DRUG DISCOVERY TODAY TARGETS

Methods to reveal domain networks

Elena Santonico, Luisa Castagnoli and Gianni Cesareni

The development and application of high-throughput technology to study protein interactions has led to the construction of complex interaction maps, the correct interpretation of which is crucial to the identification of targets for drug development. Here we propose that a more informative description of protein interaction networks can be achieved by considering explicitly the modular nature of proteins. In this representation, proteins are drawn as covalently linked modular domains binding to their target sites in partner proteins. Families of conserved modules that bind to relatively short peptides mediate a large fraction of the non-covalent interactions linking different proteins in the network. As these interactions are often involved in the propagation of signal transduction, determining the recognition specificity of each domain family member is an essential step toward a functional description of the global interactome.

Protein interactions preside over cell physiology and people believe that a complete and quantitative understanding of the protein interaction mesh could eventually enable us to predict the response of a cell to any given stimulus. This idea has fuelled the development of several high-throughput strategies aimed at a comprehensive description of the protein interaction network (interactome) in model organisms [1–4]. The first complete analysis of the yeast interactome was achieved independently by two groups using both the two-hybrid method, albeit in slightly different formats [1,2]. Somewhat disappointingly, the overlap between the two inferred interactomes is only 10 to 20%.

More recently, two reports have demonstrated that the classical pull-down (co-immunoprecipitation) method, coupled with identification of the co-precipitated proteins by mass spectrometry, can be carried out at a proteome level [3,4]. Still, the overlap between the networks obtained by the yeast two-hybrid approach and by the pull-down approach is only ~15%. This analysis indicates that the available experimental approaches either target specific and different areas of the interactome and/or are affected by a large number of false positives and false negatives.

Despite these limitations, these four reports have for the first time allowed us to glimpse the global interactome of an organism, which represents an impressive achievement.

The complex graph in Figure 1 depicts a selected fraction of the available information about the protein interaction network of a yeast cell [5]. Circles represent proteins and the lines connecting the circles indicate that the two connected proteins have been found to interact, in some condition, in the yeast cell.

Although we recognize that such an intricate graph contains valuable biological information, we are often startled by the complexity of the connections and suffer from the lack of a methodological framework to map the traditional pathway representation we are used to. This is partly because of

Elena Santonico
Luisa Castagnoli
Gianni Cesareni*
Department of Biology, University of Rome Tor Vergata, Via della Ricerca Scientifica, Rome, Italy
*e-mail: Cesareni@uniroma2.it

1359-6446/05/$ – see front matter PII: S1359-6446(05)03513-0

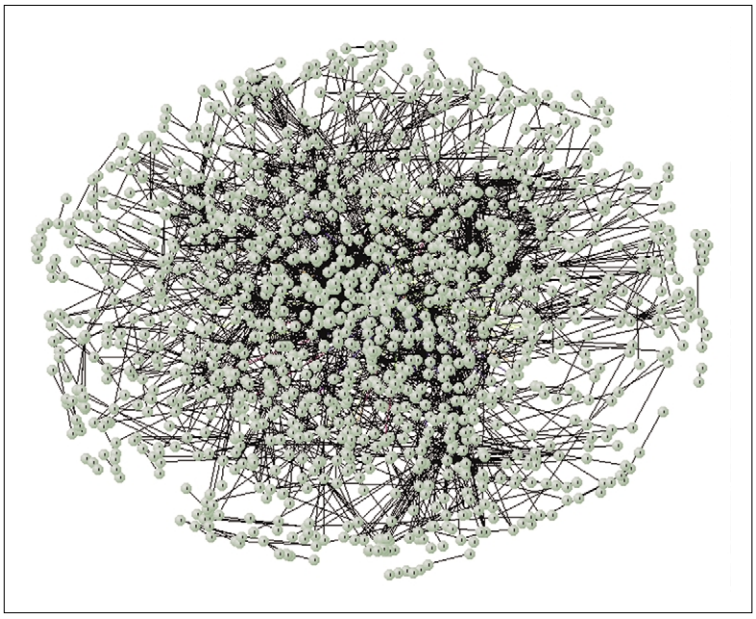

FIGURE 1

Graphical representation of the yeast interaction network. To avoid overcrowding only 2500 high confidence interactions are shown [5]. The available information about protein interactions in yeast comprises more than 14,000 interactions. Proteins are represented as circles and whenever there is evidence of physical interaction, they are connected by edges (lines). The graph was obtained using the Visant software [45].

the perception that a substantial fraction of the edges (connecting lines) in this graph could be artefactual. Conversely, many well-characterized and functionally relevant interactions are missed by the above mentioned high-throughput approaches [6]. For instance, the complex purification approach only detects relatively stable complexes that can stand the purification procedure. Even if entirely accurate, this representation has limitations. To use an analogy, it would be equivalent to having the complete road map of a large city without any information as to the traffic flow or to which paths represent large traffic arteries and which ones are narrow one-way alleys. The challenge is to transform this complex and apparently inextricable diagram into a set of connected biological pathways.

Eventually, both the basic and the applied biologist would hope to be able to trace a thicker line along the edges mediating any specific cell response to any given stimulus. This would help us to understand how to perturb this interaction chain so as to modulate a physiological or pathological cell response.

To achieve this goal one needs to take into account different types of genomic information and to use it to filter out the interactions, which, although biochemically possible, are biologically unlikely. For instance, the two partner proteins might not be expressed at the same time or in the same tissue. Alternatively, despite being co-expressed they might be segregated in different cell compartments. Thanks to the availability of increasingly larger datasets of protein co-expression and co-localization, it is becoming feasible to identify inferred interactions that are likely to be false positives because either the two partner proteins are not expressed in the same cell type or because the two partners are segregated in different cellular compartments. Moreover, information about mRNA and protein concentrations under different physiological conditions and subcellular contexts [7–10] can be used to this end.

To date, this information is only available and systematically organized for the yeast *Saccharomyces cerevisiae;* however, we can expect similar data to be produced in the next few years for mammals. The challenge is to develop frameworks for the integration of information from high-throughput data obtained by different approaches and to combine it with the results of traditional biochemical approaches [11,12].

Domain network versus protein networks

Another feature that limits our ability to interpret large interaction datasets, such as the one pictured in Figure 1, is that a large number of proteins in the network has the potential to interact with many partner proteins at a time. These 'hubs' are more numerous than we would expect if the protein network were assembled at random and represent a typical feature of biological networks [13]. From the representation in Figure 1, however, it is not clear how many protein partners these hubs would be able to concurrently engage with. Furthermore, it is practically impossible to predict the composition of the multiprotein complexes that are formed *in vivo*. We can partly solve this ambiguity by considering the modular nature of proteins.

It has become increasingly clear over the past decade that unrelated proteins frequently share significant portions of sequence similarity [14–16]. These regions of similar sequence often function as independently folded 'modules' or 'domains' capable of performing specific tasks. Thus, a large number of functionally diverse proteins can be thought of as molecules built by combining a limited number of structurally stable folded domains. Within these multidomain proteins, each domain is capable of autonomous function, and some of them also mediate the formation of complexes with partner proteins forming what we would call a domain network.

Thus, to obtain a clearer representation of the protein network we can split each protein into domains whereby each domain can interact with several partners, but only one at a time and not simultaneously (Figure 2). This representation is information-rich and allows clearer conclusions to be drawn on the biological outcome of interfering with any particular edge in a network. For instance, the simple network in Figure 2 represents several interactions that must occur for a given 'stimulus' to be propagated from a receptor A to the two effectors E and F to modulate a specific 'function'. The classical 'protein network' representation on the left does not allow us to predict the outcome of inhibiting, for instance with a

drug, the interaction of C with A, because it cannot be anticipated whether the drug will also affect the interaction of B with A. By contrast, the two alternative domain networks on the right incorporate explicitly the information about the modular structures of the six nodes of the graph and use edges to connect domains instead of proteins, allowing two different conclusions to be reached. In the hypothesis of network a) it is not possible to affect signal transduction by targeting protein A with a single drug, whereas in the hypothesis of network b), a competitor molecule binding to the receptor pocket of domain A2 will prevent the propagation of the information to the effector protein E.

In other words, a more informative representation of the protein interaction network is one in which each node of the graph is fragmented into several covalently linked nodes, each representing a protein domain. In this picture, the edges entering a node represent mutually exclusive interactions and link to protein partners that compete for binding to the same domain (Figure 2). The results of existing high-throughput experiments do not, however, support such a representation because they do not provide enough information about the topology of the interaction. This limitation calls for the development and refinement of strategies that have a high-throughput potential and are specifically aimed at the characterization of domain networks.

Protein interaction domains binding to short linear peptides

Structural analysis of functional protein complexes is consistent with the notion that many proteins interact via extended surfaces, including residues that are far apart in the protein primary structure and only come together upon protein folding. However, pioneering work on the Src kinase, in the late eighties and early nineties, demonstrated for the first time the importance of relatively small protein recognition domains in mediating the formation of protein complexes by binding to short linear peptides [17–20]. SH3 and SH2 domains are the prototypes of a growing number of domain families that share the property of binding to relatively short peptides. In fact, a large proportion of interactions in the protein network is supported by members of such domain families each characterized by preferences for binding to peptides sharing specific sequence or structure characteristics. For instance, SH3 domains mostly bind to proline-rich peptides containing a PxxP motif (in single-letter amino acid code where x is any amino acid) folded in a proline type II helix [21], whereas SH2 domains have affinity for peptides containing a phosphorylated tyrosine residue [22].

These protein domains act as independent binding modules displaying a single receptor surface that can accommodate relatively short peptides, normally binding

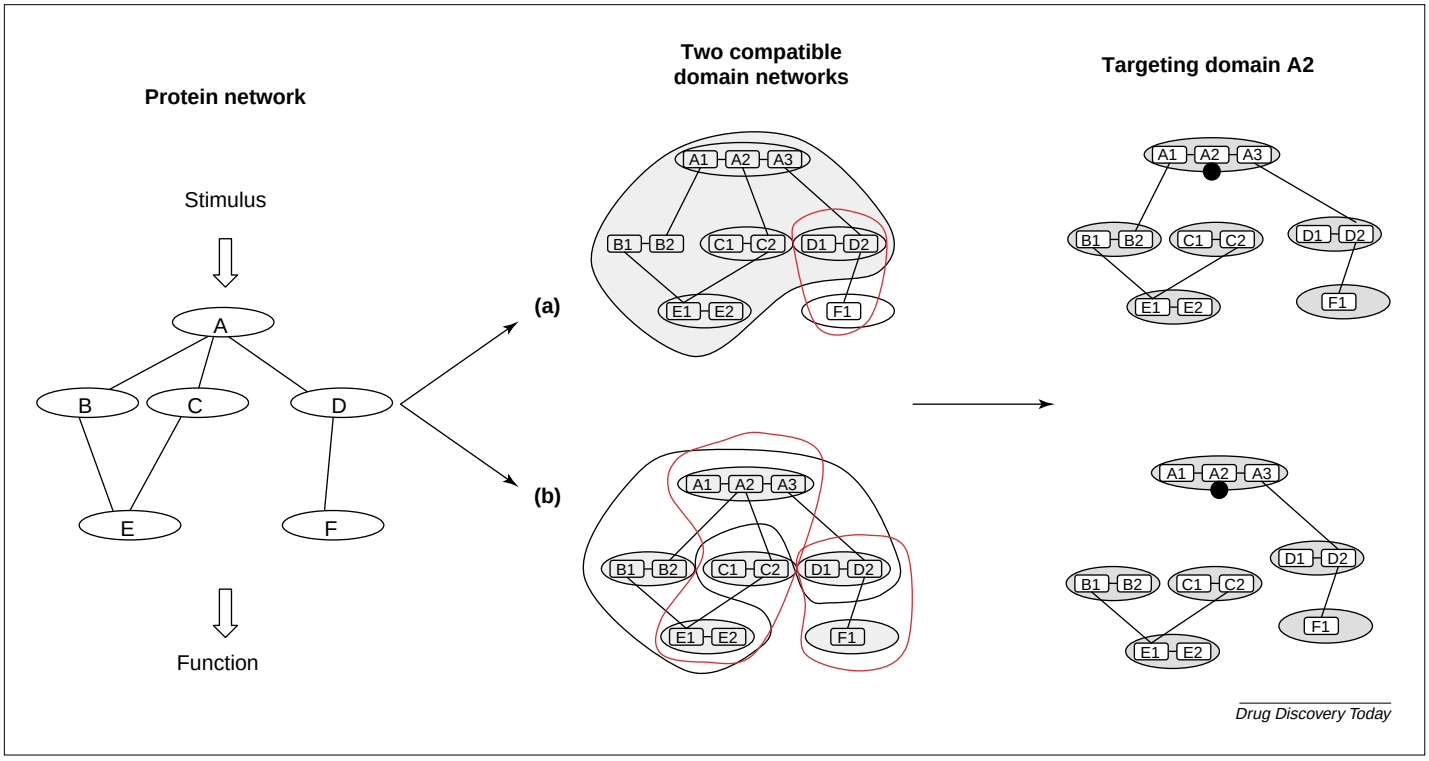

FIGURE 2

Protein networks versus domain networks. On the left-hand side is a schematic representation of a simple protein interaction network linking a receptor protein A to the effectors E and F. In the center are depicted two alternative domain networks consistent with the protein network on the left. In this representation proteins are depicted as ovals including rectangles representing protein interaction domains. Unlike the network on the left, edges connect specific domains and not proteins as a whole. Thick lines include proteins that can form a complex; different complexes are shadowed with different tones of grey. Not all possible complexes are shown. The two domain networks differ in a single connection; however, the complexes that can form are strikingly different. This is reflected on the far right by the very different outcome of targeting with an inhibitor (small black circle) the recognition pocket of domain A2.

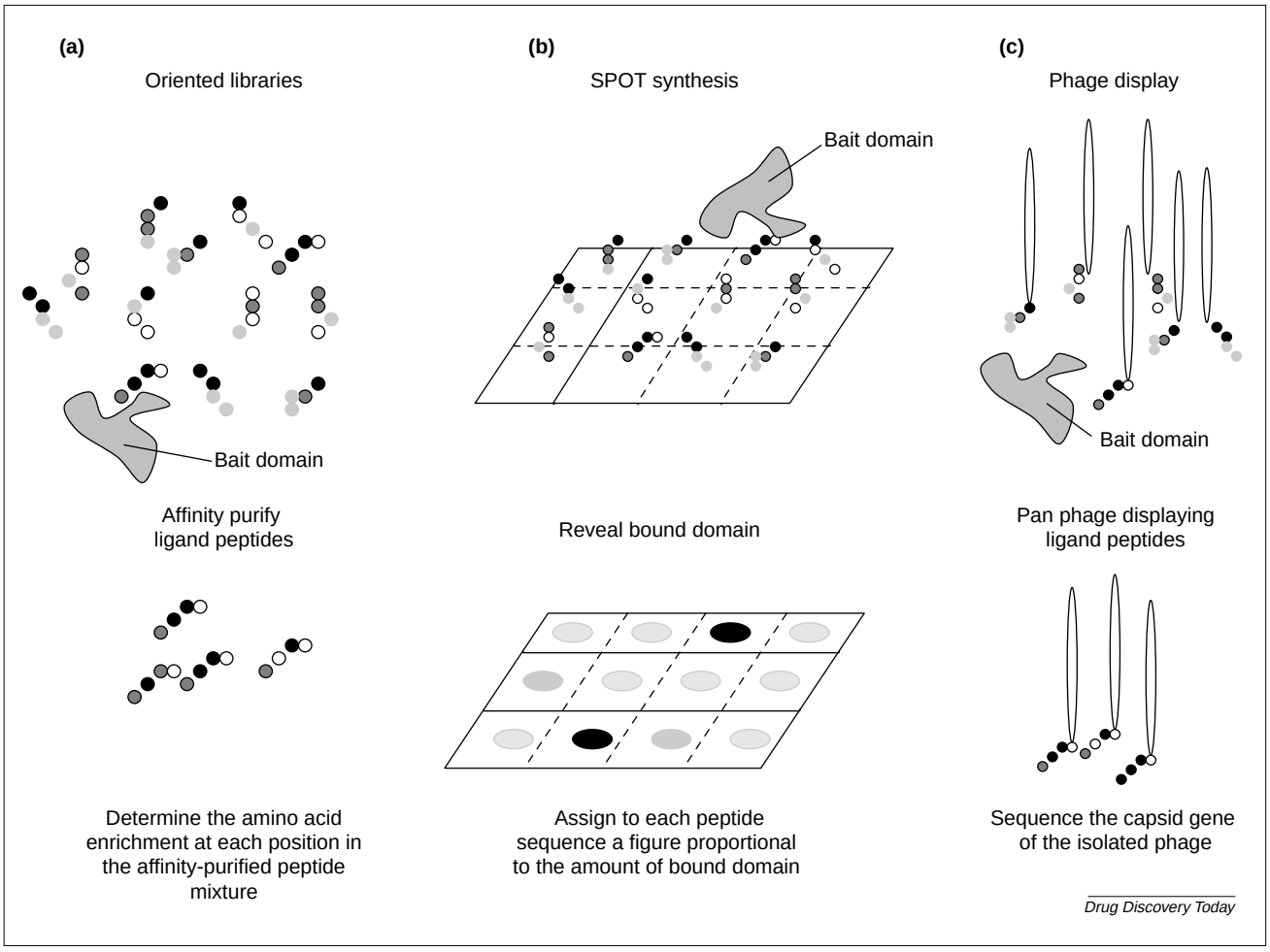

FIGURE 3

Outline of the three approaches to investigate domain recognition specificity. Small circles represent amino acids covalently linked to form peptides. The peptides to be tested are presented to the bait domain **(a)** in solution (oriented peptide libraries), **(b)** bound to cellulose membranes (SPOT synthesis) or **(c)** bound to phage capsids (phage display). After panning or screening, the peptide recognition specificity of the domain under study is inferred from comparison of the amino acid sequence of the binding peptides.

their receptor domains in an extended conformation. The interactions mediated by these domains are dynamic, with dissociation constants often in the micromolar range, and display a low stoichiometry, meaning that only a fraction of each protein is engaged in any specific interaction. For these reasons, methods based on the purification of complexes and the characterization of co-purified proteins often fail to reveal partners even when the proteins are proven, by other methods, to have a physiologically relevant physical interaction [23].

In the following, we will describe experimental approaches that have high-throughput potential and allow mining of detailed information about the topology and stoichiometry of domain–peptide interactions.

Methods to investigate domain recognition specificity

In principle, general methods to study protein interactions like the yeast two-hybrid and pull-down approaches can also be applied to the task of determining a domain network. Alternative strategies involve methods particularly suited to the characterization of the peptide recognition specificity of each member of a domain family. The 'oriented peptide library' approach, SPOT synthesis and phage display, the three technologies that have been employed for this purpose, are schematically illustrated in Figure 3.

The oriented peptide library approach (Figure 3a) is based on the assembly of a biased peptide pool where most amino acid positions are degenerate while certain amino acids at the 'orienting' positions are conserved [24–26]. For instance, in the case of a library designed to study recognition specificity mediated by domains binding phosphotyrosines, the orienting positions will be occupied by a phosphotyrosine in each peptide in the mixture. Similarly, peptides to characterize SH3 binding may contain two 'orienting' prolines separated by two residues of degenerate composition (PxxP). The degenerate peptide pool is incubated in solution with the domain of interest and the adsorbed peptides are then sequenced to determine enrichment of any particular amino acid at any position. Using this approach, the preferred contexts of a phosphorylated residue in the peptide targets of several phosphopeptide binding domains have been determined

[25,26]. Although this method has proved to be very powerful, in essence only the group that originally developed the method has used it extensively. This is probably because it requires some skill in peptide chemistry and instruments not available in every standard molecular biology laboratory. A disadvantage of this approach is that the binding peptides are analyzed in a pool. Consequently, a domain binding two classes of peptides, typified by different consensus sequences, with comparable affinities will be characterized by an amino acid preference matrix that is an average of the matrices of the two consensus sequences. As a result, the algorithm utilizing the preference matrix determined by this approach will be less effective in inferring binding peptides belonging to either of the two classes.

A second approach, based on the chemical synthesis of peptides, is known as the SPOT synthesis method and was originally developed by Frank [27] (Figure 3b). This method is based on the ability to synthesize thousands of peptides in an array format on cellulose membranes and to test them for binding to a domain of interest. The array format offers the advantage of being able to test peptides independently and to associate a figure to each peptide that correlates with the dissociation constant. Unlike the other two approaches described here (oriented peptide libraries and phage display), this method is capable of determining precisely which peptides among a large collection bind to a domain of interest and which ones do not. On the other hand, this approach is limited by the number of peptides that can be synthesized on a membrane of reasonable size in a realistic time (~10^4 peptides). Given this restriction, the SPOT synthesis approach is not readily applicable to the problem of determining the recognition specificity of new domains in the absence of any *a priori* information, as one can only explore a relatively small fraction of the sequence space.

This task is better approached by a third technology, phage display, which consists of panning libraries of 10^9–10^{10} peptides of random sequence displayed on bacteriophage capsids (Figure 3c). Displayed peptides that have affinity for the domain used as a bait can be enriched and purified, and the sequence of the displayed peptide can be deduced from the DNA sequence of the capsid gene [28]. Sequence alignment of the selected peptides makes it possible to identify the consensus sequence of a 'preferred ligand' of the bait domain. The application of phage display to the determination of domain recognition specificity was pioneered by Sparks *et al.* [29] and Rickles *et al.* [30]. Since then, a number of groups has successfully used this approach to determine the recognition consensus of several SH3, SH2, PTB, WW, EH and GYF domains [31–36].

It needs to be pointed out that the strategies described in this section only aim to find all the peptides in a proteome that have the 'biochemical' potential to bind any specific domain and do not address the functional relevance of these inferred interactions.

In addition, both the oriented peptide library and the phage-display approaches are based on several assumptions. The most fundamental one is that protein recognition domains have a single molecular pocket acting as a receptor for a short peptide in the partner protein. As a corollary, the peptides selected by the *in vitro* approaches provide information about the sequence characteristics of the physiological targets, as it is highly unlikely that the *in vitro* and the *in vivo* targets bind to different pockets in the receptor domain.

Although this assumption may seem somewhat arbitrary, the recent success of approaches exploiting panning of random peptide libraries to identify physiological partners lends credibility to this strategy. For instance, we have recently used phage display, in a genomic approach, to determine the peptide ligand consensus of 20 SH3 domains of the yeast *S. cerevisiae* [37]. A search of the proteome by a combination of regular expression and position-specific scoring matrices has allowed a network of 394 interactions connecting 206 proteins to be inferred.

Combining phage display and the SPOT synthesis approach

Both the oriented peptide library and the phage display approaches, when compared with SPOT synthesis, have the advantage of being selective techniques that offer the possibility to explore large repertoires of peptide sequences (>10^8) containing a significant fraction of sequence space. On the other hand, they only provide general recognition rules while the identification of putative ligands in the proteome requires an inference step.

The peptide profile used to scan a proteome is typically derived from a collection of 10–20 peptides selected by panning phage displayed peptide libraries. This peptide collection is often sufficient to identify those few positions that are conserved among the ligands. However, it does not provide statistically significant information about the remaining positions, which are apparently variable but often mask more subtle recognition rules. Furthermore, it is often observed that some peptides, which do not exactly match the strictly conserved pattern, bind to the receptor domain with similar or marginally lower affinity. As a consequence, the inference algorithms have a propensity to over-predict. Conversely, they miss some physiological partners that do not contain an exact match to the consensus.

These limitations became clear in the large-scale characterization of the yeast interactome supported by the family of SH3 domains. A comparison of the network inferred by phage-display experiments with the network obtained by searching the potential SH3 targets by the yeast two-hybrid approach revealed networks of comparable size, but they had only ~25% of the interactions in common. This suggests that, on the one hand, the two experimental approaches lead to overprediction (i.e. a large number of false positives), whereas on the other hand neither approach could discover all the potential targets in

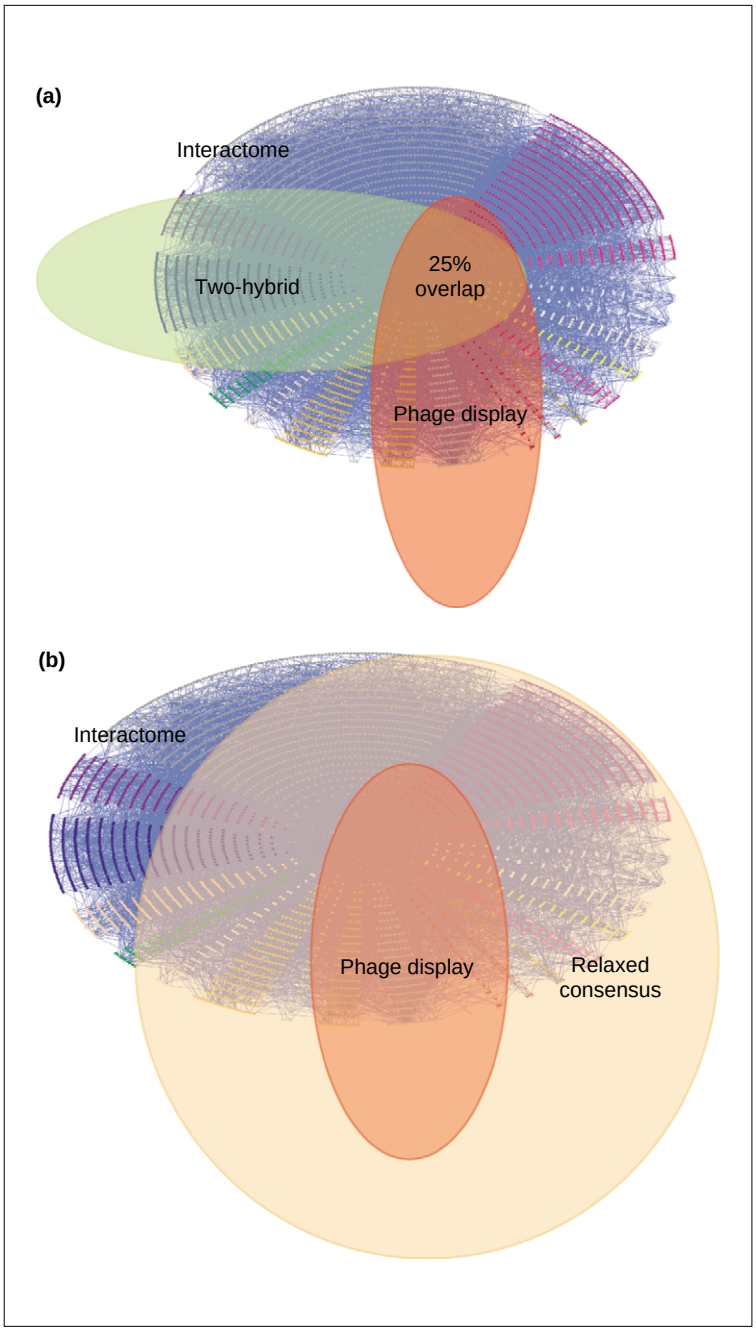

FIGURE 4

Interaction space covered by the different approaches for investigating domain recognition specificity. The network in the background represents the fraction of the physiological interactome that we aim to cover with the different approaches. **(a)** Coverage and accuracy of the networks inferred by phage display (red) and by the yeast two-hybrid (green) approaches. As the two technologies are orthogonal, the inferred networks are largely non-overlapping because they are biased by different false positives (interactions that are predicted but not observed in the physiological network) and false negatives (physiological interactions that are not revealed by the methods). The overlap between the two-hybrid and phage-display networks is all included in the 'physiological interactome'. **(b)** Schematic representation of the relationship between the networks inferred by a strict (red) and a relaxed (sepia) phage-display consensus. The phage display consensus is a strict consensus that often misses physiological targets that may not have a perfect match with the consensus sequence defined by the high-affinity peptides selected by phage display. By relaxing the consensus most of the 'physiological interactions' are included at the cost of including many false positives. In the WISE (Whole Interactome Scanning Experiment) approach the false positives are filtered out by testing them as single peptides in an array format (SPOT synthesis approach).

the proteome (i.e. a large number of false negatives; Figure 4a). Interestingly, the small overlap network consisted of 'reliable' interactions that were validated by *in vivo* co-immunoprecipitation of the inferred partners, suggesting that exploring the same protein–protein interaction space with two 'orthogonal' approaches removes false positives.

To overcome the inherent specificity and accuracy limits of the network inferred by the phage-display approach, we have recently designed a novel two-step strategy that combines the relative strengths of a selective approach (such as panning combinatorial peptide libraries displayed on phage) with the quantitative analysis that can be achieved by screening a large number of peptides arrayed at high density on a solid support [38]. We refer to this approach as WISE for Whole Interactome Scanning Experiment.

This method uses a combination of phage display and SPOT synthesis to reveal all the peptides in a proteome that have the potential to bind to any domain of interest. The first step consists of the identification of the preferred ligand of the domain of interest by panning phage displayed peptide repertoires. Based on the 'strict' consensus identified by comparing the sequences displayed on the selected phage, a relaxed consensus is subsequently designed (Figure 4b). The stringency of the relaxed consensus is such that a search in the proteome of interest identifies between 5000 and 10,000 peptides. The peptides matching the relaxed consensus are then synthesized at high density on a cellulose membrane by SPOT synthesis and probed with the domain of interest. Using this approach peptides binding to each domain can be readily identified and protein partners inferred.

This strategy presents several useful features:

- It combines the complementary strengths of phage display and SPOT synthesis.
- It takes full advantage both of the genomic information that is being accumulated and of the array format in which all the possible targets are equally represented.
- It provides a high level of detail on the interaction topology.
- It is not affected by the variability of protein concentrations inside the cell and is very sensitive (interactions up to 100 μM can be detected).
- Interactions that depend on peptide modifications (e.g. phosphorylation) can also be studied.
- The output is semi-quantitative.

By applying this approach to the SH3 domain family, we were able to show that the inference based on phage display, when compared with the results of the SPOT approach, can reach an accuracy (percentage of correct prediction) of up to ~50% at a coverage (percentage of interactions predicted) of 40%.

Conclusions

As discussed in detail in a recent review published in this journal, biological networks are essential instruments for target discovery and validation and for designing strategies

to interfere rationally with cell physiology [39]. Most drugs aim at disrupting specific edges in graphs like the one shown in Figure 1. However, the functional interpretation of networks obtained by downloading interaction data deposited in public databases [40–43] is limited, because essential informations is often missing. To model interactions that are relevant for physiological pathways we would need to know about partner co-expression and co-localization and about topology and stoichiometry of the protein complexes.

We have pointed out here how the combination of phage display and peptide SPOT synthesis can contribute to the production of high-throughput semi-quantitative data that relieves one of these limitations by modeling explicitly, in the interaction network, the domain organization of proteins. Network nodes are therefore represented as a covalent assembly of modular domains, while edges link domains rather than proteins to their targets in partner proteins. This representation is richer in information and can affect our interpretation of signal propagation along the network.

The recent development of strategies to scan an entire proteome in search of binding peptides for SH3 and WW domains [38,44] is likely to produce a wealth of binding data over the next few years. WISE and related approaches can be easily extended to a variety of protein interaction domains, including those binding to modified peptides, thereby offering a new powerful proteomic tool to help complete a full description of the cell interactome. This, combined with results from global analyses of protein expression and localization [8,9], should soon put us in the position to have a functional perspective of protein networks.

Acknowledgements

We thank the Italian Association for Cancer Research.

References

1 Ito, T. *et al.* (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. U. S. A.* 98, 4569–4574
2 Uetz, P. *et al.* (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* 403, 623–627
3 Gavin, A.C. *et al.* (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415, 141–147
4 Ho, Y. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415, 180–183
5 Han, J.D. *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 430, 88–93
6 von Mering, C. *et al.* (2002) Comparative assessment of large-scale data sets of protein-protein interactions. *Nature* 417, 399–403
7 Kumar, A. *et al.* (2002) Subcellular localization of the yeast proteome. *Genes Dev.* 16, 707–719
8 Huh, W.K. *et al.* (2003) Global analysis of protein localization in budding yeast. *Nature* 425, 686–691
9 Ghaemmaghami, S. *et al.* (2003) Global analysis of protein expression in yeast. *Nature* 425, 737–741
10 Simpson, J.C. *et al.* (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* 1, 287–292
11 Karaoz, U. *et al.* (2004) Whole-genome annotation by using evidence integration in functional-linkage networks. *Proc. Natl. Acad. Sci. U. S. A.* 101, 2888–2893
12 von Mering, C. *et al.* (2003) STRING: a database of predicted functional associations between proteins. *Nucleic Acids Res.* 31, 258–261
13 Yook, S.H. *et al.* (2004) Functional and topological characterization of protein interaction networks. *Proteomics* 4, 928–942
14 Pawson, T. and Nash, P. (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300, 445–452
15 Bork, P. *et al.* (1997) Cytoplasmic signalling domains: the next generation. *Trends Biochem. Sci.* 22, 296–298
16 Cesareni, G. *et al.*, eds (2004) *Modular Protein Domains*. Wiley-VCH Verlag GmbH & Co. KGaA
17 Sadowski, I. *et al.* (1986) A noncatalytic domain conserved among cytoplasmic protein-tyrosine kinases modifies the kinase function and transforming activity of Fujinami sarcoma virus P130gag-fps. *Mol. Cell. Biol.* 6, 4396–4408
18 Ren, R. *et al.* (1993) Identification of a ten-amino acid proline-rich SH3 binding site. *Science* 259, 1157–1161
19 Cicchetti, P. *et al.* (1992) Identification of a protein that binds to the SH3 region of Abl and is similar to Bcr and GAP-rho. *Science* 257, 803–806
20 Mayer, B.J. *et al.* (1991) The noncatalytic src homology region 2 segment of abl tyrosine kinase binds to tyrosine-phosphorylated cellular proteins with high affinity. *Proc. Natl. Acad. Sci. U. S. A.* 88, 627–631
21 Cesareni, G. *et al.* (2002) Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett.* 513, 38–44
22 Pawson, T. (2004) Specificity in signal transduction: from phosphotyrosine-SH2 domain interactions to complex cellular systems. *Cell* 116, 191–203
23 Castagnoli, L. *et al.* (2004) Selectivity and promiscuity in the interaction network mediated by protein recognition modules. *FEBS Lett.* 567, 74–79
24 Songyang, Z. *et al.* (1994) Use of an oriented peptide library to determine the optimal substrates of protein kinases. *Curr. Biol.* 4, 973–982
25 Songyang, Z. and Cantley, L.C. (1995) SH2 domain specificity determination using oriented phosphopeptide library. *Methods Enzymol.* 254, 523–535
26 Yaffe, M.B. and Cantley, L.C. (2000) Mapping specificity determinants for protein-protein association using protein fusions and random peptide libraries. *Methods Enzymol.* 328, 157–170
27 Frank, R. (1992) Spot-synthesis: an easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* 48, 9217–9232
28 Scott, J.K. and Smith, G.P. (1990) Searching for peptide ligands with an epitope library. *Science* 249, 386–390
29 Sparks, A.B. *et al.* (1994) Identification and characterization of Src SH3 ligands from phage-displayed random peptide libraries. *J. Biol. Chem.* 269, 23853–23856
30 Rickles, R.J. *et al.* (1994) Identification of Src, Fyn, Lyn, PI3K and Abl SH3 domain ligands using phage display libraries. *EMBO J.* 13, 5598–5604
31 Vaccaro, P. *et al.* (2001) Distinct binding specificity of the multiple PDZ domains of INADL, a human protein with homology to INAD from *Drosophila melanogaster*. *J. Biol. Chem.* 276, 42122–42130
32 Mongiovì, A.M. *et al.* (1999) A novel peptide-SH3 interaction. *EMBO J.* 18, 5300–5309
33 Cestra, G. *et al.* (1999) The SH3 domains of endophilin and amphiphysin bind to the proline-rich region of synaptojanin 1 at distinct sites that display an unconventional binding specificity. *J. Biol. Chem.* 274, 32001–32007
34 Paoluzi, S. *et al.* (1998) Recognition specificity of individual EH domains of mammals and yeast. *EMBO J.* 17, 6541–6550
35 Dente, L. *et al.* (1997) Modified phage peptide libraries as a tool to study specificity of phosphorylation and recognition of tyrosine containing peptides. *J. Mol. Biol.* 269, 694–703
36 Freund, C. *et al.* (2003) Structural investigations of a GYF domain covalently linked to a proline-rich peptide. *J. Biomol. NMR* 27, 143–149
37 Tong, A.H. *et al.* (2002) A combined experimental and computational strategy to define protein interaction networks for peptide recognition modules. *Science* 295, 321–324
38 Landgraf, C. *et al.* (2004) Protein interaction networks by proteome peptide scanning. *PLoS Biol.* 2, E14
39 Nikolsky, Y. *et al.* (2005) Biological networks and analysis of experimental data in drug discovery. *Drug Discov. Today* 10, 653–662
40 Zanzoni, A. *et al.* (2002) MINT: a Molecular INTeraction database. *FEBS Lett.* 513, 135–140
41 Hermjakob, H. *et al.* (2004) IntAct: an open source molecular interaction database. *Nucleic Acids Res.* 32, D452–D455
42 Bader, G.D. *et al.* (2003) BIND: the Biomolecular Interaction Network Database. *Nucleic Acids Res.* 31, 248–250
43 Xenarios, I. *et al.* (2002) DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. *Nucleic Acids Res.* 30, 303–305
44 Hu, H. *et al.* (2004) A map of WW domain family interactions. *Proteomics* 4, 643–655
45 Hu, Z. *et al.* (2004) VisANT: an online visualization and analysis tool for biological interaction data. *BMC Bioinformatics* 5, 17