Comprehensive Analysis of Yeast Surface Displayed cDNA Library Selection Outputs by Exon Microarray to Identify Novel Protein-Ligand Interactions

Scott Bidlingmaier‡, Yong Wang‡, Yue Liu‡, Niu Zhang‡, and Bin Liu‡§

Phosphatidylinositides are important signaling molecules that interact with a myriad of cellular proteins, many of which remain unidentified. We previously screened a yeast surface displayed human proteome library to identify protein fragments with affinity for the phosphatidylinositides, phosphatidylinositol-4,5-bisphosphate and phosphatidylinositol-3,4,5-trisphosphate. Much of the diversity in the screened selection outputs was represented by clones present at low frequencies, suggesting that a significant number of additional phosphatidylinoside-binding protein fragments might be present in the selection outputs. In the studies described in this report, we developed a novel cDNA library analysis method and comprehensively analyzed the polyclonal selection outputs from the phosphatidylinositol-4,5-bisphosphate and phosphatidylinositol-3,4,5-trisphosphate selections using a high-density exon microarray. In addition to the nine previously reported phosphatidylinoside-binding protein fragments, we identified 37 new phosphatidylinoside-binding candidates. Nine of 37 contain known phosphatidylinoside-binding domains, whereas the remaining 28 contain no known phosphatidylinoside-binding domain. We cloned and confirmed phosphatidylinoside binding by fluorescence-activated cell sorting for 17 of these novel candidate protein fragments. Our experiments suggest that phosphatidylinoside binding by these 17 novel protein fragments is dependent on both the inositol phosphate “headgroup” and the lipid “tail.” This is in contrast with the PH domain containing fragments we tested, for which the inositol phosphate headgroup was sufficient for binding. The novel PtdIns-binding fragments come from a wide variety of proteins, including splicing factors, transcription factors, a kinase, and a polymerase. Intriguingly, 11 of the phosphatidylinoside-binding protein fragments are from nuclear proteins, including four containing homeobox domains. We found that phosphatidylinositides and double-stranded DNA oligonucleotides derived from homeobox domain target sequences compete for binding to homeobox domain-containing protein fragments, suggesting a possible mechanism for phospholipid-dependent transcriptional regulation. FACS enrichment of target-binding clones in yeast human cDNA display libraries coupled with comprehensive analysis of the selection output by DNA microarray analysis is an effective method for investigating common as well as rare protein interactions. In particular, this method is well suited for the study of small molecule/protein and drug/protein interactions. Molecular & Cellular Proteomics 10: 10.1074/mcp.M110.005116, 1–13, 2011.

Phosphatidylinositides are a minor species of phospholipids present in cellular membranes that function as important regulators of many processes, including signal transduction, ion channels and transporters, vesicle trafficking, cytoskeletal organization, and motility (1–5). In addition to cytosolic membrane phosphatidylinositides, an independent pool of nuclear phosphatidylinositides also exists. Although significantly less is known about the functions, mechanism of action, and physical form of nuclear phosphatidylinositides, they are known to play a role in transcription, pre-mRNA splicing, cell cycle regulation, differentiation, and chromatin organization (6–10).

The various biological functions of phosphatidylinositides are mediated through association with proteins containing phosphatidylinoside-binding globular domains, which bind to different species of phosphatidylinositides with varying degrees of affinity and specificity (11), or unstructured regions with clusters of basic and hydrophobic residues (12). The identification and characterization of additional protein domains or regions that bind to phosphatidylinositides is important for understanding the mechanisms by which they regulate cellular processes. In particular, the identification of nuclear phosphatidylinoside-binding proteins will be critical for increasing our understanding of how phosphatidylinositides regulate processes in the nucleus.

Generally speaking, identifying proteins that interact with small bioactive molecules such as phosphatidylinositides is an important, but often challenging and rate-limiting step in...
understanding cellular signaling pathways. Using yeast surface display techniques, heterologous protein fragments can be efficiently displayed on the surface of the Saccharomyces cerevisiae yeast cell as C-terminal fusions to the yeast a-agglutinin subunit, Aga2p (13). Yeast surface display technology has been extensively utilized for protein engineering (14, 15). We have previously described the construction of large \( \times 10^7 \) yeast surface-displayed human protein fragment libraries (16, 17). When coupled with fluorescence-activated cell sorting (FACS)\(^1\), yeast surface-displayed human protein fragment libraries can theoretically be used to identify protein fragments with affinity for any soluble molecule that can be fluorescence detected. These libraries have been used to identify human protein fragments with affinity for tyrosine-phosphorylated peptides (16), tumor-targeting antibodies (18), and phosphatidylinositides (17).

We previously screened our yeast surface human cDNA display libraries for protein fragments with affinity for the phosphatidylinositides (PtdIns), phosphatidylinositol-4,5-bisphosphate (PtdIns(4,5)) and phosphatidylinositol-3,4,5-trisphosphate (PtdIns(3,4,5)), and recovered fragments from nine unique proteins. Eight of these fragments encoded known phosphatidylinositide binding domains (7 pleckstrin homology (PH) domains and one phosphotyrosine-binding domain), demonstrating the effectiveness of the approach (17). Our analysis revealed that whereas a small number of clones dominated the selection outputs, there were nevertheless many binding clones present at low frequencies. This suggested to us that we did not capture the full diversity of the selection outputs with our original screening and sequencing, and that a significant number of additional phosphatidylinositide-binding clones might be present at a low frequency in the selection outputs. This is a general problem for all cDNA library selection-based ligand identification approaches, as monoclonal sequencing is inadequate to cover the full diversity in the selection output. In this report we describe a novel method to address this problem. We analyzed the total polyclonal selection outputs using high density human exon microarrays, capturing the full diversity of the cDNA library selection output. In addition to the nine previously reported phosphatidylinositide-binding protein fragments, we identified 17 novel phosphatidylinositide-binding protein fragments, including 11 derived from nuclear proteins. Interestingly, in contrast with the PH domain containing fragments we tested, for which the inositol phosphate “headgroup” was sufficient for binding, phosphatidylinositide binding by these novel protein fragments is dependent on the nature of the lipid “tail.” Four of the nuclear phosphatidylinositide-binding proteins contain a homeobox domain. We found that phosphatidylinositides and double-stranded DNA oligonucleotides derived from homeobox domain target sequences compete for binding to homeobox domain-containing protein fragments, suggesting a possible mechanism for phospholipid-dependent transcriptional regulation. Our results suggest that selection experiments using yeast human cDNA display libraries coupled with comprehensive analysis of the outputs using DNA microarrays is an effective method for investigating protein interactions.

**EXPERIMENTAL PROCEDURES**

**Microarray Analysis**—Plasmids were recovered from phosphatidylinositide-binding enriched pooled polyclonal yeast selection outputs using a modified QiAprep Spin Miniprep protocol incorporating a glass bead cell lysis step (Qiagen, Hilden, Germany). The isolated plasmid was transformed by electroporation into 10G Supreme bacteria (Lucigen, Middleton, WI) and plasmid DNA was purified from the pooled transformants using a Qiagen Maxi Prep kit (Qiagen). Biotinylated cRNA probes were made by in vitro transcription using Xho1 or Not1-linearized selection output plasmid as templates with the Ambion MAXScript T7 kit according to manufacturer’s protocols (Ambion, Austin, TX). The biotinylated cRNA was fragmented by sonication and 5 μg was hybridized to an Affymetrix GeneChip® Human Exon 1.0 ST Array. The GeneChip Human Exon 1.0 ST Array has over 5 million unique 25-mer oligonucleotide probes grouped into 1.4 million probe sets that represent over 1 million exon clusters (19). After processing according to manufacturer’s instructions, the data was collected using an Affymetrix GeneChip Scanner 3000 (Affymetrix, Santa Clara, CA). The raw exon microarray dataset was processed using Affymetrix Expression Console software (Affymetrix), and analyzed with Excel (Microsoft, Redmond, WA) and Integrated Genome Browser software (Affymetrix) to generate a list of phosphatidylinositide-binding candidates.

**Recovery of Candidate PtdIns-binding Clones**—Internal primers for phosphatidylinositide-binding candidates were designed based on the exon microarray data and used along with vector-specific primers to PCR amplify fragments containing the 5’ and 3’ ends of the cDNA inserts, which were cloned into TA vector (Invitrogen, Carlsbad, CA) and sequenced to define the precise 5’ and 3’ cDNA insert boundaries and determine if the cDNA encoded protein is in-frame with AGA2 in the pYD1 yeast display vector. To generate plasmids for yeast display, new primer sets based on the 5’ and 3’ boundaries of the cDNA inserts were designed with EcoRI overhangs and used to PCR amplify the cDNA inserts, which were cloned into pYD1 (Invitrogen) and verified by sequencing. The candidate cDNA insert containing pYD1 plasmids were transformed into the S. cerevisiae yeast strain EBY100 (Invitrogen) using the lithium acetate method and transformants were selected on S.D.-CAA medium (2% dextrose, 0.67% yeast nitrogen base, 0.5% casamino acids).

**FACS Analysis**—Candidate phosphatidylinositide-binding yeast clones were grown overnight in SGR-CAA (2% galactose, 2% raffinose, 0.67% yeast nitrogen base, 0.5% casamino acids) to induce surface display of the candidate phosphatidylinositide-binding protein fragments. Induced yeast clones expressing the putative phosphatidylinositide-binding protein fragments or negative control yeast displaying Xpress epitope-tagged Aga2p from the empty pYD1 were washed twice with phosphate buffered saline (PBS) and incubated for 2 h at room temperature with 500 nM biotinylated PtdIns(4,5) and PtdIns(3,4,5) (hybrid phosphatidylinositol polyphosphates, Echelon Biosciences, Salt Lake City, UT) and mouse anti-Xpress antibody (1:1000 dilution of stock solution, Invitrogen) to monitor surface dis-

---

\(^1\) The abbreviations used are: FACS, fluorescent antibody cell sorting; PtdIns, phosphatidylinositide; PtdIns(4,5), phosphatidylinositol-4,5-bisphosphate; PtdIns(3,4,5), phosphatidylinositol-3,4,5-trisphosphate; PH, pleckstrin homology; PBS, phosphate buffered saline; GST, glutathione S-transferase; FITC, fluorescein isothiocyanate.
play of the protein fragments via the Xpress epitope tag, which is immediately upstream of the inserted cDNAs in the pYD1 vector. The yeast were washed twice with PBS and incubated with 1/500 streptavidin-phycoerythrin (Invitrogen) and 1/500 Alexa fluor 647-conjugated anti-mouse secondary antibody (Jackson ImmunoResearch, West Grove, PA) in PBS for 30 min. Following two washes with PBS, the yeast were analyzed by FACS (LSRII, BD Biosciences, San Jose, CA).

Induced phosphatidylinositide-binding yeast clones were also tested for binding to 1 μM Biotin-InsP$_3$ and Biotin-PtdIns, C$_6$ (Echelon) using the same protocol. For competition experiments, the induced yeast displaying the phosphatidylinositide binding protein fragments were washed with PBS and co-incubated with 500 nM GST-Grp1 PH domain (PIP$_3$ Grip, Echelon) and 250 nM biotinylated PtdIns(3,4,5) for 2 h and then analyzed by FACS as described above. For the binding and competition experiments with the homeobox target sequence, oligos containing the same nucleotide composition (5’-gcaggc-3’ and 5’-gctgtgactatgc-3’) were synthesized and annealed. For the titration experiments, varying concentrations of biotinylated PtdIns(3) or PtdIns(3,4,5) and randomized oligos containing the same nucleotide composition (5’-gcatagtactcagac-3’ and 5’-gtctgagctactatgc-3’) were synthesized and annealed. For the titration experiments, varying concentrations of biotinylated PtdIns(3,4,5) were incubated with the homeobox domain displaying yeast for 2 h, washed with PBS, and then incubated with 1/500 streptavidin-phycoerythrin in PBS for 30 min, washed twice with PBS, and analyzed by FACS. The FITC-labeled double-stranded annealed homeobox target sequence oligos were incubated for 2 h with the HOXD4 homeobox domain displaying yeast and then washed twice with PBS and analyzed by FACS. For competition experiments, HOXD4 homeobox domain displaying yeast were co-incubated with varying concentrations of biotinylated PtdIns(3) or PtdIns(3,4,5) and 200 nM FITC-labeled double-stranded homeobox target sequence oligo for 2 h at room temperature, washed twice with PBS, and then analyzed by FACS.

RESULTS

Accessing the full diversity of yeast surface-displayed human protein fragment library selection outputs by exon microarray analysis—We have previously described the use of yeast surface-displayed human protein fragment libraries for FACS-based selection of protein fragments with affinity for the phosphatidylinositides PtdIns(4,5) and PtdIns(3,4,5) (17). Analysis of the distribution of phosphatidylinositide-binding clones in the sort outputs revealed that 77% (72/93) of the sequenced binding clones contained cDNA inserts derived from three genes (PDK1, DAB2, and SBF1). Several binding clones were present at low frequencies, suggesting that our screening did not reveal the full diversity of the sorting outputs, and that additional phosphatidylinositide-binding clones were likely to be present in the sort outputs. We therefore sought to use DNA microarray analysis to comprehensively analyze the polyclonal selection outputs and identify additional phosphatidylinositide-binding protein fragments. A diagram of the selection process and output analysis is presented in Fig. 1. Plasmid DNA was purified from polyclonal yeast from the PtdIns(4,5) and PtdIns(3,4,5) round three selection outputs and used as a template for in vitro transcription to generate biotin-labeled RNA. The labeled RNA was hybridized to the exon microarray and data was collected and analyzed.
first examined the microarray data for the presence of the nine proteins with PtdIns-binding fragments that were previously identified in the selection outputs by monoclonal screening and sequencing (17). Strong signal intensities were observed in the regions corresponding to all nine of the previously identified proteins, suggesting that the microarray data accurately reflects the composition of the cDNA inserts in the polyclonal selection outputs (Fig. 2). Further analysis of the microarray data revealed 37 new candidate phosphatidylinositide-binding protein fragments that were not identified by previous monoclonal screening and sequencing (see Fig. 3 for examples of microarray data for new candidates). Nine of these candidate protein fragments contained known phosphatidylinositide-binding domains (7 PH domains and 2 C2 domains), whereas 28 represented novel candidate phosphatidylinositide-binding protein fragments. Fourteen of the novel candidate protein fragments are from proteins whose primary localization is known or predicted to be nuclear. Thus our exon microarray-based comprehensive analysis of cDNA library selection output is highly effective in assessing the full diversity of the output, overcoming a major limitation of traditional monoclonal sequencing-based output analysis.

Confirmation of Candidate Protein Fragment Binding to PtdIns(4,5) and PtdIns(3,4,5)—We first confirmed binding of candidate protein fragments to phosphatidylinositides. We chose 17 of the novel candidate cDNA sequences for testing, giving priority to nuclear proteins or proteins with known enzymatic activity. The candidate cDNA fragments were amplified by PCR using plasmid DNA purified from the polyclonal selection outputs as template, sequenced to verify in frame fusion with Aga2p, and recloned into the yeast display vector. We then transformed the plasmids into yeast and tested for binding to biotinylated PtdIns(4,5) and PtdIns(3,4,5) (Fig. 4A) by FACS. All 17 of the yeast displayed protein fragments bound to both biotinylated PtdIns(4,5) and PtdIns(3,4,5) (Fig. 4B). The 17 confirmed novel phosphatidylinositide-binding protein fragments in addition to the nine previously reported protein fragments (17) are summarized in Table I (protein sequences for the 17 novel fragments are listed in Supplemental Table 1). Interestingly, 11 of the newly identified phosphatidylinositide-binding fragments are from nuclear proteins. Details of the 20 unconfirmed candidate phospholipid-binding protein fragments are presented in Table II.

We next tested the ability of some of the protein fragments to bind PtdIns(3,4,5) using an in vitro pulldown assay. We used glutathione S-transferase (GST) fusion proteins containing the phosphatidylinositide-binding fragments of HMGN2, HOXA5, SFRS4, and WNK1 for pulldown experiments with

\[ \text{Fig. 2. Microarray probeset intensity data corresponding to phosphatidylinositide-binding protein fragments previously identified in the selection outputs.} \]
PtdIns(3,4,5)-coated beads. Although GST alone was not pulled down by the PtdIns(3,4,5)-coated beads, GST-HMGN2, GST-HOXA5, GST-SFRS4, and GST-WNK1 were pulled down by the PtdIns(3,4,5)-coated beads but not the uncoated control beads, indicating that these protein fragments are capable of binding PtdIns(3,4,5) in vitro (Fig. 5). A diagram of the position of the phosphatidylinositide-binding regions relative to other known protein domains is shown in Fig. 6.

Analysis of Inositol Phosphate “Headgroup” and Lipid “Tail” Contributions to Binding of Phosphatidylinositide-binding Protein Fragments—To explore the relative contributions of the inositol phosphate “headgroup” and the lipid “tail” to binding, we tested 16 of the novel PtdIns-binding protein fragments and 2 PH domain-containing fragments for binding to biotinylated Ins(3,4,5) by FACS using the yeast display system. The headgroups of the biotinylated Ins(3,4,5) and the previously tested biotinylated PtdIns(3,4,5) are identical, but the structures of their hydrophobic lipid tail regions are different (compare Fig. 4A to Fig. 7A). Although the two PH domain-containing protein fragments bound to the Ins(3,4,5), no binding of the novel, non-PH domain-containing protein fragments bound was observed at the concentration tested (1 μM) (Fig. 7B). We also tested some of the clones for binding to a different form of biotinylated phosphatidylinositide, PtdIns(3,4,5)C₆, which has a shortened six carbon chain lipid tail, and obtained similar results (no binding) (Supplemental Fig. 1). These results suggest that the inositol phosphate headgroup is not sufficient for binding to these protein fragments and that binding is dependent in part on the nature of the lipid tail.

To further explore this issue, we performed competition experiments using a purified protein fragment containing the Grp1 PH domain. PH domains are known to bind to the inositol phosphate headgroups of phosphatidylinositides (21). Using the same FACS assay as in previous experiments, we found that addition of the Grp1 PH domain significantly reduced the binding of PtdIns(3,4,5) to all protein fragments we tested (Fig. 8). Thus, the PtdIns(3,4,5) cannot efficiently bind the Grp1 PH domain and any of the protein fragments we tested simultaneously, suggesting that the inositol phosphate headgroup or nearby regions are necessary for binding to the tested protein fragments. Taken together, these data suggest that the novel phosphatidylinositide-binding protein frag-

---

**Fig. 3.** Microarray probe set intensity data for selected new phosphatidylinositide-binding candidates not identified by previous monoclonal screening and sequencing. The gene exon structure is shown in green. Intensity values for exon-associated probe sets are represented by red bars below the gene exon structure. Data is visualized using Integrated Genome Browser software.
ments we tested bind to phosphatidylinositides differently than PH domains, in a manner that is dependent on the lipid tail.

PtdIns(3,4,5) and Homeobox Domain Target Sequence DNA Compete for Binding to Homeobox Domain-containing Protein Fragments—Four of the phosphatidylinositide-binding protein fragments contain a homeobox domain, a helix-turn-helix structure that binds DNA in a sequence specific manner and is commonly found in transcription factors (22). We did competition studies to determine if PtdIns(3,4,5) could compete with homeobox domain target sequence DNA for binding to the yeast displayed protein fragments is shown. Yeast displaying Xpress epitope-tagged Aga2p from the empty pYD1 vector serve as a negative control. The y axis indicates phosphatidylinositide binding whereas the x axis indicates surface expression monitored via the Xpress epitope tag, which is immediately upstream of the inserted cDNAs in the pYD1 vector.

Quantitative analysis of protein-protein (23) and protein-DNA interactions (24) has been previously reported. The apparent affinity of PtdIns(3,4,5) for the HOXD4 homeobox domain-containing protein fragment was ~50 nM whereas the apparent affinity of the homeobox domain target sequence oligo was ~100 nM. We tested the ability of increasing amounts of PtdIns(3,4,5) to reduce binding of the homeobox target sequence oligo to yeast surface-displayed HOXD4 homeobox domain. As a control, we used PtdIns(3), which does not bind to the yeast surface-displayed HOXD4 homeobox domain. We found that addition of PtdIns(3,4,5), but not PtdIns(3), could reduce binding of the homeobox target sequence oligo to yeast surface-displayed HOXD4 homeobox domain in a concentration-dependent manner (Fig. 9B). We next tested the ability of the homeobox target sequence oligo to compete with PtdIns(3,4,5) for binding to recombinant HOXA5 homeobox domain using an in vitro pulldown assay.

10.1074/mcp.M110.005116–6
The homeobox target sequence oligo, but not a randomized negative control oligo, was able to prevent the pulldown of the HOXA5 homeobox domain protein fragment by PtdIns(3,4,5)-coated beads (Fig. 9C). Thus, our experiments suggest that PtdIns(3,4,5) and homeobox domain target sequence DNA compete for binding to homeobox domain-containing protein fragments.

**DISCUSSION**

Bioactive molecules often bind to multiple cellular proteins. We have previously developed an effective approach based on screening yeast surface displayed human cDNA libraries to identify binding proteins for any ligands that can be fluorescently labeled (16–18). After enrichment of binding clones by several rounds of FACS, individual clones were tested for binding to the target ligand and then sequenced. Although effective, the distribution of binding clones in the sorting outputs suggested that the full diversity had not been uncovered by the monoclonal screening approach. Conventional monoclonal screening and sequencing approaches often identify dominant clones, leaving important biological interactions undetected because of low representation in the selection output. This is a general problem for all expression cloning strategies based on cDNA library selection. We have developed a novel method to address this problem. By analyzing the selection output using a high-density human exon array, the full diversity of the selection output has been uncovered at once. Our study shows that the combination of FACS selection using yeast human proteome surface display libraries with exon microarray analysis of the selection outputs is a powerful way of identifying protein/ligand interactions. The method is comprehensive, and not biased toward abundant proteins. The method can be used to identify protein fragments with affinity for any soluble ligand that can be fluorescently detected, including small biological molecules and drugs. The method is generally applicable to analyzing cDNA library selection outputs when many binders are expected. For example, the method can be applied to the analysis of selection outputs from yeast and phage cDNA display libraries, two-hybrid and three-hybrid libraries, mammalian cDNA expression libraries, and other viral or bacteria-based cDNA display libraries.

Various techniques have been previously used to systematically screen for proteins that bind phosphatidylinositides, including protein microarrays (25), column-based affinity purification using agarose-conjugated phosphatidylinositol

| Protein | Accession # | Subcellular location | Total aa | Expressed aa | PtdIns-binding domain (aa) |
|---------|-------------|----------------------|---------|-------------|---------------------------|
| APOH    | AAH20703    | S                    | 345     | 94–345      | PH (261–386)              |
| ARNO    | CA468084    | PM                   | 399     | 226–399     |                           |
| ATN1    | EAW88707    | N, C                 | 1190    | 1–229       |                           |
| CRABP1  | CAG33298    | C                    | 137     | 1–137       |                           |
| CRIP1A  | AAH11535    | U                    | 164     | 1–140       |                           |
| DAB2    | AAH03064    | CP                   | 771     | 2–221       | PTB (36–176)              |
| FAM71B  | NP_570969   | N, NS                | 605     | 267–546     |                           |
| GAB2    | NP_536739   | PM, C                | 676     | 1–125       | PH (9–116)                |
| HMGN2   | CAG34619    | N, C                 | 90      | 6–90        |                           |
| HOXA5   | CAG47073    | N                    | 270     | 206–270     |                           |
| HOXB6   | AAH14651    | N                    | 224     | 136–224     |                           |
| HOXC6   | CAG33235    | N                    | 235     | 150–235     |                           |
| HOXD4   | EAX11086    | N                    | 255     | 144–255     |                           |
| NUCKS1  | Q9H1E3      | N                    | 243     | 54–243      |                           |
| OSBP2   | AA18915     | PM                   | 915     | 5–290       | PH (185–272)              |
| PKD1    | AAH06339    | C, PM                | 429     | 232–429     | PH (330–394)              |
| PNPLA7  | CA41582     | ER, IM, L, N         | 1317    | 1013–1274   |                           |
| POLS    | Q5XG7       | N                    | 542     | 412–542     |                           |
| PSD     | CA12516     | I                    | 1024    | 739–1024    | PH (758–874)              |
| PTPN5   | AAH64807    | IM                   | 565     | 1–60        |                           |
| RNPS1   | AAC39791    | N, NS, C             | 305     | 1–230       |                           |
| SBF1    | AAH87612    | IM, N                | 1893    | 1638–1893   | PH (1790–1890)            |
| SFRS4   | CA14326     | NS                   | 494     | 409–494     |                           |
| SPTBN2  | NP_008877   | C                    | 2390    | 2101–2390   | PH (2221–2304)            |
| WDR60   | Q8VVS4      | U                    | 1066    | 60–260      |                           |
| WNK1    | Q9HA43      | C                    | 2382    | 924–1156    |                           |

<sup>a</sup> Previously reported clones (17).
Exon Array Analysis of cDNA library Selection Outputs

Table II

Unverified candidate phospholipid-binding protein fragments from microarray analysis of polyclonal selection outputs. NCBI Entrez protein database accession numbers for human protein matches and the expressed regions and total aa lengths are shown. Candidate clones with a cDNA encoded protein that has been determined by sequencing to be in-frame with AGA2 in the pYD1 yeast display vector are indicated in the "Verified in-frame" column. The predicted or observed subcellular localization (retrieved from the GO database (58)) of the protein is also indicated (PM, peripheral membrane; IM, integral membrane; C, cytoplasm; U, unknown; I, intracellular; N, nucleus; NS, nuclear speckle; CP, clathrin coated pit; IF, intermediate filament; G, golgi; MT, microtubule). aa, amino acids; C2, C2 domain; PH, pleckstrin homology domain

| Protein | Accession # | Subcellular location | Total aa | Expressed aa | PtdIns-binding domain (aa) | Verified in-frame |
|---------|-------------|----------------------|----------|-------------|---------------------------|------------------|
| AKT3    | CAB553537   | PM, C                | 479      | 1–128a      | PH (6–106)                | +                |
| FAM62B  | AAH13957    | IM                   | 489      | 310–489     | C2 (355–479)              | +                |
| HCF1C1  | NP_005325   | N, C                 | 2035     | 1583–1743   |                           | +                |
| MAP4    | AAH48200    | MT                   | 857      | 432–684     |                           | +                |
| PAK4    | NP_0085875  | G                    | 591      | 206–591     |                           | +                |
| PHLD1   | AAH13031    | U                    | 691      | 593–691a    | PH (584–683)              | +                |
| PLEK2   | Q8NYT0      | PM, C                | 353      | 237–353     | PH (250–352)              | +                |
| RASA3   | CAI12864    | M                    | 834      | 545–684     | PH (577–675)              | +                |
| SRRP35  | EAW45661    | N, NS                | 262      | 1–95        |                           | +                |
| SYT1    | CAI14783    | PM                   | 550      | 189–373     | C2 (255–324)              | +                |
| APP     | NP_000475   | IM, CP               | 770      | 291–407a    |                           | +                |
| ARHGODA | NP_004300   | C                    | 204      | 1–204a      |                           | +                |
| C14orf104| NP_060609 | C                    | 837      | 515–602a    |                           | +                |
| CYTH1   | NP_059430   | PM, C                | 397      | 234–397a    | PH (260–377)              | +                |
| CYTH3   | NP_004218   | PM                   | 399      | 189–399     | PH (264–381)              | +                |
| CYTH4   | NP_037517   | I                    | 394      | 233–394     | PH (259–375)              | +                |
| PARK7   | NP_009193   | C, N                 | 189      | 1–107a      |                           | +                |
| SART1   | NP_000517   | N                    | 800      | 1–220a      |                           | +                |
| SLC7A5  | NP_003477   | C, M                 | 507      | 181–347a    |                           | +                |
| SLC9A3R2| NP_004776   | M, C, N              | 326      | 139–326a    |                           | +                |

*Insert boundaries predicted from microarray data.

(26), a yeast growth rescue assay (27), λgt11-based expression cloning (28), and an in vitro transcription/translation-based expression cloning assay (29); however, with the notable exception of the protein microarray study that analyzed the yeast phosphatidylinositol-binding proteins, only a small number of phosphatidylinositol-binding proteins have been identified in each study. Including our previously published results (17) we have now used yeast surface cDNA display library screening to identify 26 confirmed phosphatidylinositol-binding protein fragments and another 20 candidate phosphatidylinositol-binding protein fragments. Although many of these protein fragments contain known or predicted phosphatidylinositol-binding domains (14 PH, 2 C2, and 1 phosophotyrosine-binding), 29 contain no previously reported phosphatidylinositol-binding domains or regions. We confirmed phosphatidylinositol binding to 17 of these protein fragments displayed on the surface of yeast cells by FACS. Thus, affinity-based selection with yeast cDNA display libraries coupled with DNA microarray analysis of the selection outputs is an effective method for investigating protein interactions.

Sequence analysis of the 17 confirmed novel phosphatidylinositol-binding protein fragments did not reveal any similarity with known PtdIns-binding domains (i.e. PH, C2) or any apparent conserved motives. Ten of the 17 protein fragments (FAM71B, HMGN2, HOX5, HOX6, HOXC6, HOXD4, POLS, RNP1, SFRS4, and WDR60) are predicted to have a significant net positive charge at neutral pH (pI > 9.0) whereas five (ATN1, CRABP1, NUCKS1, PNPLA7, and PTPN5) are predicted to have a net negative charge (pI < 6). Thus, although charge may contribute to binding with the negatively charged phosphatidylinositides in some cases, there is no absolute requirement for a strong net positive charge.

Significantly, 11 of the 17 confirmed novel phosphatidylinositol-binding protein fragments are from proteins whose primary localization and site of action is the nucleus. Phos-
Phosphatidylinositides and enzymes involved in their synthesis exist in the nucleus, and nuclear phosphatidylinositides have been functionally connected to several nuclear events, including chromatin organization, and mRNA processing and export (6–9, 30, 31). The few nuclear phosphatidylinositide-binding effector proteins that have been described include histones H1 and H3 (32), the PHD finger of ING2 (33), the nuclear receptors SF-1 and LRH-1 (34–36), Nucleophosmin/B23 (26), the PDZ domains of syntenin-2 and zonula occludens-1 and -2 regulates their localization to nuclear speckles (37, 38). It is therefore possible that the protein fragments we identified in FAM71B, RNPS1, SFRS4, and SRRP35 regulate their localization to nuclear speckles in a similar manner.

Four of the phosphatidylinositide-binding protein fragments contain a homeobox domain, which is a helix-turn-helix sequence-dependent DNA binding domain found in transcription factors (22). In our experiments, the apparent affinity of yeast displayed HOXD4 for PtdIns(3,4,5) was ~50 nM. The amount of PtdIns(4,5) in detergent-washed rat liver nuclei was measured as ~30 pmol/mg of protein (40). Using an estimate of 72 mg/ml (calculated using a value of 3.1 pg for the mass of the haploid nuclear DNA content of the rat genome (49) and a 2:1 nuclear protein:DNA mass ratio (50, 51) for the total nuclear protein concentration gives a rough estimate of 4 μM for the nuclear PtdIns(4,5) concentration, indicating that interactions in the affinity range we observed could be biologically significant. Our experiments demonstrated that phosphatidylinositides and double-stranded DNA oligonucleotides derived from homeobox domain target sequences compete for binding to homeobox domain-containing protein fragments. It is possible that nuclear phosphatidylinositides play a role in transcriptional regulation by modulating the interaction between homeobox domain proteins and target DNA sequences (Fig. 10A). A similar model has recently been proposed for the Sin3A corepressor complex proteins SAP30 and SAP30L (39). Future studies could attempt to identify mutations within the homeobox domain that selectively disrupt PtdIns binding without affecting DNA binding. If homeobox protein mutants with these properties could be found, they could be used to model homeobox domain-PtdIns interaction and investigate the possible regulation of homeobox proteins by PtdIns in cell homeostasis and tissue development.

One of the PtdIns-binding regions is from the DNA polymerase POLS. The yeast homolog of POLS, Trf4p, plays a role in sister chromatid cohesion (52) and DNA double-strand break repair (53). Intriguingly, the phosphoinositide 3-kinase catalytic subunit p110β, its substrate, PtdIns(4,5), and its product, PtdIns(3,4,5), have recently been shown to localize to DNA double-strand breaks (54). Thus, it is possible that accumulation of PtdIns at sites of DNA damage acts to recruit POLS and potentially other factors that participate in DNA repair (Fig. 10B). Additionally, it is possible that binding by PtdIns directly regulates the enzymatic activity of POLS.

Although the seventeen novel phosphatidylinositide-binding protein fragments bound to the hybrid form phosphatidylinositide...
linositide, they did not bind to Ins(3,4,5) (headgroup only), indicating that the structure of the lipid tail is also critical for binding. The hybrid form phosphatidylinositides we used in our experiments are soluble in aqueous solutions and have a diacyl lipid tail with two 16-carbon chains, and a covalently linked biotin reporter group (55). The other form of phosphatidylinositide we used, PtdIns(3,4,5) C₆, has a short six-carbon monoacyl lipid tail and did not bind to the PtdIns-binding clones that we tested. An attractive hypothesis is that the pool of nuclear phosphatidylinositides that exists in a nonmembranous, detergent-resistant form is associated with nuclear proteins, likely through interactions with their hydrophobic lipid tails. Thus, the lipid tail-dependent PtdIns-binding nuclear proteins we identified in this study could participate in sequestering phosphatidylinositides in the nucleus.

We have not attempted to define the lipid specificity of the phosphatidylinositide binding protein fragments. It is possible that they are able to bind to other lipid species that we have not studied. In addition, we studied only phosphatidylinositide binding by protein fragments but not the full-length protein. It is formally possible that other regions of the full-length protein could regulate phosphatidylinositide binding by the protein fragments that we identified. Finally, it is important to note that we have not verified the interaction between the proteins we identified and phosphatidylinositides in living cells.

---

**Fig. 7.** FACS analysis of inositol phosphate “headgroup” and lipid “tail” contributions to binding of PtdIns(3,4,5) to phosphatidylinositide-binding protein fragments. A, Chemical structure of biotinylated Ins(3,4,5). B, Testing 16 novel phosphatidylinositide-binding and 2 PH domain-containing yeast displayed protein fragments for binding to biotinylated Ins(3,4,5) (1 μM) by FACS.

**Fig. 8.** Competition between soluble Grp1 PH domain and the yeast displayed phosphatidylinositide-binding protein fragments for binding to PtdIns(3,4,5) measured by FACS. Gray bars represent PtdIns(3,4,5) binding. Black bars represent PtdIns(3,4,5) binding in the presence of soluble Grp1 PH domain. Error bars represent the standard deviation. Data was analyzed using student’s t-tests and p values were <0.05.
cells. In the future, experiments could be done to determine if the proteins we identified interact with endogenous phosphatidylinositides in vivo.

Although there are many more PH domain (258) and C2 domains (149) in the human proteome than what we recovered, it is uncertain whether they all bind strongly to phosphatidylinositides. Indeed, currently available binding data, including genome-wide analysis of PH domains in the yeast Saccharomyces cerevisiae, suggest the likelihood that most PH domains do not bind strongly to phosphatidylinositides (11). Thus, our recovery of 14 PH domains may represent a significant fraction of the total number of PH domains in the human genome that bind strongly to phosphatidylinositides.

C2 domains demonstrate highly variable affinity and specificity for phospholipids and binding is frequently calcium-dependent (56). Because we did not add calcium during our selections, it is possible that the number of C2 domains we recovered was diminished. Several proteins or domains known to bind to PtdIns(4,5) and/or PtdIns(3,4,5), such as the tubby domain (57), the PDZ domain (37, 38), nucleophosmin/B23 (26) were not recovered in our analysis. It is possible that clones containing the full domain or region necessary for binding were not present in the yeast display library, or they may not be displayed in the properly folded conformation for binding.

In summary, we developed a novel exon microarray-based method suitable for the comprehensive analysis of cDNA library polyclonal selection outputs. We identified 37 new phosphatidylinositide-binding candidates from previously generated selection outputs enriched for PtdIns-binding protein fragments (17) and confirmed phosphatidylinositide binding by FACS for 17 of the novel candidate protein fragments. Eleven of the phosphatidylinositide-binding protein fragments are from nuclear proteins, including four containing homeobox domains. We found that phosphatidylinositides can compete with target DNA binding to homeobox domains and thus could play a role in transcriptional regulation. Our studies show that FACS enrichment of target-binding clones in yeast

---

**Fig. 9.** Testing PtdIns(3,4,5) and homeobox domain target sequence DNA competition for binding to homeobox domain containing protein fragments. A, Concentration-dependent binding of biotinylated PtdIns(3,4,5) and homeobox target sequence double-stranded oligonucleotide to yeast displayed HOXD4 homeobox domain-containing protein fragment; B, Competition between PtdIns(3,4,5) or PtdIns(3) and the homeobox target sequence oligonucleotide for binding to the yeast surface-displayed HOXD4 homeobox domain; C, Measuring homeobox target sequence oligonucleotide competition with PtdIns(3,4,5) for binding to recombinant HOXA5 homeodomain by an in vitro pulldown assay with PtdIns(3,4,5)-coated beads. A randomized double-stranded oligonucleotide with the same nucleotide composition was used as a negative control.

**Fig. 10.** Models for regulation of nuclear functions by phosphatidylinositides. A, Binding of PtdIns to the homeodomain prevents binding to homeobox target sequences in chromosomal DNA or displaces bound homeodomain proteins, leading to altered gene expression. B, PtdIns accumulated at sites of double strand DNA breaks acts to recruit POLS, which participates in repairing the break.

---

Exon Array Analysis of cDNA library Selection Outputs

Molecular & Cellular Proteomics 10.3

10.1074/mcp.M110.005116–11
human cDNA display libraries coupled with comprehensive analysis of the selection output by exon microarray analysis is an effective method for investigating protein interactions.

Acknowledgments—We thank Dr. Chris Barker and Ms. Linda Ta at the Gladstone Genomics Core Facility for processing of exon microarrays and helpful discussions.

REFERENCES

1. Balla, T., Szentpetery, Z., and Kim, Y. J. (2009) Phosphoinositide signaling: new tools and insights. Physiology 24, 231–244
2. Di Paolo, G., and De Camilli, P. (2006) Phosphoinositides in cell regulation and membrane dynamics. Nature 443, 651–657
3. Lindmo, K., and Stenmark, H. (2006) Regulation of membrane traffic by phosphoinositide 3-kinases. J. Cell Sci. 119, 605–614
4. Logan, M. R., and Mandato, C. A. (2006) Regulation of the actin cytoskeleton by PI(4,5)P2 in yeast. Biochim. Biophys. Acta 1761, 505–508
5. Ye, K., and Ahn, J. Y. (2008) Nuclear phosphoinositide signaling. Front. Biosci. 13, 540–548
6. Ivan, R. F. (2006) Nuclear inositol signalling – expansion, structures and clarification. Biochim. Biophys. Acta 1761, 505–508
7. Jones, D. R., and Divate, N. (2004) Linking lipids to chromatin. Curr. Opin. Genet. Dev. 14, 196–202
8. Lemmon, M. A. (2008) Membrane recognition by phospholipid-binding domains. Nat. Rev. Mol. Cell Biol. 9, 99–111
9. McCauloug, S., and Murray, D. (2005) Plasma membrane phosphoinositide organization by protein electrostatics. Nature 438, 605–611
10. Bader, E. T., and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. 15, 553–557
11. Pepper, L. R., Cho, Y. K., Boder, E. T., and Shusta, E. V. (2006) A decade of yeast surface display technology: where are we now? Comb. Chem. High Throughput Screen 11, 127–134
12. Gai, S. A., and Wittrup, K. D. (2007) Yeast surface display for protein engineering and characterization. Curr. Opin. Struct. Biol. 17, 467–473
13. Bidlingmaier, S., and Liu, B. (2006) Construction and application of a yeast surface-displayed human cDNA library to identify post-translational modification-dependent protein-protein interactions. Mol. Cell Proteomics 5, 533–540
14. Bidlingmaier, S., and Liu, B. (2007) Interrogating Yeast Surface-Displayed Human Proteome to Identify Small Molecule-binding Proteins. Mol. Cell Proteomics 6, 2012–2020
15. Bidlingmaier, S., He, J., Wang, Y., An, F., Feng, J., Barbone, D., Gao, D., Franc, B., Broadus, V. C., and Liu, B. (2009) Identification of MCAM/CD146 as the target antigen of a human monoclonal antibody that recognizes both epithelioid and sarcomatoid types of mesothelioma. Cancer Res. 69, 1570–1577
16. Abdueva, D., Wing, M. R., Schaub, B., and Triche, T. J. (2007) Experimental comparison and evaluation of the Affymetrix exon and U133Plus2 GeneChip arrays. PLoS One 2, e913
17. Pellerin, I., Schnabel, C., Catron, K. M., and Abate, C. (1994) Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster. Mol. Cell. Biol. 14, 4532–4545
18. Thomas, C. C., Deak, M., Alessi, D. R., and van Aalten, D. M. (2002) High-resolution structure of the pleckstrin homology domain of protein kinase b/akt bound to phosphatidylinositol (3,4,5)-trisphosphate. Curr. Biol. 12, 1256–1262
19. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994) Homeodomain-DNA recognition. Cell 78, 211–223
20. VanAntwerp, J. J., and Wittrup, K. D. (2000) Fine affinity discrimination by yeast surface display and flow cytometry. Biotechnol. Prog. 16, 31–37
21. Jarjour, J., West-Foye, H., Certo, M. T., Hubert, C. G., Doyle, L., Getz, M. M., Stoddard, B. L., and Scharenberg, A. M. (2009) High-resolution profiling of homing endonuclease binding and catalytic specificity using yeast surface display. Nucleic Acids Res. 37, 6871–6880
22. Zhu, H., Bilgin, M., Bangham, R., Hall, D., Casamayor, A., Bertone, P., Lan, N., Jansen, R., Bidlingmaier, S., Houfek, T., Mitchell, T., Miller, P., Dean, R., Gerstein, M., and Snyder, M. (2007) Global analysis of protein activities using proteome chips. Science 318, 2101–2105
23. Ahn, J. Y., Liu, X., Cheng, D., Peng, J., Chan, P. K., Wade, P. A., and Ye, K. (2005) Nuclear phosphoinosit/B3a, a nuclear PI(4,5)(P)3 receptor, mediates the antiaapoptotic actions of NF5 by inhibiting CAD. Mol. Cell 18, 435–445
24. Isakoff, S. J., Cardozo, A., Andreieu, J., Li, Z., Ferguson, K. M., Abagyan, R., Ledimbo, M. R., and Skolnik, E. Y. (2002) Solution and analysis of PH domain-containing targets of phosphatidylinositol 3-kinase using a novel in vivo assay in yeast. EMBO J. 17, 5347–5348
25. Klarlund, J. K., Guilherme, A., Holik, J. J., Virbasius, J. V., and Lichtenberg, M. P. (1997) Signaling by phosphoinositide-3,4,5-trisphosphate through proteins containing pleckstrin and Sec7 homology domains. Science 275, 1927–1930
26. Yu, H., Wu, K., Orikasa, C., and Takenawa, T. (1998) Phosphatidylinositol 4,5-bisphosphate reverses the inhibition of RNA transcription caused by histone H1. Eur J Biochem. 251, 281–287
27. Gozani, O., Karuman, P., Jones, D. R., Ivanov, D., Cha, J., Lugovskoy, A. A., Baird, C. L., Zhu, H., Field, S. J., Lessnick, S. L., Villasenor, J., Mehrotra, B., Chen, J., Rao, V. R., Brugge, J. S., Ferguson, C. G., Payrastre, B., Myszka, D. G., Cantley, L. C., Wager, D., Divate, N., Prestwich, G. D., and Knecht, J. (2003) The PH domain finger of the chimeric orphan receptor ING3 functions as a nuclear phosphonositide receptor. Cell 114, 99–111
28. Krylova, I. N., Sabin, E. P., Moore, J., Xu, R. X., Waitt, G. M., MacKay, J. A., Juzumie, D., Bynum, J. M., Madaus, K., Montana, V., Lebedeva, L., Suzawa, M., Williams, J. D., Williams, S. P., Gu, R. K., Thompson, J. W., Fietterick, R. J., Willson, T. M., and Ingraham, H. A. (2005) Structural analyses reveal phosphatidyl inositols as ligands for the NPR orphan receptors SF-1 and LRH-1. Cell 120, 343–355
29. Li, Y., Choi, M., Cavaey, G., Daugherty, J., Suino, K., Kovach, A., Bingham, N. C., Kliwer, S. A., and Xu, H. E. (2005) Crystallographic identification and functional characterization of phospholipids as ligands for the orphan nuclear receptor steroidogenic factor-1. Mol. Cell 17, 491–502
30. Ortlund, E. A., Lee, Y., Solomon, I. H., Hager, J. M., Saft, R., Choi, Y., Guan, Z., Tripathy, A., Raetz, C. R., McDonald, D. P., Moore, D. D., and Redinbo, M. R. (2005) Modulation of human nuclear receptor LRH-1 activity by phospholipids and SHP. Nat Struct Mol. Biol. 12, 357–363
31. Mortier, E., Uytens, G., Leenaerts, I., Hennes, F., Heung, M. Y., Degeest, G., David, G., and Zimmermann, P. (2005) Nuclear speckles and nucleoli targeting by PI2-PDZ domain interactions. EMBO J. 24, 2566–2565
32. Meerschaert, K., Tun, M. P., Remue, E., De Ganck, A., Boucherie, C., Vanloo, B., Degeest, G., Vandekervkhoche, J., Zimmermann, P., Bhardwaj, N., Lu, H., Cho, W., and Grueterman, J. (2009) The PDZ2 domain of zonula occludens-1 and -2 is a phosphoinositide binding domain. Cell Mol. Life Sci. 66, 3951–3966
33. Víri, K. M., Jánis, J., Siggara, T., Heinonen, T. N., Valjakka, J., Bulyk, M. L., Mäki, M., and Lohi, O. (2009) DNA-binding and -bending activities of SAP30L and SAP30 are mediated by a zinc-dependent module and monophosphoinositides. Mol. Cell. Biol. 29, 342–356
Vann, L. R., Wooding, F. B., Irvine, R. F., and Divecha, N. (1997). Metabolism and possible compartmentalization of inositol lipids in isolated rat-liver nuclei. *Biochem. J.* 327 (pt. 2), 569–576.

Osborne, S. L., Thomas, C. L., Gschmeissner, S., and Schiavo, G. (2001) Nuclear PtdIns(4,5)P2 assemblies in a mitotically regulated particle involved in pre-mRNA splicing. *J. Cell Sci.* 114, 2501–2511.

Watt, S. A., Kular, G., Fleming, I. N., Downes, C. P., and Lucocq, J. M. (2002) Subcellular localization of phosphatidylinositol 4,5-bisphosphate using the pleckstrin homology domain of phospholipase C delta1. *Biochem. J.* 363, 657–666.

Boronenkov, I. V., Loijens, J. C., Umeda, M., and Anderson, R. A. (1998) Phosphoinositide signaling pathways in nuclei are associated with nuclear speckles containing pre-mRNA processing factors. *Mol. Biol. Cell* 9, 3547–3560.

Gregory, T. R., Nicol, J. A., Tamm, H., Kullman, B., Kullman, K., Leitch, I. J., Murray, B. G., Kapraun, D. F., Greilhuber, J., and Bennett, M. D. (2007) Eukaryotic genome size databases. *Nucleic Acids Res.* 35, D332–338.

Oud, P. S., Reubsaet-Veldhuizen, J. A., Beck, H. L., Pahlplatz, M. M., Hesselmans, G. H., Hermens, H. G., Tas, J., James, J., and Vooijs, G. P. (1986) DNA and nuclear protein measurement in columnar epithelial cells of human endometrium. *Cytometry* 7, 325–330.

Gönülüü, U., and Kato, H. (1995) Nuclear RNA and nuclear protein content of tumor cell in adenocarcinoma of the lung. *Lung Cancer* 13, 19–19.

Wang, Z., Castaño, I. B., De Las, Peñas, A., Adams, C., and Christman, M. F. (2000) Pol kappa: A DNA polymerase required for sister chromatid cohesion. *Science* 289, 774–779.

Loyer, P., Trembley, J. H., Lahti, J. M., and Kidd, V. J. (1998). The RNP protein, RNPS1, associates with specific isoforms of the p34cdc2-related PITSLRE protein kinase in vivo. *J. Cell Sci.* 111 (Pt 11), 1495–1506.

Sapra, A. K., Anko, M. L., Grishina, I., Lorenz, M., Pabis, M., Poser, I., Rollins, J., Weiland, E. M., and Neugebauer, K. M. (2009) SR protein family members display diverse activities in the formation of nascent and mature mRNPs in vivo. *Mol. Cell* 34, 179–190.

Cowper, A. E., Cáceres, J. F., Mayeda, A., and Sorensen, A. S. (2001) Serine-arginine (SR) protein-like factors that antagonize authentic SR proteins and regulate alternative splicing. *J. Biol. Chem.* 276, 48908–48914.