

A Combined Experimental and Computational Strategy to Define Protein Interaction Networks for Peptide Recognition Modules

Amy Hin Yan Tong,^{1*} Becky Drees,^{2*} Giuliano Nardelli,^{3*}
 Gary D. Bader,^{4*} Barbara Brannetti,³ Luisa Castagnoli,³
 Marie Evangelista,⁵ Silvia Ferracuti,³ Bryce Nelson,⁵
 Serena Paoluzi,³ Michele Quondam,³ Adriana Zucconi,³
 Christopher W. V. Hogue,⁴ Stanley Fields,^{2,6†}
 Charles Boone,^{1,5†} Gianni Cesareni^{3†}

Peptide recognition modules mediate many protein-protein interactions critical for the assembly of macromolecular complexes. Complete genome sequences have revealed thousands of these domains, requiring improved methods for identifying their physiologically relevant binding partners. We have developed a strategy combining computational prediction of interactions from phage-display ligand consensus sequences with large-scale two-hybrid physical interaction tests. Application to yeast SH3 domains generated a phage-display network containing 394 interactions among 206 proteins and a two-hybrid network containing 233 interactions among 145 proteins. Graph theoretic analysis identified 59 highly likely interactions common to both networks. Las17 (Bee1), a member of the Wiskott-Aldrich Syndrome protein (WASP) family of actin-assembly proteins, showed multiple SH3 interactions, many of which were confirmed in vivo by coimmunoprecipitation.

Peptide recognition modules mediate many protein-protein interactions critical for the assembly of complexes and pathways that coordinate specific biochemical functions (1). These modules bind to ligands containing a core structural motif; for example, SH3 and WW domains recognize proline-rich peptides, EH domains bind to peptides containing the NPF motif, and SH2 and PTB domains bind to peptides containing a phosphorylated tyrosine (2–4). For particular modules within the same family, binding-partner specificity is determined by key residues flanking the core binding motif (5). Although the complete genome sequence for an organism pro-

vides all of the potential peptide recognition modules and binding partners, a major challenge is to use these data to construct protein-protein interaction networks in which every module is linked to its cognate partners. Here we apply a four-step strategy for the derivation of protein-protein interaction networks mediated by peptide recognition modules:

1) Screen random peptide libraries by phage display to define the consensus sequences for preferred ligands that bind to each peptide recognition module.

2) On the basis of these consensus sequences, computationally derive a protein-protein interaction network that links each peptide recognition module to proteins containing a preferred peptide ligand.

3) Experimentally derive a protein-protein interaction network by testing each peptide recognition module for association to each protein of the inferred proteome in the yeast two-hybrid system.

4) Determine the intersection of the predicted and experimental networks and test in vivo the biological relevance of key interactions within this set.

Because this strategy identifies ligands that bind directly to specific peptide recognition modules and defines interacting partners from the intersection of data sets derived independently, we anticipate that the resultant network will be enriched for physiologically relevant interactions.

We applied this approach to *Saccharomyces cerevisiae* SH3 domains as a test case. With the SH3 domain of the protein kinase Src as a query sequence for ψ -BLAST analysis (6), 24 SH3 proteins were identified within the predicted *S. cerevisiae* proteome (7). Apart from Fus1, which controls cell fusion during mating, and Pex13, which participates in peroxisome biogenesis, most yeast SH3 proteins have been implicated in either signal transduction (Bem1, Boi1, Boi2, Cdc25, Sdc25, and Sho1) or reorganization of the cortical actin cytoskeleton (Abp1, Bud14, Cyk3, Hof1, Myo3, Myo5, Rvs167, and Sla1) (8). A set of eight SH3 proteins [Bbc1 (Mti1), Bzz1, Nbp2, Yfr024c, Ygr136w, Yhl002w, Ypr154w, and Ysc84] remains to be characterized. Bem1 and Bzz1 contain 2 SH3 domains and Sla1 contains 3, with a total of 28 SH3 domains analyzed in this study.

Step 1: We used phage display to select SH3 domain ligands from a random amino acid nonapeptide library (7) and screened all but four SH3 domains (Bem1-2, Cdc25, Sla1-1, and Sla1-2), which could not be expressed in a soluble form as glutathione-S-transferase (GST)-SH3 fusion proteins in *Escherichia coli*. After three selection cycles, positive clones were sequenced, and a consensus ligand was determined for 20 different SH3 domains (Fig. 1). Four SH3 domains—Bud14, Sdc25, Cyk3, and Hof1—did not select a ligand from the nonapeptide library, suggesting that they may not bind to a simple linear peptide with micromolar affinity. To further explore the subset of peptides containing the PxxP motif, we screened a biased library (xxxxPxxPxxxx) (7); however, the same SH3 domains failed to select a preferred ligand. In general, the ligand-binding surface of SH3 domains binds to a core PxxP ligand motif. Class I peptides conform to the consensus RxLPPZP (Z, hydrophobic residues or Arg) and bind in an orientation opposite to that of class II peptides, Px#PxR (9). Most of the yeast SH3 domains selected proline-rich peptides that aligned with the typical Class I or Class II consensus sequence (Fig. 1). Because of ancient chromosomal duplications, several SH3 proteins occur as pairs of paralogs (Myo3/Myo5, Yfr024c/Ysc84, and Ygr136w/Ypr154w). The SH3 domains of paralogs selected highly similar peptides, resulting in a similar consensus (Fig. 1). A few SH3 domains selected peptides conforming to a highly unusual consensus. Bem1-1 SH3 domain selected peptides containing a Ppx-VxPY and Fus1 SH3 domain selected peptides with an RxxR (s/t)(s/t) SI consensus.

Step 2: We used the consensus sequences to search the yeast proteome for potential natural SH3 ligands. For 18 SH3 domains, we compiled a position-specific scoring matrix (PSSM) by calculating the frequency with which each amino acid was found at each

¹Banting and Best Department of Medical Research and Department of Molecular and Medical Genetics, University of Toronto, Toronto, Ontario, Canada M5G 1L6. ²Department of Genome Sciences and Department of Medicine, University of Washington, Box 357730, Seattle, WA 98195, USA. ³Department of Biology, University of Rome Tor Vergata, Via Della Ricerca Scientifica, 00133 Rome, Italy. ⁴Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Ontario, Canada M5S 1A8 and Department of Biochemistry, University of Toronto, Ontario, Canada M5G 1L6. ⁵Department of Biology, Queens University, Kingston, Ontario, Canada K7L 3N6. ⁶Howard Hughes Medical Institute, University of Washington, Box 353360, Seattle, WA 98195, USA.

*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: fields@u.washington.edu, charlie.boone@utoronto.ca, giovanni.cesareni@uniroma2.it

position of the selected nonapeptides. The PSSM contained 9 columns (one for each peptide position) and 20 rows (one for each amino acid). To infer the ligands, we first defined a basic consensus pattern—for example, RxxPxxP or PxxPxR—for each SH3 domain, and then used the PSSM to score all yeast peptides containing the consensus pattern. Peptides with the top 20% scores were considered potential ligands (7).

Because many of the yeast SH3 domain proteins have functionally connected roles in signal transduction and actin assembly, we tested whether they could be represented as a network of interacting proteins (10). The data were first imported into the Bio-molecular Interaction Network Database

(BIND) (11), then formatted with BIND tools (7) and exported for visualization in the Pajek package (12), a program originally designed for the analysis of social networks. The resulting protein-protein interaction map derived from the phage-display analysis (Fig. 2A) contains several known interactions [e.g., Sho1 SH3-Pbs2 (13) and Rvs167 SH3-Abp1 (14)].

Abstracting the network as a graph permits analysis of the interactions with graph theoretical algorithms. Proteins are represented as nodes in the graph and interactions are represented as edges connecting the nodes. A subset of interconnected proteins in which each protein has at least k interactions (where k is an integer) forms a

k -core. These cores represent proteins that are associated with one another by multiple interactions, as may occur in a molecular complex. The k -cores for the phage-display network were computed by using a core finding function in BIND (11) and colored accordingly (Fig. 2A). The most highly connected core of the phage-display network was a single six-core subgraph, i.e., each protein in the subgraph has at least six interactions with other proteins in the subgraph (Fig. 2B). This core may represent a single complex; however, because the network does not take into account temporal expression or protein localization information, other interpretations are possible.

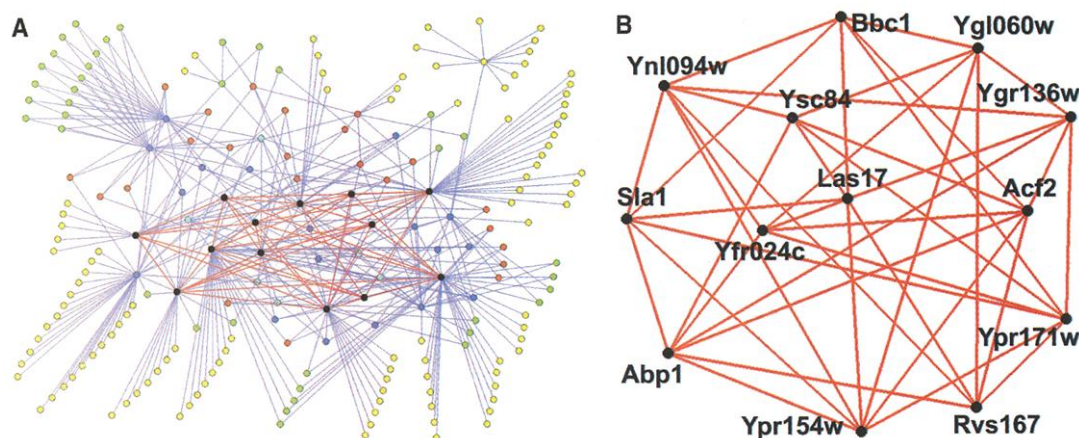
To assess the significance of this six-core, we constructed models of the phage-display network by randomly permuting its interactions. Modeling 1000 different random networks resulted in an average core number of 4.01 (SD = 0.12); therefore, the observation of a highly connected six-core within the phage-display network was unlikely to occur by chance. The proteins within the six-core include several SH3 proteins—Abp1, Sla1, and Rvs167 (8, 14)—involved in cortical actin assembly; Las17, the yeast homolog of human Wiskott-Aldrich Syndrome protein (WASP), which binds to and activates the Arp2/3 actin nucleation complex (15–19); Acf2 (Pcal), a protein required for Las17-dependent reconstitution of actin assembly in vitro (15); and several SH3 proteins of uncharacterized function: Bbc1, Yfr024c, Ypr154w, Ygr136w, and Ysc84.

Step 3: To derive a second protein-protein interaction network for comparison with the predicted phage-display network, we conducted a series of two-hybrid screens (20) with 18 different SH3 domain proteins as well as several proline-rich targets (Bbc1, Bni1, Las17, and Vrp1) as bait (7). We screened many of these proteins or protein

	Class I	Class II	Unusual
Bem1-1			P P x V x P Y
Fus1			R x x R s t s t S l
Abp1		rk x x p x x P x rk P x w #	
Myo3	P x @ p P	P x x x P	
Myo5	P x @ p P	P x x x P	
Pex13	R x l P x # P		
Sla1-3	h R x p P x p P		
Sho1	s kr x L P x x P		
Ygr136w	R x rk # x l P	P x # P x R p	
Ypr154w	@ kr R P p # x l P	P P # P x R P	
Yhl002w	y R p # P x x P		f R x x x h Y t
Ysc84		P x L P x R	
Yfr024c		P p L P x R P	
Rvs167	R x # P x p P	P P # P P R	
Bzz1-1	K kr x P P p x P		
Bzz1-2	kr kr p P P p # P		
Bbc1	R kr x P x p P	p kr # P x R P	
Boi1	R x x P x x P	p P R x P r R #	
Boi2		p P R n P x R #	
Nbp2	P x R P a P x x P		

Fig. 1. Consensus sequence of yeast SH3 peptide ligands. The consensus peptides were derived from an alignment of the selected phage-display peptides (x, any amino acid; lowercase letters, residues conserved in 50 to 80% of the selected peptides; uppercase letters, residues conserved in more than 80% of the selected peptides). Abbreviations for the amino acid residues are as follows: A, Ala; H, His; K, Lys; L, Leu; N, Asn; P, Pro; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr; #, hydrophobic residues; @, aromatic residues. The consensus sequences corresponding to Class I peptides, first column; Class II peptides, second column; unaligned, third column.

Fig. 2. (A) Yeast SH3 domain protein-protein interaction network predicted by means of phage display–selected peptides. In total, 394 interactions and 206 proteins are shown; a network with each gene name labeled is included in the supplementary material (7). The proteins are colored according to their k -core value (6-core, black; 5-core, cyan; 4-core, blue; 3-core, red; 2-core, green; 1-core, yellow), identifying subsets of interconnected proteins in which each protein has at least k interactions. Here, lower core numbers encompass all higher core numbers (e.g., a 4-core includes all the nodes in the 4-core, 5-core, and 6-core). The interactions of the 6-core subgraph are high-lighted in red. (B) The 6-core subgraph derived from the phage-display protein-protein interaction network, expanded to all



low identification of individual proteins. The 6-core subset contains eight SH3 domain proteins (Abp1, Bbc1, Rvs167, Sla1, Yfr024c, Ysc84, Ypr154w, and Ygr136w) and five proteins predicted to bind to at least six different SH3 domains (Las17, Acf2, Ypr171w, Ygl060w, and Ynl094w).

domains against both a genome-wide array of yeast Gal4 activation domain–open reading frame fusions and conventional two-hybrid libraries. In addition, we assayed directly for two-hybrid interactions between the SH3 domains and several proline-rich targets. Most of the resulting interactions (Fig. 3A) have not been reported previously. For example, only seven of the interactions within this network were identified by previous large-scale two-hybrid screens (20–22), indicating that these screens were far from saturating and suggesting that thousands of two-hybrid interactions remain to be identified for the yeast proteome.

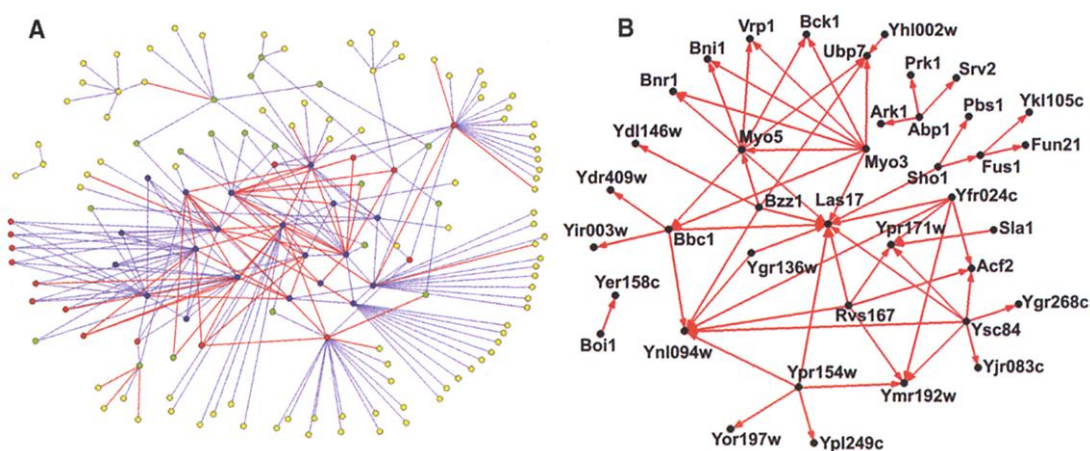
Step 4: We determined the common elements of the phage-display and two-hybrid interaction networks by finding the

intersection of the data sets, where the elements of the data sets are binary protein-protein interactions and the interaction comparisons were considered reflexive (i.e., $A-B = B-A$). Only a subset of the interactions within the two networks is expected to overlap (23). In particular, the phage-display and two-hybrid analysis should identify different sets of false-positive interactions, excluding them from the overlap network. In total, 59 interactions in the phage-display network were also found in the two-hybrid network (Fig. 3B). To determine the significance of this overlap, we created random phage-display networks by keeping the SH3-containing proteins and the number of interactions they participate in as a constant and randomly picking

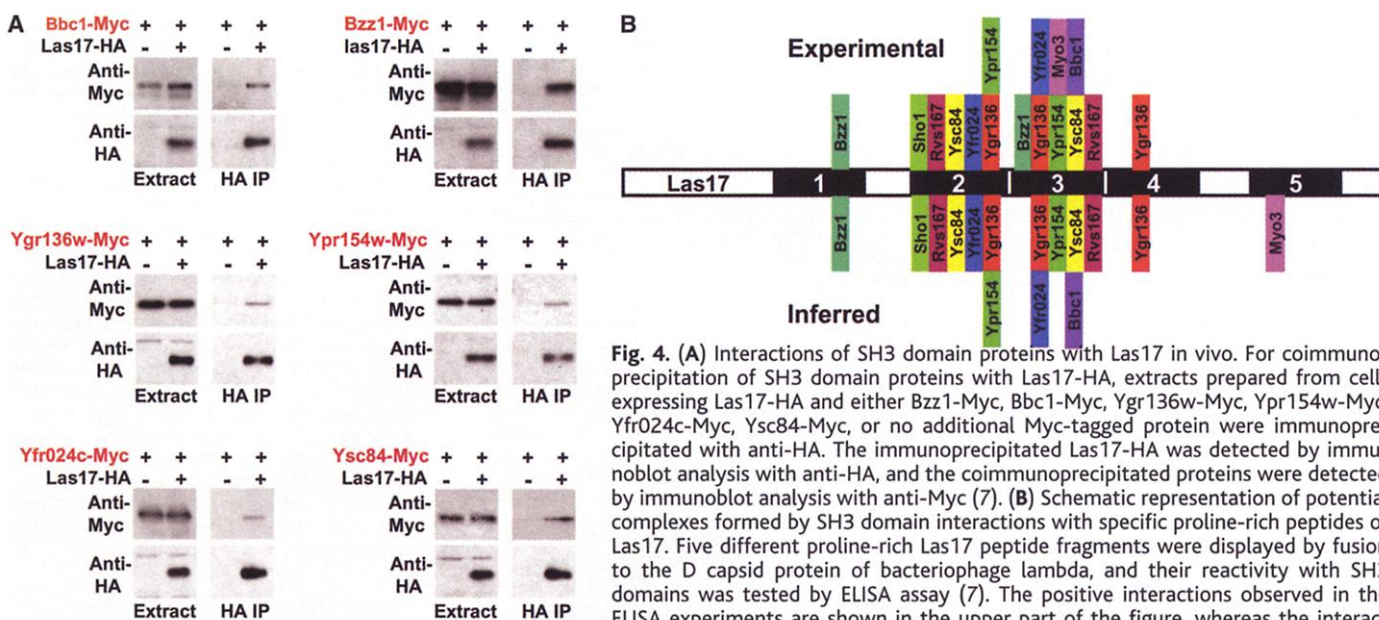
interacting partners from the yeast proteome (7). In 1000 random networks with an average of 206 proteins (SD = 4.05), the average overlap was 0.84 interactions (SD = 1.01). Thus, the phage-display analysis was highly enriched for interactions common to the two-hybrid network. Further, the overlap network was enriched for literature-validated interactions (24), over threefold compared with the two-hybrid network and over fivefold compared with the phage-display network, suggesting that most of these SH3 domain interactions are likely to be physiologically relevant.

To examine the in vivo relevance of some of the interactions predicted by this strategy (Fig. 3B), we focused on further analysis of the WASP homolog Las17, which localizes

Fig. 3. (A) Two-hybrid SH3 domain protein-protein interaction network. Two-hybrid results, based largely on screens with SH3 domains as bait, generated a network containing 233 interactions and 145 proteins. A network with each gene name labeled is included in the supplementary material (7). Proteins are colored according to their *k*-core value (see Fig. 2A). The largest core of the two-hybrid network is a single 4-core (blue nodes). Interactions common to the phage-display network are highlighted in red. **(B)** Overlap of the protein-protein interaction networks derived from the phage display and two-hybrid screens. Expanded view of the common two-hybrid protein-protein interaction network derived from the phage display and two-hybrid screens. All of these interactions are common to both networks.



directly by SH3 domains. The arrows point from an SH3 domain protein to the target protein. Additional evidence to support the relevance of several of these interactions is provided in the supplementary material (7).



tions inferred by phage display are shown in the lower part. The fragment boundaries are Las17-1 (153-190), Las17-2 (306-336), Las17-3 (339-366), Las17-4 (374-403), and Las17-5 (423-476), respectively. For the Myo3/Myo5 paralog pair, only Myo3 was tested by ELISA assay.

to cortical actin patches and interacts directly with several proteins involved in actin assembly. The network overlap predicts that the SH3 domains of 10 proteins may bind to a central proline-rich region of Las17, including three known binding partners Myo3, Myo5, and Rvs167 (16–18); proteins identified previously by two-hybrid screens Yfr024c, Ygr136w, Ypr154w, and Ysc84 (19–22); and previously unidentified partners Bbc1 (25), Bzz1, and Sho1. This extensive set of interactions appears to be specific for Las17 because other actin-assembly proteins with proline-rich regions (Bni1, Bnr1, and Vrp1) were predicted to bind to only the SH3 domains of Myo3 and Myo5. The Las17 interactions appear to occur in vivo, because Myc epitope-tagged versions of six predicted binding partners coimmunoprecipitated with hemagglutinin (HA) epitope-tagged Las17 (Las17-HA) when expressed at normal amounts in yeast (Fig. 4A). In the case of the Bzz1-Las17 interaction, genetic and localization experiments further confirmed its physiological relevance (26). Thus, at least nine different SH3 proteins associate with Las17 in vivo. Most of these proteins are highly conserved (8), suggesting that analogous complexes may occur for WASP-like proteins of higher eukaryotes.

The motifs derived from the phage-display experiments also predict the region of the target protein that binds the SH3 domain (Fig. 4B). To test this prediction, we displayed five Las17 proline-rich peptide fragments as fusions to the D capsid protein on bacteriophage lambda (7) and analyzed the binding of these fragments to a panel of SH3 domains in an enzyme-linked immunosorbent assay (ELISA). Apart from Myo3, whose best predicted target, in the Las17-5 fragment, was not confirmed experimentally, the phage-display ligand algorithm consistently predicted the Las17 fragment that showed the strongest binding (Fig. 4B). These findings indicate that Las17 contains multiple binding sites of comparable affinity for several SH3 domains and suggest that Las17 may form one or more complexes containing multiple SH3 domain proteins.

The strategy described here has several features that make it particularly effective in the identification of relevant protein-protein interaction networks. First, both phage-display and two-hybrid analysis take full advantage of genomic information. Second, the two approaches are highly orthogonal in their respective strengths and weaknesses. Phage display uses in vitro binding and short synthetic peptides, whereas two-hybrid analysis uses in vivo binding and native proteins or protein domains. Third, the combined strategy is rapid and general. It can be implemented

readily for other peptide recognition modules, apart from those that bind to ligands with cell type-specific modifications, and other organisms with a sequenced genome. Fourth, this method predicts precise binding sites.

References and Notes

1. T. Pawson, J. D. Scott, *Science* **278**, 2075 (1997).
2. R. Ren, B. J. Mayer, P. Cicchetti, D. Baltimore, *Science* **259**, 1157 (1993).
3. A. E. Salcini et al., *Genes Dev.* **11**, 2239 (1997).
4. M. F. Moran et al., *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8622 (1990).
5. S. Paoluzi et al., *EMBO J.* **17**, 6541 (1998).
6. A. S. Altschul et al., *Nucleic Acids Res.* **25**, 3389 (1997).
7. The following supplementary material, presented as a description, figure, or table, is available on Science Online at www.sciencemag.org/cgi/content/full/1064987/DC1: supplementary fig. 1 (S1), an alignment of the yeast SH3 domains; supplementary note 1 (S2), materials and methods for the phage-display analysis; supplementary table 1 (S3), the selected phage-display peptides that bind to each SH3 domain; supplementary note 2 (S4), the method for creating the position-specific scoring matrix used to rank the yeast peptides containing ligand consensus sequences; supplementary table 2 (S5), the binding and nonbinding peptides for each SH3 domain; supplementary table 3 (S6), the ligands predicted by the profile analysis; supplementary note 3 (S7), the methods used for visualization of protein-protein interaction networks; supplementary fig. 2 (S8), the phage-display network with each node assigned a gene name; supplementary note 4 (S9), the two-hybrid materials and methods; supplementary table 4 (S10), the positive interactions derived from the two-hybrid assays; supplementary fig. 3 (S11), the two-hybrid interactions derived from directed assays of SH3 domains with proline-rich targets; supplementary fig. 4 (S12), the two-hybrid network with each node assigned a gene name; supplementary note 5 (S13), evidence to support the protein-protein interactions predicted in the overlap network; supplementary note 6 (S14), the materials and methods used for the Las17 coimmunoprecipitation experiments; supplementary note 7 (S15), the materials and methods used to construct the plasmids encoding Bzz1W531S (476–633), Bzz1W615S (476–633), and Bzz1W531S,W615S (476–633); supplementary fig. 5 (S16), the cortical patchlike localization of Bzz1-GFP (green fluorescent protein), Bbc1-GFP, Ysc84-GFP, Ynl094w-GFP, and the synthetic-lethal phenotype associated with bzz1Δ vrp1Δ double mutants; supplementary fig. 6 (S17), the ELISA assay results for phage-display analysis of the Las17 proline-rich peptides; supplementary table 5 (S18), the positive two-hybrid interactions and their BIND identification numbers.
8. References for all the genes mentioned in this study can be found at the following databases: Proteome (www.proteome.com/), Saccharomyces Genome Database (<http://genome-www.stanford.edu/Saccharomyces/>), and the Munich Information Center for Protein Sequences (www.mips.biochem.mpg.de/). Yeast protein-protein interaction data sets can be found at Proteome and the Munich Information Center for Protein Sequences.
9. B. J. Mayer, *J. Cell Sci.* **114**, 1253 (2001).
10. B. Schwikowski, P. Uetz, S. Fields, *Nature Biotechnol.* **18**, 1257 (2000).
11. G. D. Bader et al., *Nucleic Acids Res.* **29**, 242 (2001).
12. <http://vlado.fmf.uni-lj.si/pub/networks/pajek/>
13. T. Maeda, M. Takekawa, H. Saito, *Science* **269**, 554 (1995).
14. T. Lila, D. G. Drubin, *Mol. Biol. Cell* **8**, 367 (1997).
15. T. Lechler, R. Li, *J. Cell Biol.* **138**, 95 (1997).
16. D. Winter, T. Lechler, R. Li, *Curr. Biol.* **9**, 501 (1999).

17. M. Evangelista et al., *J. Cell Biol.* **148**, 353 (2000).
18. K. Colwill, D. Field, L. Moore, J. Friesen, B. Andrews, *Genetics* **152**, 881 (1999).
19. A. Madania et al., *Mol. Biol. Cell* **10**, 3521 (1999).
20. P. Uetz et al., *Nature*, **403**, 601 (2000).
21. T. Ito et al., *Proc. Natl. Acad. Sci. U.S.A.* **98**, 4569 (2001).
22. B. L. Drees et al., *J. Cell Biol.* **154**, 549 (2001).
23. The phage-display analysis should identify a subset of the natural interactions mediated directly by short peptides, missing those that require either longer peptides or further stabilization by noncontiguous residues; moreover, some of the ligands predicted by phage display may not be surface-exposed within a folded protein. In the case of the two-hybrid interactions, there is a recognized potential for false-positives, in part due to overexpression and nuclear targeting of the fusion proteins, and because we are conducting screens of yeast proteins in yeast cells, the interactions may not be direct. Further, some of the SH3 domains screened by two-hybrid interactions were not included in the phage-display network and vice versa because some bait proteins yielded results only in one assay.
24. YPD (www.proteome.com/DB-demo/intro-to-YPD.html) contained 89 interactions that involved one of the SH3 domain proteins that we studied. Of these interactions, 17 were found in the phage-display network (394 interactions in total), 16 in the two-hybrid network (233 interactions in total), and 13 in the overlap network (59 interactions in total). Thus, the overlap network maintains most of the literature-validated interactions of the phage-display and two-hybrid network.
25. The SH3 domains of yeast type I myosins bind to a proline-rich ligand in Bbc1 (Mti1) (Fig. 3B) (27), and a Bbc1 deletion, bbc1Δ, suppresses the temperature sensitivity and endocytosis defects of vrp1 mutants (27).
26. We found that Bzz1-GFP concentrated in cortical patchlike structures (7), which localized to sites of polarized growth, suggesting that Bzz1 may be targeted to cortical actin patches, like components of the Las17-Vrp1-Myo3/5 complex (16, 17). Moreover, a similar patchlike localization pattern was observed for Ynl094w-GFP, Bbc1-GFP, and Ysc84-GFP (7), supporting the predicted Bzz1-Ynl094w, Bbc1-Las17, Bbc1-Ynl094w, and Ysc84-Las17 interactions (Fig. 3B). From a synthetic lethal screen in which a strain carrying a complete deletion of the BZZ1 gene, MATα bzz1Δ, was crossed to an ordered array of ~2000 different MATα gene-deletion strains (28), we found that bzz1Δ vrp1Δ double-mutant meiotic progeny showed reduced fitness specifically (7). This genetic interaction suggests that Bzz1 functionally associates with Bee1-Vrp1-Myo3/5 actin-assembly complex.
27. J. Mochida, T. Yamamoto, K. Fujimura-Kamada, K. Tanaka personal communication.
28. E. H. Winzler et al., *Science* **285**, 901 (1999).
29. We thank B. Kay for the PxxP biased library; A. Burgess, K. Kennedy, C. Lee, N. Melo, A. Parsons, and I. Pot for help with the conventional two-hybrid screens; and A. Davidson for comments on the phage-display network. Supported by grants from the Canadian Institute of Health Research (CIHR) and the National Cancer Institute of Canada (C.B.); CIHR (C.H.); AIRC, the European Union framework 4 program, and the CNR target project Biotechnology (G.C.); and the National Center for Research Resources (P41 RR11823) (S.F.). C.B. is an International Research Scholar of the Howard Hughes Medical Institute, and S.F. is an investigator of the Howard Hughes Medical Institute.

1 August 2001; accepted 6 December 2001

Published online 13 December 2001;

10.1126/science.1064987

Include this information when citing this paper.