Construction and Application of a Yeast Surface-displayed Human cDNA Library to Identify Post-translational Modification-dependent Protein-Protein Interactions*

Scott Bidlingmaier and Bin Liu‡

Although post-translational modifications such as phosphorylation mediate fundamental biological processes within the cell, relatively few methods exist that allow proteome-wide identification of proteins that interact with these modifications. We constructed a yeast surface-displayed human cDNA library and utilized it to identify protein fragments with affinity for phosphorylated peptides derived from the major tyrosine autophosphorylation sites of the epidermal growth factor receptor or focal adhesion kinase. We identified cDNAs encoding the Src homology 2 domains from adapter protein APS, phosphoinositide 3-kinase regulatory subunit 3, SH2B, and tensin, demonstrating the effectiveness of this approach. Our results suggest that large libraries of functional human protein fragments can be efficiently displayed on the yeast surface. In addition to the analysis of post-translational modifications, yeast surface-displayed human cDNA libraries have many potential applications, including identifying targets and defining potential cross-reactive proteins for small molecules or drugs. Molecular & Cellular Proteomics 5:533–540, 2006.

Post-translational modifications such as phosphorylation, acetylation, and sulfonation regulate many fundamental biological processes within the cell by mediating protein-protein interactions (1–3). For example, phosphorylation on specific amino acids (tyrosine, serine, or threonine) can directly control the assembly of multiprotein complexes by recruiting phosphospecific binding proteins (1, 4–6). Despite the critical importance of post-translational modifications in regulating cellular physiology, relatively few methods exist that allow proteome-wide identification of proteins that interact with these modifications.

Various techniques have been developed to screen large expressed protein libraries for proteins or protein fragments with specific binding properties, including the yeast two-hybrid system and phage display (7–15). However, these systems have limitations. Because it relies on the internal co-expression of “bait” and “prey” fusion proteins, the two-hybrid system cannot be used to identify proteins that bind to externally synthesized or modified proteins or compounds. More recently, variations of the two-hybrid system have been developed for the analysis of post-translational modification-dependent protein-protein interactions and small molecule-protein interactions (16, 17). However, these methods require the design and construction of assay-specific components and face uncertainties in efficiency and specificity of modifications that take place inside the yeast cell (16). Phage display is limited by potential expression bias against eukaryotic proteins expressed in a prokaryotic host and the low number of fusion proteins displayed on each phage particle (8, 12).

To address these limitations, we constructed a yeast surface-displayed library of human cDNA fragments. Heterologous protein fragments can be efficiently displayed at high copy levels on the Saccharomyces cerevisiae cell wall, and yeast surface display technology has been successfully used to affinity mature human antibody fragments and map antibody-binding epitopes (18–20). Because yeast protein expression pathways are similar to those found in mammalian cells, human protein fragments displayed on the yeast cell wall are likely to be properly folded and functional. Coupled with fluorescence-activated cell sorting (FACS)1 (18, 20), yeast surface-displayed cDNA libraries potentially allow the selection of protein fragments or domains with affinity for any soluble molecule that can be fluorescently detected. In this report we demonstrate the utility of this approach by identifying human protein fragments with affinity for tyrosine-phosphorylated peptides derived from the major autophosphorylation sites of the epidermal growth factor receptor (EGFR) and focal adhesion kinase (FAK).

From the Department of Anesthesia and UCSF Comprehensive Cancer Center, University of California, San Francisco, California 94110

Received, September 21, 2005, and in revised form, November 21, 2005

Published, MCP Papers in Press, December 1, 2005, DOI 10.1074/mcp.M500309-MCP200

© 2006 by The American Society for Biochemistry and Molecular Biology, Inc.

This paper is available on line at http://www.mcponline.org

Additional information is available in the Supporting Material.
**EXPERIMENTAL PROCEDURES**

**Construction of a Yeast Surface-displayed Testis cDNA Library**—A random primed, size-selected (0.3–1.2 kb; average size, 0.75 kb) adult human testes cDNA library with $5 \times 10^6$ primary clones (Invitrogen) was digested with EcoRI, ligated into EcoRI-digested and dephosphorylated pYD1 yeast display vector (Invitrogen), and transformed into 10G Supreme competent bacteria (Lucigen, Middleton, WI). Greater than $6 \times 10^7$ transformants were pooled to form the library, and the inserts of 20 randomly picked clones were sequenced to verify the quality and diversity of the library. Plasmid was prepared from the library using a Qiagen maxiprep kit (Qiagen, Hilden, Germany) and transformed into the Saccharomyces cerevisiae strain EBY100 (Invitrogen), and transformants were selected on SD-CAA medium (2% dextrose, 0.67% yeast nitrogen base, 0.5% casamino acids). Greater than $6 \times 10^7$ transformants were pooled and harvested by resuspending in SD-CAA + 15% glycerol, aliquoted, and stored at $-80 \, ^\circ C$.

**Fluorescence-activated Cell Sorting for Selection of Phosphopeptide-binding Clones**—The yeast library was grown in SR-CAA (SD-CAA with 2% raffinose in place of dextrose) at 30 °C to an $A_{600}$ of 5. To induce expression of testis cDNA products on the yeast surface, the yeast were reactivated at an $A_{600}$ of 0.5 in SRG-CAA (SR-CAA + 2% galactose) and grown at 30 °C for 16–36 h. To monitor induction, expression of the V5 epitope, the cDNA insert must have an ORF along its entire length that is in frame with both the upstream AGA2 coding region and the downstream V5 coding region. The P2 gate shows yeast that express the V5 tag and therefore contain fully in-frame cDNA inserts. Typically about 4–6% of cells from the induced library were V5-positive, suggesting that about 12–18% of induced cells express an Aga2p-fused ORF that runs the entire length of the insert. GPI, glycosylphosphatidylinositol.

---

**Fig. 1. Construction of yeast surface-displayed human cDNA library.** A. Size-selected (0.3–1.2 kb) human testis cDNAs were cloned into the yeast display vector pYD1 to create a yeast surface-displayed library containing $5 \times 10^6$ members. Human protein fragments were expressed as fusions with Aga2p. The cDNA inserts were flanked by epitope tags for monitoring the level of expression (Xpress™ tag) and the amount of in-frame fusion (V5 tag). PCR analysis showed that the cDNA fragments range from 0.4 to 1.2 kb. B. The expression level of in-frame cDNA fusions was monitored using a monoclonal anti-V5 epitope antibody. To express the V5 epitope, the cDNA insert must have an ORF along its entire length that is in frame with both the upstream AGA2 coding region and the downstream V5 coding region. The P2 gate shows yeast that express the V5 tag and therefore contain fully in-frame cDNA inserts. Typically about 4–6% of cells from the induced library were V5-positive, suggesting that about 12–18% of induced cells express an Aga2p-fused ORF that runs the entire length of the insert. GPI, glycosylphosphatidylinositol.
washed twice with PBS, and incubated in 500 μl of PBS with 10 μM biotinylated, tyrosine-phosphorylated peptides EGFRpY1173 (STAENAELYLRVAPQS) or FAKpY397 (SVSETDDYAEIDDEE) (phosphorylated residues in bold) for 4 h at 4 °C. The corresponding non-phosphorylated, non-biotinylated peptides were added at a final concentration of 40 μM to compete away non-phosphospecific binding. Cells were washed twice with PBS and incubated with 500 μl of 1:500 diluted phycoerythrin-conjugated streptavidin (SA-PE) (Invitrogen/Biosource, Camarillo, CA) for 20 min at 4 °C. Cells were washed twice with PBS, sorted by flow cytometry (FACSAria, BD Biosciences), and recovered on SD-CAA plates. Approximately 5 x 10^7 cells were analyzed in the first round selections. In subsequent rounds, allophycoecyanin-conjugated streptavidin (SA-APC) (Invitrogen/Molecular Probes, Eugene, OR) was alternated with SA-PE to minimize the selection of clones that bind the detection reagent. After four rounds of sorting, individual clones were picked, induced, and tested by FACS (LSRII, BD Biosciences) for binding to both phosphorylated and non-phosphorylated peptides.

**Isolation and Sequencing of Plasmids from Yeast Clones**—Plasmids were recovered from yeast clones exhibiting phosphospecific binding.

---

**Fig. 2.** Selection of phosphopeptide-binding clones from a yeast surface-displayed human cDNA library by FACS. Enrichment of EGFRpY1173- and FAKpY397-binding clones through four rounds of FACS. Bound peptide was detected with SA-PE. The sort window is shown for the first round. In the fourth round, the **upper left quadrant** was sorted, and the output was used for individual clone analysis.

**Fig. 3.** Relationship between SH2 domains and cloned cDNA ORFs. In all cases, the ORF (solid black bar) contains the entire predicted SH2 domain (labeled box). The tensin molecule is not drawn to scale.
**FIG. 4.** Phosphorylation-dependent binding of selected clones. Induced yeast clones were tested for binding with either phosphorylated or non-phosphorylated peptides (10 μM). Bound peptide was detected with SA-PE. Typically 40–60% of the yeast population was induced.
binding using a modified QIAprep Spin Miniprep protocol that incorporates a glass bead cell lysis step (Qiagen). Isolated plasmids were transformed into DH5α cells and purified, and the cDNA inserts were sequenced. Public gene and protein databases were searched for matches to each cDNA insert.

**K<sub>D</sub> Measurements**—Induced phosphopeptide-binding clones were incubated with various concentrations of EGFRpY1173 or FAKpY397 for 16 h at 4 °C. After two PBS washes cells were incubated with 1:500 SA-PE for 20 min at 4 °C. After two PBS washes, cells were analyzed by FACS (LSRII, BD Biosciences). Mean fluorescence intensity data were plotted and analyzed with GraphPad Prism software (GraphPad Software, San Diego, CA) using a one site binding, non-linear regression curve fit algorithm.

**RESULTS**

**Construction of a Yeast Surface-displayed Testis cDNA Library**—We constructed an inducible library of 5 × 10<sup>6</sup> human testis cDNA fragments displayed on the yeast surface as C-terminal fusions to the yeast α-agglutinin subunit, Aga2p (Fig. 1A). The cDNA fragments were derived from a library generated by random priming and ranged in size from 0.3 to 1.2 kb. We reasoned that protein fragments of this size (~100–400 amino acids) would be expressed efficiently by the yeast yet remain large enough to encompass stable functional domains. Sequencing analysis of 20 randomly picked clones was performed to verify the quality and diversity of the library. All of the 20 clones analyzed were unique and matched expressed sequence tags in public databases (data not shown). To determine the efficiency of protein display on the yeast surface after induction, expression of the V5 epitope located downstream of the cloning site in the pYD1 vector was monitored. For the V5 epitope to be expressed on the yeast surface, the cDNA insert must have an ORF along its entire length that is in frame with both the upstream AGA2 coding region and the downstream V5 coding region. Typically about 4–6% of cells from the induced library were V5-positive (Fig. 1B). Because only one-third of clones with a full-length in-frame Aga2p-fused ORF will also be in frame with the V5 epitope, the actual number of clones with a full-length Aga2p-fused ORF will be ~3 times the number of V5-positive clones. Thus, we estimated that about 12–18% of induced cells express an Aga2p-fused ORF that runs the entire length of the insert.

**Selection, Identification, and Characterization of Phosphopeptide-binding Clones**—Surface expression of the human cDNA library was induced, and the yeast cells were incubated with 10 μM biotinylated, tyrosine-phosphorylated peptides derived from the major autophosphorylation sites of either EGFR (EGFRpY1173) or FAK (FAKpY397). To compete away non-phosphospecific binders, the corresponding non-phosphorylated, non-biotinylated peptides (40 μM) were included in the incubations. Binding was detected with SA-PE or SA-APC, and binding clones were enriched through four rounds of FACS. Very few binding clones (<0.5%) were present in the initial library population (Fig. 2). After three rounds of selection, >15% (for EGFRpY1173 sort) and >40% (for FAKpY397 sort) of the population bound the peptides (Fig. 2). Individual yeast clones from the fourth round population were screened by FACS for phosphospecific binding to either EGFRpY1173 or FAKpY397. Plasmids from phosphospecific peptide-binding clones were recovered, retransformed into yeast to verify the results of the primary screen, and sequenced to determine the identity of their cDNA inserts.

Four unique phosphopeptide-binding clones were identified. Clones expressing fragments encompassing the SH2 domains of adapter protein APS and phosphoinositide 3-kinase regulatory subunit 3 (PIK3R3) were recovered from the EGFRpY1173 sort (Fig. 3). These clones bind EGFRpY1173 but not the non-phosphorylated control peptide EGFRnP1173 (Fig. 4). The same APS-expressing clone identified by the EGFRpY1173 sort as well as clones expressing the SH2 domains of SH2B and tensin were recovered from the FAKpY397 sort (Fig. 3). These clones bind FAKpY397 but not the non-phosphorylated control peptide FAKnP397 (Fig. 4). The expressed regions and accession numbers for the human proteins matching each clone are shown in Table I.

To determine binding affinity and specificity, we used FACS to measure the equilibrium dissociation constants (K<sub>D</sub>) of the four yeast surface-displayed phosphopeptide-binding clones for both EGFRpY1173 and FAKpY397. Although the APS and PIK3R3 clones have similar affinities for both EGFRpY1173 and FAKpY397, the tensin and SH2B clones have much greater affinity for the FAKpY397 phosphopeptide (Table I, and see Fig. 5 for data plots). Although the tensin and SH2B clones bind to FAKpY397 with K<sub>D</sub> of <3 μM, binding to EGFRpY1173 was not observed until a concentration of 5 μM, and no saturation was evident even at the highest concentration tested (20 μM) (Fig. 5).

**DISCUSSION**

Here we have described the construction of a yeast surface-displayed human cDNA library and demonstrated its utility in the study of post-translational modifications by identifying protein fragments with phosphospecific affinity for tyrosine-phosphorylated peptides derived from the major autophosphorylation sites of EGFR and FAK.
phosphorylation sites of EGFR and FAK. Not surprisingly, human protein fragments encompassing SH2 domains were recovered. Notably most of these represent previously unreported interactions. For example, although the adapter protein APS has been shown to interact with a number of receptor tyrosine kinases including the insulin receptor (21) and the

**Fig. 5.** Plots of fitted $K_D$ data for APS, PIK3R3, SH2B, and tensin clones. Induced yeast were incubated with increasing amounts of either EGFRpY1173 or FAKpY397, and binding was detected with SA-PE. Data were fitted, and $K_D$ values were calculated using GraphPad Prism (GraphPad Software). No measurable binding to the non-phosphorylated peptides was observed. MFI, mean fluorescence intensity.
stem cell factor (c-Kit) (22), its association with autophosphorylation sites of EGFR or FAK has not been reported previously. Similarly although tensin localizes to integrin-mediated focal adhesions (23) and coprecipitates with FAK (24), a direct interaction between the SH2 domain of tensin and FAKpY397 has not been reported previously. Several previously reported interactions were not identified by the sorting. For example, the SH2 domains of protein-tyrosine phosphatase SHP-1 and adapter protein Shc have been reported to interact with phosphotyrosine 1173 of EGFR (25, 26). There are several possible explanations for this. First, because we sorted using a peptide concentration of 10 µM, it is possible that low affinity interactions were not detected. Indeed the binding affinity of an EGFRpY1173 phosphopeptide to the SH2 domain of Shc has not been reported to be 60 µM (27). Because the lowest binding affinity recovered by our sorting was 3.3 µM, it is likely that a 60 µM interaction would not be detectable under our sorting conditions. Second, because only about 200 yeast clones were screened from each of the fourth round sorting populations and most of the clones were recovered only once, it is possible that a more extensive screening of the final round populations would yield additional binding clones. Finally because the library was derived from a tissue-specific cDNA source, it is possible that some genes are not represented in the library. To address this issue we are currently constructing a larger library with cDNA inserts derived from a variety of different tissue sources.

We displayed a large library of human protein fragments on the yeast surface, making them accessible to probing by ligands of diverse chemical compositions. There are potential problems associated with displaying cytoplasmic proteins on the yeast surface. The fusion proteins pass through the yeast secretory pathway and are therefore exposed to an oxidizing environment that may affect proper folding. In addition, some fusion proteins may be glycosylated, which could alter their binding properties. The SH2-domain-containing fragments identified by this study bind to phosphotyrosine peptides specifically, suggesting that they are functional and have folded properly on the yeast surface. This is consistent with the previous reports on functional display of heterologous proteins on the yeast surface (18, 28). Coupled with FACS, yeast surface-displayed cDNA libraries can be used to select protein fragments or domains with affinity for any soluble molecule that can be fluorescently labeled. In addition to the analysis of post-translational modifications, yeast surface-displayed human cDNA libraries have many potential applications, including identifying targets and defining potential cross-reactive proteins for small molecules or drugs and the selection of cellular proteins with novel enzymatic activities.

Acknowledgment—We thank Dr. James D. Marks (University of California at San Francisco) for use of the FACS-Aria.

* This work was supported in part by National Institutes of Health Grant R21 DK086429-01 and Department of Defense Grant W81XWH-05-1-0027. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed: Dept. of Anesthesiology, 1001 Potrero Ave., Rm. 3C38, San Francisco, CA 94110. Tel.: 415-206-6973; Fax: 415-206-6276; E-mail: Liub@anesthesia.ucsf.edu.

REFERENCES


22. Wollberg, P., Lennartsson, J., Gottfridsson, E., Yoshimura, A., and...