

## Global Analysis of Protein Activities Using Proteome Chips

Heng Zhu,<sup>1</sup> Metin Bilgin,<sup>1</sup> Rhonda Bangham,<sup>1</sup> David Hall,<sup>2</sup> Antonio Casamayor,<sup>1</sup> Paul Bertone,<sup>1</sup> Ning Lan,<sup>2</sup> Ronald Jansen,<sup>2</sup> Scott Bidlingmaier,<sup>2</sup> Thomas Houfek,<sup>3</sup> Tom Mitchell,<sup>3</sup> Perry Miller,<sup>4</sup> Ralph A. Dean,<sup>3</sup> Mark Gerstein,<sup>2</sup> Michael Snyder<sup>1,2\*</sup>

<sup>1</sup>Department of Molecular, Cellular, and Developmental Biology, and <sup>2</sup>Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA.

<sup>3</sup>Fungal Genomics Laboratory, North Carolina State University, Campus Box 7251, Raleigh, NC 27695–7251, USA.

<sup>4</sup>Department of Anesthesiology, Yale University, New Haven, CT 06520, USA. \*To whom correspondence should be addressed. E-mail: michael.snyder@yale.edu

**To facilitate studies of the yeast proteome, we have cloned 5800 open reading frames and overexpressed and purified their corresponding proteins. The proteins were printed onto slides at high spatial density to form a yeast proteome microarray and screened for their ability to interact with proteins and phospholipids. We identified many new calmodulin and phospholipid-interacting proteins; a common potential binding motif was identified for many of the calmodulin-binding proteins. Thus, microarrays of an entire eukaryotic proteome can be prepared and screened for diverse biochemical activities. They could also be used to screen protein-drug interactions and to detect posttranslational modifications.**

A daunting task in the post-genome sequencing era is to understand the functions, modification, and regulation of every encoded protein (1). Currently, much effort is devoted toward studying gene, and hence protein, function and regulation by analyzing mRNA expression profiles, gene disruption phenotypes, two-hybrid interactions, and protein subcellular localization (2). Although these studies are useful, much information about protein function can be derived from the analysis of biochemical activities (3–7). In principle, the biochemical activities of proteins can be systematically probed by producing proteins in a high throughput fashion and analyzing the functions of hundreds or thousands of protein samples in parallel using protein microarrays (5, 6, 8). Major hurdles in screening an entire proteome array have been the ability to generate the necessary expression clones and also the expression and purification of proteins in a high-throughput fashion.

We have constructed a yeast proteome microarray containing approximately 80% yeast proteins and screened it for a number of biochemical activities. We first built a high-quality collection of 5800 yeast ORFs (93.5% of the total) cloned into a yeast high-copy expression vector using recombination cloning (9). The yeast proteins are fused to GST-HisX6 at their amino termini and expressed in yeast using the inducible *GAL1* promoter (5, 9). The yeast expression strains contain individual plasmids in which the correct yeast ORFs have been shown to be fused in-frame to GST by DNA sequencing. The proteins were expressed in yeast to help ensure that the proteins were modified and folded properly. Using a 96-well format, 1152 samples were purified at once from yeast extracts using glutathione-agarose beads (10). 0.1% Triton was included in the lysis buffer and washes to ensure that the purified proteins were free of lipids. The quality and quantity of the purified proteins were

monitored using immunoblot analysis of 60 random samples (Fig. 1A). Greater than 80% of the strains produced detectable amounts of fusion proteins of the expected molecular weight.

To prepare the proteome chips, 6566 protein preparations representing 5800 different yeast proteins were printed in duplicate onto glass slides using a commercially available microarrayer. Our initial experiments used aldehyde-treated microscope slides (6) in which fusion proteins attach to the surface through primary amines at their N-termini or other residues of the protein. In subsequent experiments, we spotted proteins onto nickel-coated slides, in which the fusion proteins attach through their HisX6 tags and presumably uniformly orient away from the surface. Although both slides were successful, the nickel-coated slides gave superior signals for our particular protein preparations (Fig. 1B).

To determine how much fusion protein was covalently attached to different glass surfaces and the reproducibility of the protein attachment, we probed the chips with anti-GST antibodies. Over 93.5% of the protein samples gave signals significantly above background (i.e. greater than 10 fg of protein), and 90% of the spots contain 10 fg to 950 fg of protein. Our results also demonstrate that it is feasible to spot 13,000 protein samples in one half the area of a standard microscope slide with excellent resolution (Fig. 1C). To test the reproducibility of the protein spotting, the signals from each pair of duplicated spots were compared with one another; 95% of the signals were within 5% of the average (10).

The proteome chips were tested by probing for several protein-protein interactions and protein-lipid interactions. To test for protein-protein interactions, the yeast proteome was probed with biotinylated calmodulin in the presence of calcium (11). Calmodulin is a highly conserved calcium-binding protein involved in many calcium-regulated cellular processes and has many known partners (12). The bound biotinylated protein was detected using Cy3-labeled streptavidin. As a control we also probed with Cy3-labeled streptavidin alone. These studies identified six known calmodulin targets (Fig. 2A): Cmk1p and Cmk2p are the type I and II calcium/calmodulin-dependent serine/threonine protein kinases (12), Cmp2p is one of the two yeast calcineurins (13), Dst1p plays a role in transcription elongation (14), Myo4p is a class V myosin heavy chain (15), and Arc35p is a component of the Arp2/3 actin-organizing complex (16). Arc35p was recently shown to interact with calmodulin in a two-hybrid study (17); thus our data confirm that Arc35p and calmodulin interact in vitro. Of the six known calmodulin targets that we did not detect, two are not

in our collection and the rest were not detectable in the GST probing experiments. In addition to known partners, the calmodulin probe identified 33 additional potential partners. These include many different types of proteins (Table 1) (10), consistent with a role for calmodulin in many diverse cellular processes.

Sequence searching (5) revealed that 14 of the 39 calmodulin-binding proteins contain a motif whose consensus is I/L-Q-X-X-K-K/X-G-B, where X is any residue and B is a basic residue (Fig. 2B). A related sequence in myosins, I-Q-X-X-X-X-K-X-X-X-R, has been shown previously to bind calmodulin (18). Thus, we demonstrate that the domain is found in many calmodulin-binding proteins. Presumably the other targets that lack this motif have other calmodulin-binding sequences (10).

In addition to the calmodulin-binding targets, we also identified one protein, Pyc1p, which bound Cy3-labeled streptavidin. Pyc1p encodes a pyruvate carboxylase 1 homolog that contains a highly conserved biotin attachment region (19). Thus, as predicted by its sequence, Pyc1p is biotinylated *in vivo*. With appropriate detection assays, we expect that proteome chips can identify many types of posttranslational modification of proteins.

To test whether proteome chips could be used to identify activities which might not be accessible by other approaches, such as protein-drug interactions and protein-lipid interactions, we screened for phosphatidylinositol (PI) binding proteins. PIs are important constituents of cellular membrane and also serve as second-messengers that regulate diverse cellular processes, including growth, differentiation, cytoskeletal rearrangements, and membrane trafficking (20). Because they are often present only transiently and in low abundance within cells, PIs have not been characterized extensively, and little is known about which proteins bind different phospholipids (20).

Five types of PI liposomes and one liposome lacking PIs were used to probe the proteome chips. Each contains phosphatidylcholine (PC) with 1% (w/w) N-(biotinoyl)-1,2-dihexadecanoyl-sn-glycero-3-phosphoethanolamine, triethylammonium salt (biotin DHPE); the biotinylated lipid serves as a label that can be detected by Cy3-streptavidin (21). In addition to PC, the five other liposomes contain either 5% (w/w) PI(3)P, PI(4)P, PI(3,4)P<sub>2</sub>, PI(4,5)P<sub>2</sub>, or PI(3,4,5)P<sub>3</sub> (Fig. 2A). All of these phospholipids have been found in yeast except PI(3,4,5)P<sub>3</sub> (20).

The six liposomes identified a total of 150 different protein targets that produced signals significantly higher than the background; an algorithm was devised to assist in the identification of positive signals (22). Fifty-two (35%) of the lipid-binding proteins correspond to uncharacterized proteins. Of the 98 known proteins, 45 proteins are membrane-associated and either have, or are predicted to have, membrane-spanning regions (23, 24). This includes integral membrane proteins, those with lipid modifications (e.g. the glycosylphosphatidylinositol (GPI) anchor proteins Tos6p and Sps2p (23) and prenylated proteins (Gpa2p and the mating pheromone  $\alpha$ -factor) (25)), as well as peripherally-associated proteins (e.g. Kcc4p and Myo4p which are at the cell periphery (15, 26)). Eight others are involved in lipid metabolism (e.g. Bpl1p) or inositol ring phosphorylation (e.g. Kcs1p), or predicted to be involved in membrane or lipid function (e.g. Ylr020cp has homology to triacylglycerol lipase). Of the 52 uncharacterized proteins, thirteen (25%) are predicted to be associated with membranes (24) and others contain basic stretches as might be expected for electrostatic

interactions with negatively charged lipids. Surprisingly, 19 of the lipid-binding proteins are kinases, and 17 of these are protein kinases.

The phospholipid-binding proteins were sorted into whether they bound lipids strongly or weakly based on the phospholipid-binding signal relative to the amount of GST (Fig. 3) (22). We found that more (72%) of the strong lipid binding proteins (Fig. 3, A and B) were characterized relative to the weakly binding proteins (54%) (Fig. 3, C and D) and more strong lipid binding proteins are known or predicted to be membrane-associated, relative to the weaker binding proteins (Fig. 3, Membrane Column). Interestingly, 13 of 17 of the protein kinases bind very strongly to the PIs. We further grouped the proteins by whether they preferentially bound one or more PI over PC. 99 proteins bound to PC as well or nearly as well as to the PIs (PI/PC < 1.3) (Fig. 3, B and D). However, 49 proteins bound to one or more PIs preferentially (PI/PC > 1.3) (Fig. 3, A and C). Analysis of the strong PI-binding proteins revealed that many of them specifically bound particular PIs. For example, Stp22p, which is required for vacuolar targeting of plasma membrane proteins such as Ste2p and Can1p, preferentially binds PI(4)P (27). Nine protein kinases specifically bind PI(4)P and PI(3,4)P<sub>2</sub> strongly and one binds these lipids weakly. Atp1p, a subunit of the F1-ATP synthase of the mitochondrial inner membrane, also binds PI(3,4)P<sub>2</sub> (28). Sps2p which is localized to the prospore membrane (29) preferentially interacts with PI(3)P. Preferential binding of Myo4p to PI(4,5)P<sub>2</sub> may be important for its interaction at the cell cortex and/or its regulation. No strong lipid binding targets were found that specifically bound PI(3,4,5)P<sub>3</sub> although some proteins bound both this lipid and others (Fig. 3). These results demonstrate that many membrane-associated proteins including integral membrane proteins and peripherally associated proteins preferentially bind specific phospholipids *in vivo*.

Several proteins involved in glucose metabolism were identified as phospholipid binding. This includes a) three enzymes involved in sequential glycolytic steps (phosphoglycerate mutase (Gpm3p), enolase (Eno2p) and pyruvate kinase (Cdc19p/Pyk1p)), b) hexokinase (Hxk1p) and c) two protein kinases (Snf1p and Rim15p). Although unexpected, previous studies indicate that some of these might interact with lipids. Hxk1p binds zwitterion micelles which stimulate its activity (30), and Eno2p is secreted suggesting an interaction with membranes (31). We speculate that either phospholipids regulate steps involved in glucose metabolism or many steps of glucose metabolism occur on membrane surfaces. In this latter case, the phospholipids would serve as a scaffold to efficiently carry out glycolytic steps.

Six proteins not expected to be involved in membrane function or lipid signaling, Rim15p, Eno2p, Hxk1p, Sps1p, Ygl059wp, and Gcn2p, were further tested for PI-binding using two types of standard assays (30). For three proteins, Rim15p, Eno2p, and Hxk1p, PI(4,5)P<sub>2</sub> liposomes were first adhered to a nitrocellulose membrane; different amounts of the GST fusion proteins and a GST control were used to probe the membrane, and bound proteins were detected using anti-GST antibodies. As shown in Fig. 4A, each yeast fusion protein tightly bound PI(4,5)P<sub>2</sub> whereas GST alone did not. We also carried out the reverse assay for GST fusion proteins of Rim15p, Sps1p, Ygl059wp, and Gcn2p (30). Different amounts of these purified proteins were spotted onto nitrocellulose and probed with the six different liposomes

(Fig. 4B); the bound liposomes were detected using an HRP-conjugated streptavidin. As with the microarrays, liposomes bound to each protein, but not the BSA control. Sps1p bound all five PI containing liposomes nearly equally. Rim15p, Gcn2p, and Ygl059wp exhibited different affinities to different liposomes (see Fig. 4B for Rim15p); PI(3)P and PI(4)P and PI(3,4)P<sub>2</sub> bound strongest. In each case, a linear correlation between the binding signal and the level protein was revealed (10). In summary, these results demonstrate that PI-binding proteins identified in the proteome array also bind lipids in conventional assays.

One concern about our experiments is that because proteins are purified from yeast, we might detect indirect interactions through associated proteins. Most of the interactions that we detect are expected to be direct or at least tightly associated with the protein of interest as proteins were prepared using stringent conditions, and for seven samples contaminating bands were not detected using Coomassie staining. Another limitation is that properly folded extracellular domains and secreted proteins are likely to be underrepresented in our collection since GST and a HisX6 tag are fused at the N terminus. Thus, proteins with a signal peptide may not be delivered to the secretory pathway and may not be folded or modified properly, although we did detect three signal peptide-containing proteins, suggesting that at least some are produced and contain functional domains. Another limitation is that not all interactions are detected because not all proteins are readily overproduced and purified in this high throughput approach; we expect that 80% of the arrayed yeast proteins are full-length and at reasonable levels for screening.

Regardless, the use of proteome chips has significant advantages over existing approaches. Random expression libraries are incomplete, the clones are often not full-length, and the libraries are tedious to screen. A recent alternative approach is to generate defined arrays and screen them using a pooling strategy (4). The pooling strategy requires two steps, the actual number of proteins screened is not known, and does not work well when large numbers of reacting proteins exist, as each pool will test positive. Another method for detecting interactions is the two-hybrid approach (2). In this approach interactions are typically detected in the nucleus, which limits the types of interactions that can be detected. The advantage of the proteome chip approach is that a comprehensive set of individual proteins can be directly screened in vitro for a wide variety of activities including protein-drug interactions, protein-lipid interactions and enzymatic assays using a wide range of in vitro conditions. Furthermore, once the proteins are prepared, proteome screening is significantly faster and cheaper. Using similar procedures it is clearly possible to prepare protein arrays of 10-100,000 proteins for global proteome analysis in humans and other eucaryotes.

## References and Notes

1. S. Fields, Y. Kohara, D. J. Lockhart, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 8825 (1999); A. Goffeau *et al.*, *Science* **274**, 563 (1996).
2. P. Ross-Macdonald *et al.*, *Nature* **402**, 413 (1999); J. L., DeRisi, V. R. Iyer, P. O. Brown, *Science* **278**, 680 (1997); E. A. Winzler *et al.*, *Science* **285**, 901 (1999); P. Uetz *et al.*, *Nature* **403**, 623 (2000); T. Ito *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **97**, 1143 (2000).
3. H. Zhu, M. Snyder, *Curr. Opin. Chem. Biol.* **5**, 40 (2001).
4. M. R. Martzen *et al.*, *Science* **286**, 1153 (1999).
5. H. Zhu *et al.*, *Nature Genet.* **26**, 283 (2000).
6. G. MacBeath, S. L. Schreiber, *Science* **289**, 1760 (2000).
7. A. Caveman, *J. Cell Sci.* **113**, 3543 (2000).
8. P. Arenkov *et al.*, *Anal. Biochem.* **278**, 123 (2000).
9. D. A. Mitchell, T. K. Marshall, R. J. Deschenes, *Yeast* **9**, 715 (1993). The expression vector pEGH was created by inserting an RGS-HisX6 epitope tag between the GST gene and the polycloning site of pEG(KG). The yeast ORFs were cloned using the strategy described previously (5), except every step was done in a 96-well format. Plasmid DNAs confirmed by DNA sequencing were reintroduced into both yeast (Y258) and *E. coli* (DH5 $\alpha$ ). The library contains 5800 unique ORFs.
10. For details of 96-well format protein purification protocol, a full list of results from all the experiments, and the design of the positive identification algorithms, please visit our public web site (<http://bioinfo.mbb.yale.edu/proteinchip>) and supplementary material at *Science Online* ([www.sciencemag.org/cgi/content/full/1062191/DC1](http://www.sciencemag.org/cgi/content/full/1062191/DC1)).
11. Biotinylated calmodulin (CalBiochem, USA) was added to the proteome chip at 0.02  $\mu$ g/ $\mu$ l in PBS buffer with 0.1 mM calcium and incubated in a humidity chamber for one hour at room temperature. 0.1 mM calcium was present in buffers in all subsequent steps. The chip was washed three times with PBS at RT. Cy3-conjugated streptavidin (Jackson IR, USA) (1:5000 dilution) was added to the chip and incubated for 30 min at RT. After extensive washing, the chip was spun dry and scanned using a microarray scanner; the data was subsequently acquired with the GenePix array densitometry software (Axon, USA).
12. S. S. Hook, A. R. Means, *Annu. Rev. Pharmacol. Toxicol.* **41**, 471 (2001).
13. M. S. Cyert, R. Kunisawa, D. Kaim, J. Thorner, *Proc. Natl. Acad. Sci. U.S.A.* **88**, 7376 (1991).
14. D. A. Stirling, K. A. Welch, M. J. Stark, *EMBO J.* **13**, 4329 (1994).
15. F. Bohl, C. Kruse, *EMBO J.* **19**, 5514 (2000); E. Bertrand *et al.*, *Mol. Cell* **2**, 437 (1998).
16. D. C. Winter, E. Y. Choe, R. Li, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7288 (1999).
17. C. Schaerer-Brodbeck, H. Riezman, *Mol. Biol. Cell* **11**, 1113 (2000).
18. K. Homma, J. Saito, R. Ikebe, M. Ikebe, *J. Biol. Chem.* **275**, 34766 (2000).
19. J. Menendez, J. Delgado, C. Gancedo, *Yeast* **14**, 647 (1998).
20. G. Odorizzi, M. Babst, S. D. Emr, *Trends Biochem. Sci.* **25**, 229 (2000); D. A. Fruman *et al.*, *Annu. Rev. Biochem.* **67**, 481 (1998); T. F. Martin, *Annu. Rev. Cell Dev. Biol.* **14**, 231 (2000); S. Wera, J. C. T. Bergsma, *FEMS Yeast Res.* **1406**, 1 (2001).
21. Liposomes were prepared using standard methods (52). Briefly, appropriate amounts of each lipid in chloroform were mixed and dried under nitrogen. The lipid mixture was resuspended in TBS buffer by vortexing. The liposomes were created by sonication. To probe the proteome chips, 60  $\mu$ l of the different liposomes were added onto different chips. The chips were incubated in a humidity chamber for one hour at RT. After washing with TBS buffer for three times, Cy3-conjugated streptavidin (1:5000 dilution) was added to the chip and incubated for 30 min at RT.
22. Positives were identified using a combination of the GenePix software which computes a local intensity



background for each spot and a series of algorithms we developed. Details can be found at <http://bioinfo.mbb.yale.edu/proteinchip> and at [www.sciencemag.org/cgi/content/full/1062191/DC1](http://www.sciencemag.org/cgi/content/full/1062191/DC1).

23. M. C. Costanzo *et al.*, *Nucleic Acids Res.* **29**, 75 (2001).
24. M. Gerstein, *Proteins* **33**, 518 (1998).
25. K. Ansari *et al.*, *J. Biol. Chem.* **274**, 30052 (1999).
26. Y. Barral, M. Parra, S. Bidlingmaier, M. Snyder, *Genes Dev.* **13**, 176 (1999).
27. Y. Li, T. Kane, C. Tipper, P. Spatrick, D. D. Jenness, *Mol. Cell Biol.* **19**, 3588 (1999).
28. I. Arnold *et al.*, *J. Biol. Chem.* **274**, 36 (1999).
29. S. Chu *et al.*, *Science* **282**, 699 (1998).
30. A. Casamayor *et al.*, *Curr. Biol.* **9**, 186 (1999); R. Guerra, M. L. Bianconi, *Biosci. Rep.* **20**, 41 (2000).
31. M. Pardo *et al.*, *Yeast* **15**, 459 (1999).
32. We thank K. Nelson and S. Dellaporta for providing invaluable help. We also thank A. Kumar, G. Michaud, and C. Costigan for providing comments on the manuscripts. This research is supported by grants from NIH. H. Z. and R. J. were supported by a postdoctoral fellowship from the Damon Runyon-Walter Winchell Foundation and by an IBM Graduate Research Fellowship, respectively.

2 May 2001; accepted 13 July 2001

Published online 26 July 2001; 10.1126/science.1062191

Include this information when citing this paper.

**Fig. 1.** GST::yeast proteins were purified in a 96-well format. (A) 60 samples were examined by immunoblot analysis using anti-GST antibodies; 19 representative examples are shown. Greater than 80% of the preparations produce high yields of fusion protein. (B) 6566 protein samples representing 5800 unique proteins were spotted in duplicate on a single nickel-coated microscope slide. The slide was probed with anti-GST antibodies (10). (C) An enlarged image of one of the 48 blocks is depicted to the right of the proteome chip.

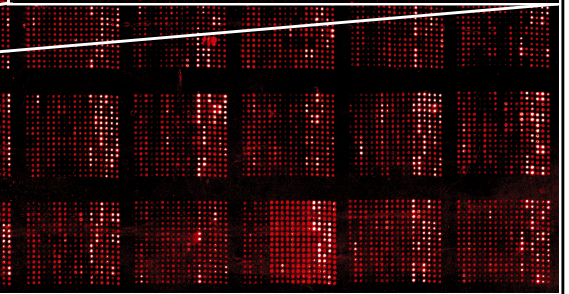
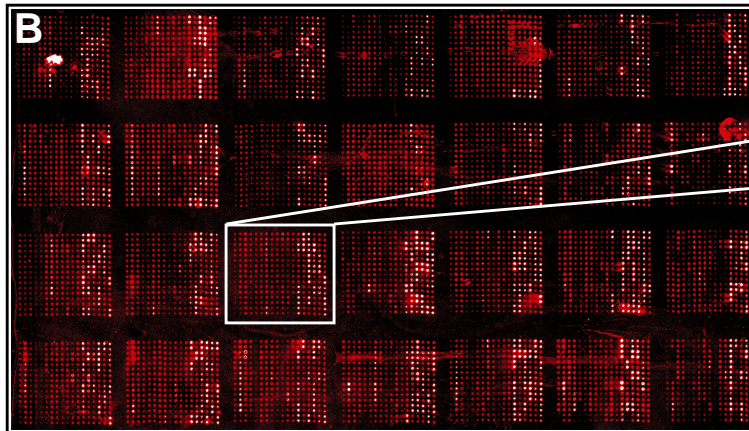
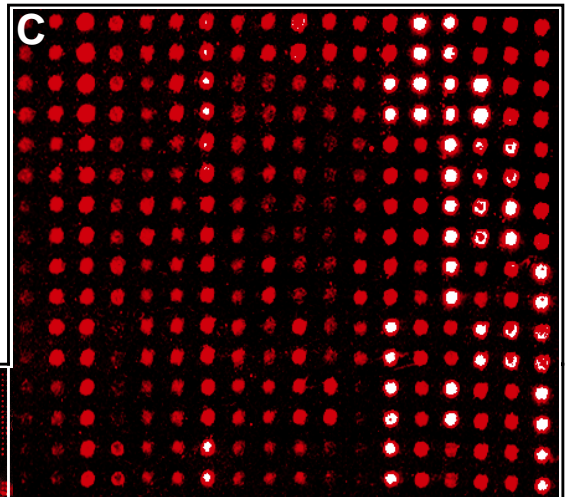
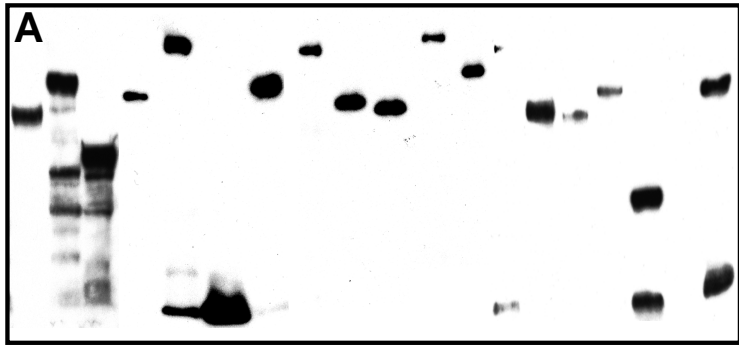
**Fig. 2.** (A) Examples of different assays on the proteome chips. Proteome chips containing 6566 yeast proteins were spotted in duplicate and incubated with the biotinylated probes indicated. The positive signals in duplicate (green) are in the bottom panel; the top panel shows the same yeast protein preparations of a control proteome chip probed with anti-GST antibodies (red). The upper panel shows the amounts of GST fusion proteins as detected by the anti-GST antibodies (red). (B) A putative calmodulin-binding motif; identified by searching for amino acid sequences that are shared by the different calmodulin targets (10). 14 of 39 positive proteins share a motif whose consensus is I/L-Q-X-K-K/X-G-B, where X is any residue and B is a basic residue. The size of the letter indicates the relative frequency of the amino acid indicated.

**Fig. 3.** Analysis of the phosphatidylinositol lipid-binding proteins. To determine the PI-binding specificity of 150 positive proteins, their binding signals were normalized against the corresponding binding signals of PC. Based on the ratios (PI/PC), the proteins were grouped into four categories: (A) 30 strong and specific, (B) 43 strong and nonspecific, (C) 19 weak and specific, and (D) 58 weak and nonspecific PI-binding proteins. The green color intensity represents the PI/PC signal ratio as shown by the scale in the figure. The first column to the right of the PI/PC binding ratios indicates

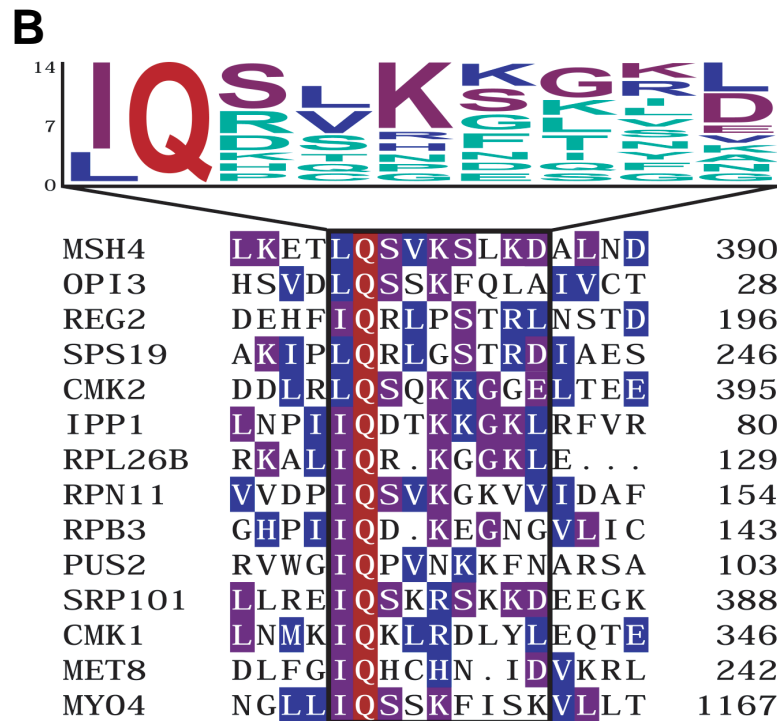
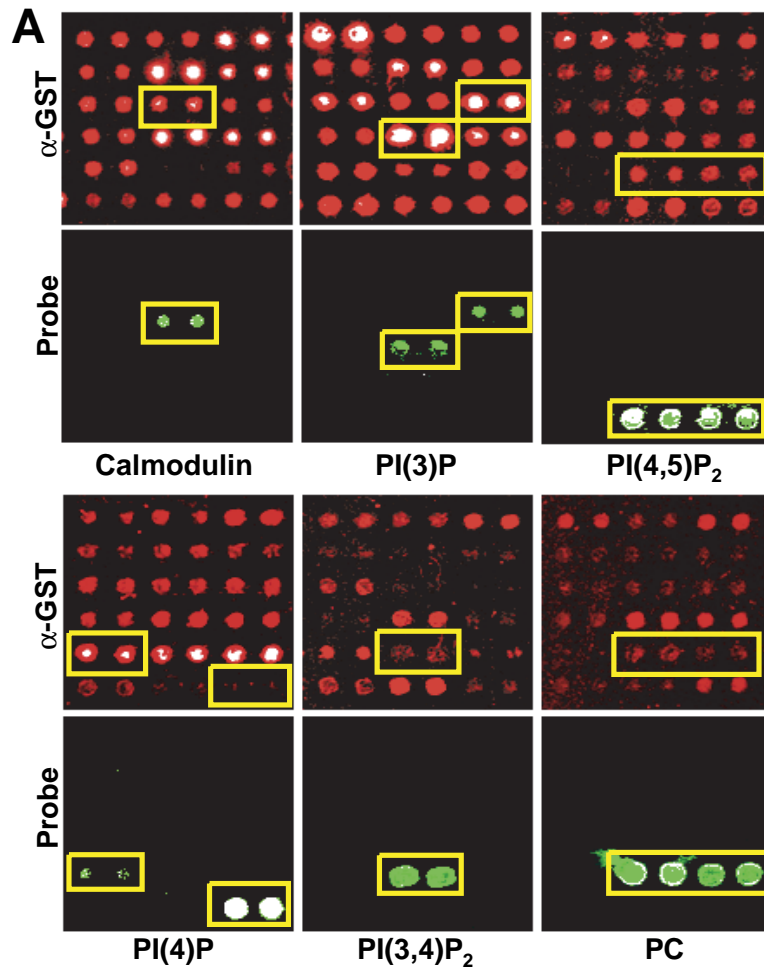
the maximum binding signal intensity (open boxes) and its confidence interval (solid horizontal lines); the numbers indicate the log of the values. Grey, yellow, light-yellow, and red boxes in columns to the right of confidence interval column identify membrane-associated proteins, protein kinases, other kinases, and uncharacterized ORFs, respectively.

**Fig. 4.** Conventional methods confirm protein-lipid interactions detected by the proteome microarrays (30). (A) PI(4,5)P<sub>2</sub> liposomes were first adhered to a nitrocellulose membrane; a dilution series of Rim15p, Eno2p, and Hxk1p, and a GST control were used to probe the membrane. The bound proteins were detected using the anti-GST antibodies and an ECL kit. (B) A reverse assay was carried out to test potential protein-lipid interactions. The proteins were prepared and spotted onto nitrocellulose filters in a dilution series and probed with the six different liposomes. As a control, the six liposomes were also added to the membrane. After extensive washing, the bound liposomes were detected using an HRP-conjugated streptavidin and an ECL kit.

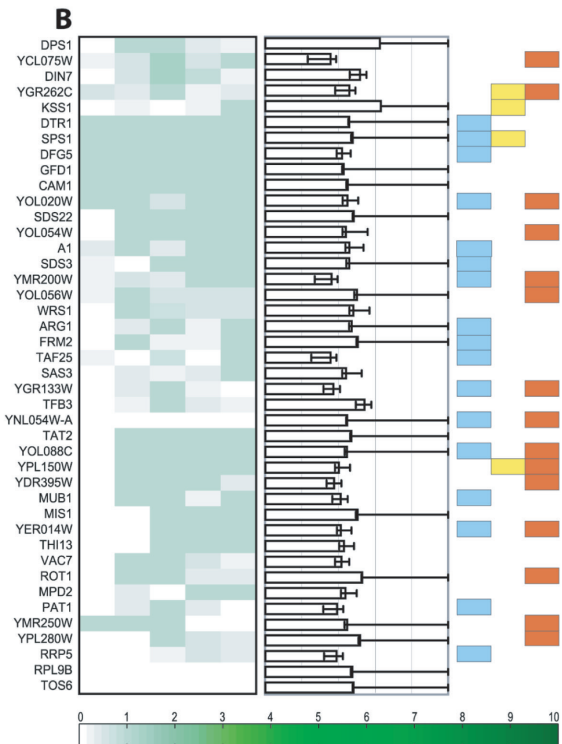
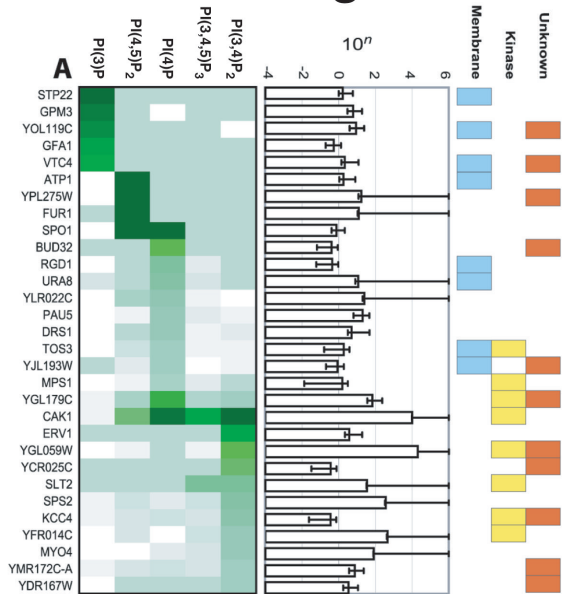




— 2 mm



# Strong



# Weak

