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,2 Silvia Ferracuti,3 Bryce Nelson,5 Serena Paoluzzi,3 Michele Quondam,3 Adriana Zucconi,3 Christopher W. V. Hogue,4 Stanley Fields,2,6† Charles Boone,1,5† Gianni Cesareni3†

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 communoprecipitation. We applied this approach to *Saccharomyces cerevisiae* SH3 domains as a test case. With the SH3 domain of the protein kinase Srlc as a query sequence for ω-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, Csk3, Ho1, Myo3, Myo5, Rvs167, and Slal1) (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 Slal 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, Slal-1, and Slal-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, Csk3, and Ho1—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 PxP motif, we screened a biased library (xxxxPxPxxxxx) (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 PxP 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-PxR motif with an RxRx(L/S)x(S/L) 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 position.
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, RxxPXP 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 (Pca1), a protein required for the 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

**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 subsectors 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 allow 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 6 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...
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 Yrf024c, 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-S5 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.