruyama, I.N. (2000) Affinity selection of cDNA libraries by lambda phage surface display. *Gene* **256**, 229–236
23. Cheng, X., Kay, B.K., and Juliano, R.L. (1996) Identification of a biologically significant DNA-binding peptide motif by use of a random phage display library. *Gene* **171**, 1–8
24. Choo, Y. and Klug, A. (1994) Toward a code for the interactions of zinc fingers with DNA: selection of randomized fingers displayed on phage. *Proc. Natl. Acad. Sci. USA* **91**, 11163–11167
25. Rebar, E.J. and Pabo, C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science* **263**, 671–673

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012

26. Zhang, Y., Pak, J.W., Maruyama, I.N., and Machida, M. (2000) Affinity selection of DNA-binding proteins displayed on bacteriophage lambda. *J. Biochem.* **127**, 1057–1063
27. Ellenberger, T.E., Brandl, C.J., Struhl, K., and Harrison, S.C. (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* **71**, 1223–1237
28. Luisi, B.F., Xu, W.X., Otwinowski, Z., Freedman, L.P., Yamamoto, K.R., and Sigler, P.B. (1991) Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* **352**, 497–505
29. Laughon, A. and Gesteland, R.F. (1982) Isolation and preliminary characterization of the *GAL4* gene, a positive regulator of transcription in yeast. *Proc. Natl. Acad. Sci. USA* **79**, 6827–6831
30. Keegan, L., Gill, G., and Ptashne, M. (1986) Separation of DNA binding from the transcription-activating function of a eukaryotic regulatory protein. *Science* **231**, 699–704
31. Giniger, E., Varnum, S.M., and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell* **40**, 767–774
32. Pan, T. and Coleman, J.E. (1990) GAL4 transcription factor is not a "zinc finger" but forms a $Zn(II)_2Cys_6$ binuclear cluster. *Proc. Natl. Acad. Sci. USA* **87**, 2077–2081
33. Pfeifer, K., Kim, K.S., Kogan, S., and Guarente, L. (1989) Functional dissection and sequence of yeast HAP1 activator. *Cell* **56**, 291–301
34. Sambrook, J., Fritsch, E.F., and Manuatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
35. Murakami, Y., Naitou, M., Hagiwara, H., Shibata, T., Ozawa, M., Sasanuma, S., Sasanuma, M., Tsuchiya, Y., Soeda, E., and Yokoyama, K. (1995) Analysis of the nucleotide sequence of chromosome VI from *Saccharomyces cerevisiae*. *Nat. Genet.* **10**, 261–268
36. Christian, R.B., Zuckermann, R.N., Kerr, J.M., Wang, L., and Malcolm, B.A. (1992) Simplified methods for construction, assessment and rapid screening of peptide libraries in bacteriophage. *J. Mol. Biol.* **227**, 711–718
37. Ohno, S., Ohno-Iwashita, Y., Suzuki, K., and Imahori, K. (1977) Purification and characterization of active component and active fragment of colicin E3. *J. Biochem.* **82**, 1045–1053
38. Sano, M., Ohyama, A., Takase, K., Yamamoto, M., and Machida, M. (2001) Electrophoretic mobility shift scanning using an automated infrared DNA sequencer. *Biotechniques* **31**, 1056–1058, 1060, 1062
39. Toda, T., Sano, M., Honda, M., Rimoldi, O.J., Yang, Y., Yamamoto, M., Takase, K., Hirozumi, K., Kitamoto, K., Minetoki, T., Gomi, K., and Machida, M. (2001) Deletion analysis of the enolase gene (*enoA*) promoter from the filamentous fungus *Aspergillus oryzae*. *Curr. Genet.* **40**, 260–267
40. James, P., Halladay, J., and Craig, E.A. (1996) Genomic libraries and a host strain designed for highly efficient two-hybrid selection in yeast. *Genetics* **144**, 1425–1436
41. Arnold, J., Cuticchia, A.J., Newsome, D.A., Jennings, W.W., III, and Ivarie, R. (1988) Mono- through hexanucleotide composition of the sense strand of yeast DNA: a Markov chain analysis. *Nucleic Acids Res.* **16**, 7145–7158
42. Boder, E.T. and Wittrup, K.D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology* **15**, 553–557
43. Boder, E.T. and Wittrup, K.D. (2000) Yeast surface display for directed evolution of protein expression, affinity, and stability. *Methods Enzymol.* **328**, 430–444

Downloaded from http://jb.oxfordjournals.org/ at Islamic Azad University on September 29, 2012