PROTEINS: Structure, Function, and Bioinformatics 54:49–57 (2004)

Detection of Functional Modules From Protein Interaction Networks

Jose B. Pereira-Leal,[1] Anton J. Enright,[2] and Christos A. Ouzounis[1]*
[1]*Computational Genomics Group, The European Bioinformatics Institute, Cambridge, United Kingdom*
[2]*Computational Biology Center, Memorial Sloan-Kettering Cancer Center, New York, New York*

ABSTRACT **Complex cellular processes are modular and are accomplished by the concerted action of functional modules (Ravasz et al., Science 2002;297:1551–1555; Hartwell et al., Nature 1999;402: C47–52). These modules encompass groups of genes or proteins involved in common elementary biological functions. One important and largely unsolved goal of functional genomics is the identification of functional modules from genomewide information, such as transcription profiles or protein interactions. To cope with the ever-increasing volume and complexity of protein interaction data (Bader et al., Nucleic Acids Res 2001;29:242–245; Xenarios et al., Nucleic Acids Res 2002;30:303–305), new automated approaches for pattern discovery in these densely connected interaction networks are required (Ravasz et al., Science 2002;297:1551–1555; Bader and Hogue, Nat Biotechnol 2002;20:991–997; Snel et al., Proc Natl Acad Sci USA 2002;99:5890–5895). In this study, we successfully isolate 1046 functional modules from the known protein interaction network of *Saccharomyces cerevisiae* involving 8046 individual pair-wise interactions by using an entirely automated and unsupervised graph clustering algorithm. This systems biology approach is able to detect many well-known protein complexes or biological processes, without reference to any additional information. We use an extensive statistical validation procedure to establish the biological significance of the detected modules and explore this complex, hierarchical network of modular interactions from which pathways can be inferred. Proteins 2004;54:49–57.** © 2003 Wiley-Liss, Inc.

Key words: functional modules; protein interactions; genomics; yeast; bioinformatics

INTRODUCTION

Cluster analysis is an obvious choice of methodology for the extraction of functional modules from protein interaction networks. Clustering can be defined as the grouping of objects based on their sharing of discrete, measurable properties. A variety of clustering algorithms have been developed and successfully used in diverse fields. In functional genomics, clustering algorithms have been devised for multiple tasks, such as mRNA expression analysis and the detection of protein families.[1,2]

Our aim is to detect biologically meaningful patterns in the entire known protein interaction network of *Saccharomyces cerevisiae*. This species was chosen because it is a well-studied model organism for which large quantities of interaction data are already available.[3] It was previously noted that the most reliable protein interaction assignments are those supported by more than one method.[4,5] Accordingly, we have devised a simple additive weighting scheme for our data set that rewards repeated observations of interactions within and (in particular) between experimental methodologies and provides us with a reliable confidence measure (see Materials and Methods).

RESULTS AND DISCUSSION

Because our aim is to isolate functionally coordinated interactions (modules), after computing weights for each interaction, we express the network of proteins (nodes) connected by interactions (edges) as a network of connected interactions [Fig. 1(a) and (b)]. In more general terms, this procedure takes a graph ***G***, consisting of edges connecting nodes, and produces its associated line graph ***L(G)*** in which edges now represent nodes and nodes represent edges.[6] This simple procedure is commonly used in graph theory,[7] and the line graph generated has a number of advantages for graph clustering: 1) it does not sacrifice information content because the original bidirectional network can be recovered, 2) it takes into account the higher-order local neighborhood of interactions, and 3) hence, it is more highly structured than the original graph [Fig. 1(c) and (d)].

This increase in local structure is illustrated by a fivefold increase in the overall graph-clustering coefficient after transformation of the original graph into the corresponding line graph (from 0.06 to 0.33). The clustering coefficient is a parameter that indicates, for each node, the amount of adjacent nodes that are also adjacent to each other.[8] Subsequently, in this line graph of interactions, each node represents an interaction between two proteins, and each edge represents pairs of interactions connected by a common protein. The increase in local structure in

Grant sponsor: European Molecular Biology Laboratory; Grant sponsor: Fundação para a Ciência e Tecnologia-Portugal; Grant sponsor: UK Medical Research Council; Grant sponsor: IBM Research.

J.B.P.-L. and A.J.E. contributed equally to this work.

*Correspondence to: Christos A. Ouzounis, Computational Genomics Group, The European Bioinformatics Institute, Cambridge CB10 1SD, UK. E-mail: ouzounis@ebi.ac.uk

Received 28 January 2003; Accepted 14 May 2003

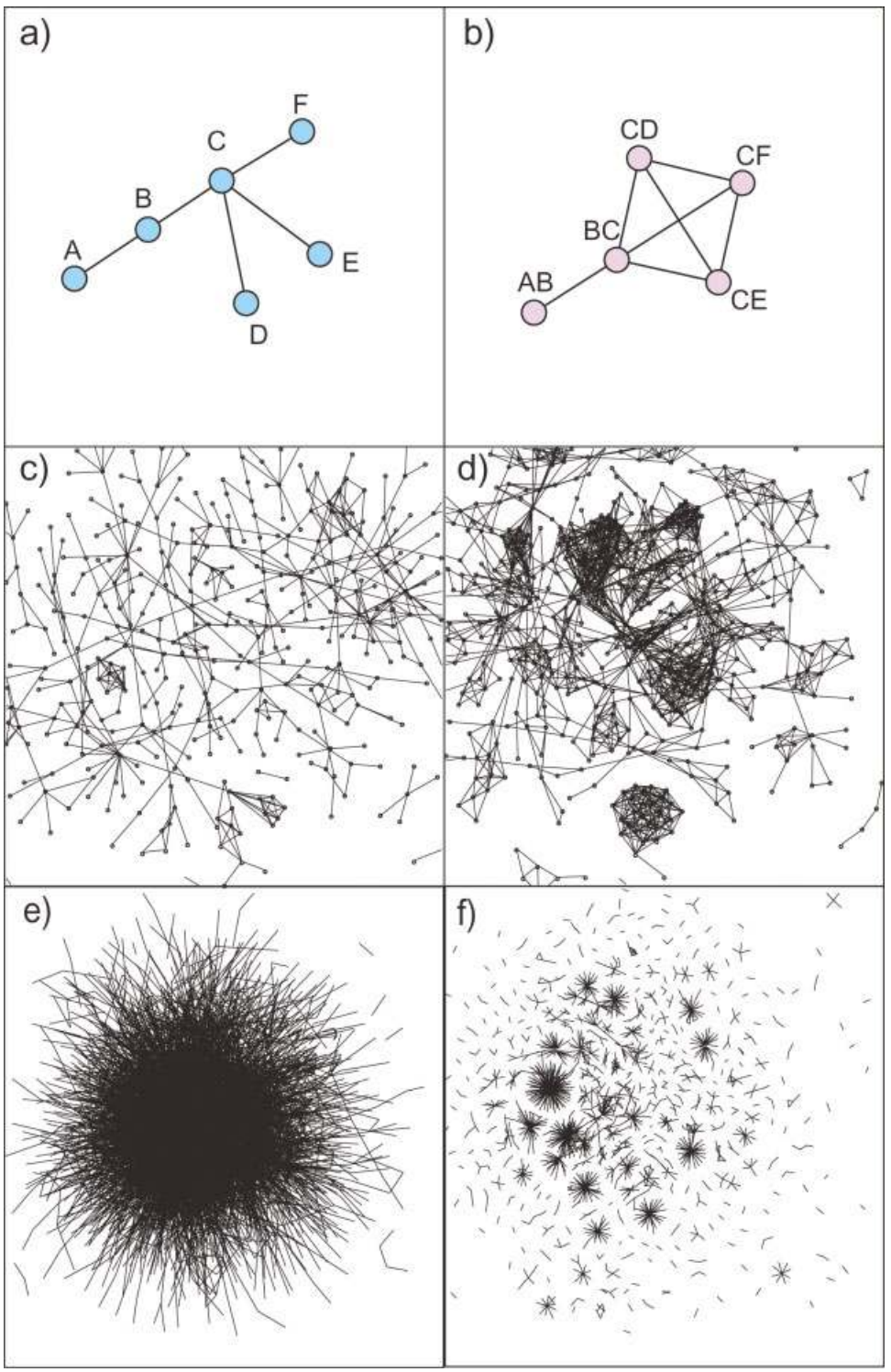

Fig. 1. Transforming a network of proteins to a network of interactions. **a:** Schematic representation illustrating a graph representation of protein interactions: nodes correspond to proteins and edges to interactions. **b:** Schematic representation illustrating the transformation of the protein graph connected by interactions to an interaction graph connected by proteins. Each node represents a binary interaction and edges represent shared proteins. Note that labels that are not shared correspond to terminal nodes in (a)—in this particular case, A, D, E, and F in edges AB, CD, CE, CF. **c:** Graph illustrating a section of a protein network connected by interactions. **d:** Graph illustrating the increase in structure as an effect of transforming the protein graph in (c) to an interaction graph. **e:** Graph representation of Yeast protein interactions in DIP. **f:** Graph representing a pruned version of (e) with the reconstituted interactions after transformation and clustering, as described in Materials and Methods. These graphs were produced by using BioLayout.[22]

this higher-order representation generates a graph that is better suited to clustering algorithms such as TribeMCL,[2] based on graph clustering by flow simulation.[9] Clustering the interaction network with use of this method yields a total of 1046 clusters [Fig. 1(e) and (f)]. These clusters range in size from 2 to 292 components (average size is 8.05) and form a scale-free protein network, in accordance with the expected topology of protein interaction networks.[10–12]

Another distinct advantage of this approach is that it produces an overlapping graph partitioning of the interaction network. This implies that proteins may be present in multiple functional modules. Many clustering approaches cannot place elements in multiple clusters, which can be unrealistic for biological systems, where proteins may participate in multiple cellular processes and pathways. In this analysis, each protein was on average present in 2.1 clusters. Furthermore, because clusters may be overlapping (i.e., sharing proteins), very few interactions are discarded by the method. In total, only 76 interactions involving 146 proteins, which were weakly connected to the main interaction network, were discarded by the clustering method. These 76 interactions represent <1% of the interactions in the initial interaction data, indicating that the use of the protocol described here does not result in a significant loss of data.

To test our hypothesis that detected clusters represent functional modules, containing proteins that are involved in the same coordinated functional process, we make use of a variety of external biological information. Our rationale is that proteins in valid functional modules will possess consistent functional annotations indicating their common involvement in the same biological process, whereas an incorrect cluster will contain many proteins with disparate or conflicting annotations indicating random assignment to that module.

For each protein in a cluster, we obtain manually derived regulatory and metabolic classifications (*KEGG*), automatic functional classifications (*GQFC*), and also cellular localization information (*LOC*) (from KEGG, GeneQuiz, and MIPS, respectively; see Materials and Methods). On average, the coverage of clusters is 20% for regulatory and metabolic roles in KEGG, 45% for functional classes in GeneQuiz, and 48% for cell localization in MIPS. It is of interest that the coverage does not appear to be dependent on the cluster size (see http://www.ebi.ac.uk/research/cgg/proteinets/).

We then analyze the homogeneity of these annotations for each cluster and each classification scheme. These independent and distinct classifications guarantee that cluster validation is not biased by any single classification scheme. To test the significance of the overall clustering result and not just of individual clusters, we obtain the sum of all scores for all clusters and compare this score with the distribution of equivalent summed scores from a set of 1000 randomized clustered networks with the same topology (Fig. 2). Keeping the same cluster size distribution ensures that we do not introduce any score bias resulting from cluster size effects. The summed scores, for each of the three classification schemes, are significantly larger than score sums obtained from randomized clusterings (Fig. 2). This result is highly significant because for each classification scheme, not one of the 1000 random expected scores was higher than the observed result (p values of $P_{KEGG} < 0.001$, $P_{GQFC} < 0.001$, and $P_{Loc} < 0.001$; z score values of $Z_{KEGG} = 60.9$; $Z_{GQFC} = 30.5$; $Z_{LOC} = 39.5$). This result clearly indicates that the clustering obtained is distinctly nonrandom and that our algorithm detects biologically meaningful modules from the protein interaction network (Fig. 2).

To further substantiate that detected clusters correspond to functional modules (i.e., their members participate in common biological processes), we ask whether detected clusters are also enriched in pairs of proteins shown previously to be functionally associated, using genetic interaction information (see Materials and Methods). We observe a statistically significant ($p < 0.001$) selection for pairs of genetic interactors in the same clusters and in directly associated clusters compared with random clusters (see http://www.ebi.ac.uk/research/cgg/proteinets/) (see Materials and Methods). This result strongly suggests that detected clusters are specifically enriched in proteins involved in the same biological process and further substantiates that they correspond to valid functional modules.

Finally, we observe that many detected clusters represent previously characterized functional modules. One example is a high-scoring cluster (cluster 46; $R_{Kegg} = 1.00$, $R_{MIPS} = 1.00$, $R_{GQFC} = 0.76$) containing the mitochondrial F1-F0 ATP synthase complex (EC 3.6.1.34).[13,14] In this cluster, all known protein interactions related to the formation of this complex are recovered. In addition, two chaperones (Atp11p, Atp12p), which are essential for F1 complex formation, are also recovered, as are four other proteins, Inh1p and three proteins of unknown function—YBR271W, YDR322B, and YML081B. The Inh1p protein provides connections from this module to other modules involved in the cytoskeleton, budding, and cell division (e.g., clusters 69, 77, 147, and 664). Inh1p is known to be an endogenous inhibitor of the F1F0 ATP synthase, and this role has been well documented.[14] This complex of 17 proteins forming 36 distinct interactions is extracted from an interaction neighborhood of 149 proteins forming 170 interactions (considering only the first neighbors of the complex members). The recovery of this vital functional module from a large, complex, and highly connected network is nontrivial and further illustrates the power of this approach.

Similarly, many other well-known protein complexes are found in the detected clusters [e.g., the prereplicative complex (cluster 99)[15] or the oligosaccharyl transferase complex (cluster 140)[16]] (see Figs. 3 and 4 for more examples). To determine to what extent known complexes were recovered as functional modules, we made use of the manually curated MIPS definitions of protein complexes.[17] We observe that we recover all complexes containing at least three proteins and that, on average, these complexes are recovered with 69% recall. Of these, 30% are recovered

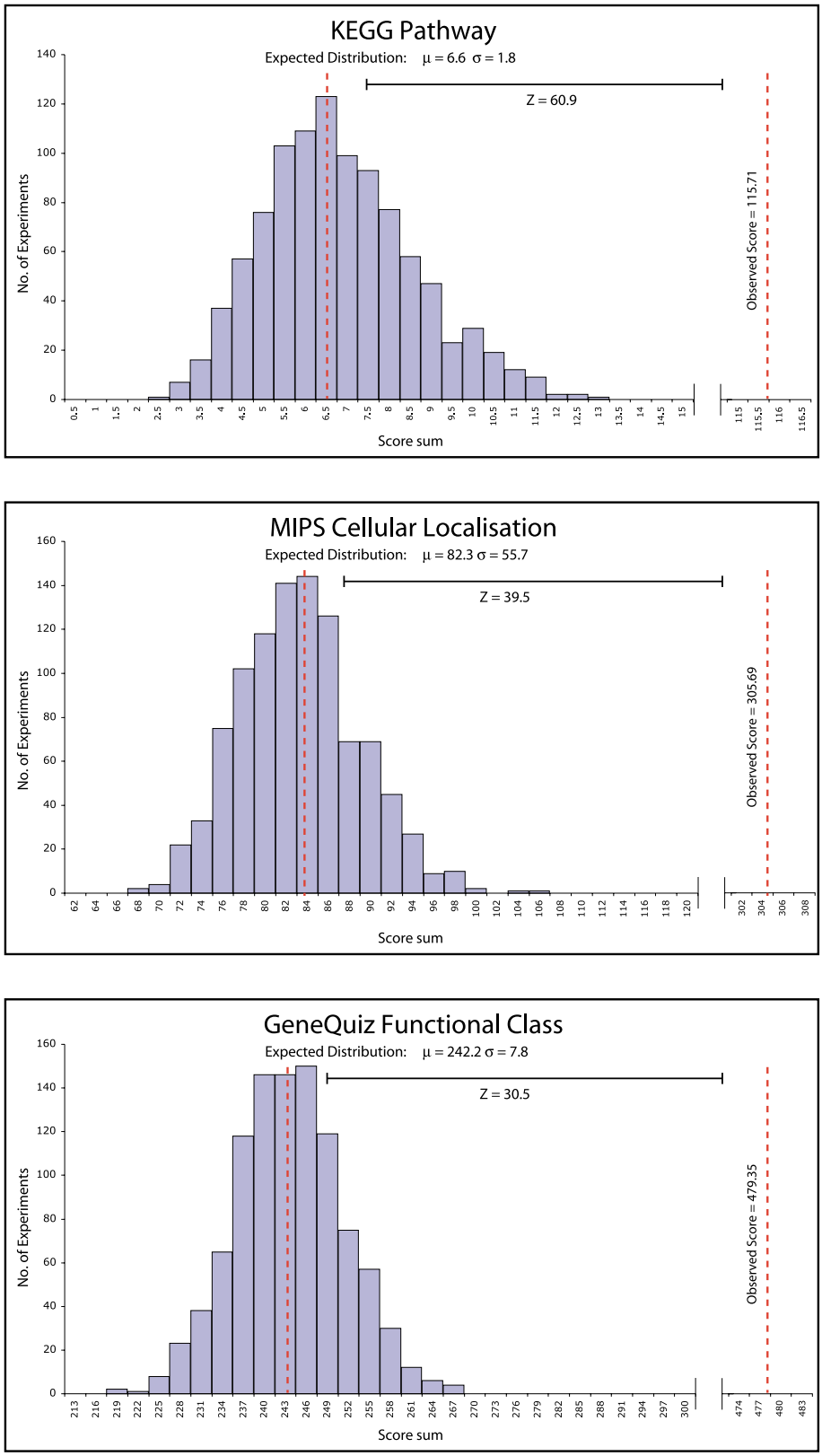
KEGG Pathway
Expected Distribution: μ = 6.6 σ = 1.8
Z = 60.9
Observed Score = 115.71
No. of Experiments
Score sum
MIPS Cellular Localisation
Expected Distribution: μ = 82.3 σ = 55.7
Z = 39.5
Observed Score = 305.69
No. of Experiments
Score sum
GeneQuiz Functional Class
Expected Distribution: μ = 242.2 σ = 7.8
Z = 30.5
Observed Score = 479.35
No. of Experiments
Score sum

Figure 2.

with 100% recall, and only ~30% of the complexes are recovered at recall values <50% (not shown). The detection of curated protein complexes strongly suggests that many high-scoring modules, which may not have previously been characterized, are valid and represent interesting and experimentally testable functional associations.

Our results clearly show that detected clusters 1) contain proteins with consistent functional classifications and cellular localizations, 2) are enriched in proteins of known functional connectedness, and 3) include many examples of previously described functional modules. These results strongly suggest that the detected clusters correspond to biologically meaningful functional modules. The complementarity between graph weighting, graph transformation, and graph flow simulation clustering appears to be the key factor to the successful isolation of biologically significant modules. Each of the three steps contributes heavily toward the final result. If any of these three stages are omitted, the final clustering result is poor, because the final clustering scores are lower than those obtained with the complete protocol. For example, graph flow clustering of the raw interaction data—without its associated line-graph—results in the detection of clusters whose scores were decreased by a factor of 2.4, 1.3, and 2.5, respectively for KEGG, GQFC, and LOC.

Analysis of the detected functional modules allows the exploration of both the functional repertoire of proteins and the biological processes in which they participate. Furthermore, the detection of a module containing well-characterized proteins along with proteins of unknown function, represents a functional prediction for the unknown proteins. An illustrated example of this type of analysis for proteins contained in a single module (cluster 55) and involved in vacuolar transport and fusion is shown in Figure 3. Another key feature of our clustering procedure is that proteins can be present in more than one functional module. Many of the detected modules contain proteins that are present in other modules; hence, two modules containing a common protein can be linked. This one-to-many relationship can be exploited for pathway discovery. Illustrating this type of application is the recovery of a signal transduction pathway controlling cell wall biogenesis, from the membrane protein (Fks1p) to the transcription factors activated by this pathway (Swi4p, 6p, and Rlm1p) (Fig. 4). This pathway was recovered as a set of two clusters, linked by two proteins (Pkc1p and Smd3p) (Fig. 4).

More generally, the automatic assignment of proteins to multiple clusters by the present approach allows us to gain an insight into how elementary biological units interact to form complex cellular networks. The graph representing interactions between functional modules is highly connected, illustrating that few biological processes are isolated units, but form a complex web of functional interactions within the cell. A subsection of this complex graph is shown in Figure 5. This graph shows the connections between 40 functional modules manually labeled according to their basic functional categories. For instance, it illustrates how several structural modules are linked, via signaling modules, to cell cycle regulatory modules and to protein-trafficking modules. In the same figure, one can see how the RNA polymerase module connects to transcriptional regulatory modules. Thus, analysis of protein interaction data can produce testable predictions of interconnectedness between functional modules and the high-level organization of these modules into the entire network of cellular processes.

In conclusion, using a novel, fully automated and unsupervised discovery protocol, involving transformation and clustering of networks, we are able to identify biologically relevant functional modules from protein interaction data. The method has a predictive aspect, because modules can subsequently be used to place poorly characterized proteins into their functional context according to their interacting partners within a module. This predictive power extends to a higher level, because connections between functional modules can be used to examine the organization and coordination of multiple complex cellular processes and determine how they are organized into pathways. Given the ever-increasing amount of interaction data available, we expect that the approach described herein will prove useful in the ongoing efforts to explore the protein interaction universe and understand how functional building blocks are assembled into an entire living system.

Fig. 2. Module validation using biological classification schemes. Distribution of score sums of 1000 sets of randomized clusters compared to the sum averages of the transformed and clustered network, for each of the three classification schemes, namely (**a**) pathway membership, (**b**) cellular localization, and (**c**) functional class assignment (see Materials and Methods). Each randomized cluster set is generated by taking the clusters obtained from the clustering of the line graph of protein interactions and randomly shuffling protein identifiers between clusters. This ensures that the cluster size distribution of random cluster sets remains constant (and hence the number of proteins), but that cluster membership of proteins is random. For each randomized set, we compute the score sum for each classification scheme as described for the line graph analysis. All three panels are shown on the same scale (*y* axis); however, the absolute values obtained (*x* axis) are highly dependent on the number of classes being assigned (KEGG = 110; MIPS = 15; GQFC = 17). Classification schemes with larger numbers of classes typically produce lower scores according to our scoring scheme. Conversely, lower coverage results in lower scores. For this reason, *z* scores are used to assess the significance for each classification scheme, being comparable across schemes.

MATERIALS AND METHODS

Protein Interaction Data

The protein interaction network is derived from the yeast subset of the Database of Interacting Proteins (DIP) (Jan 2002 release).[18] The DIP database is appropriate because it contains curated interactions from both small- and large-scale experimental studies. This data set consists of 8046 physical interactions involving 4081 yeast proteins. Most of these interactions have been derived by yeast two-hybrid screen. Of these interactions, 6586 are supported by a single observation and all other interactions are supported by at least two experimental observations.

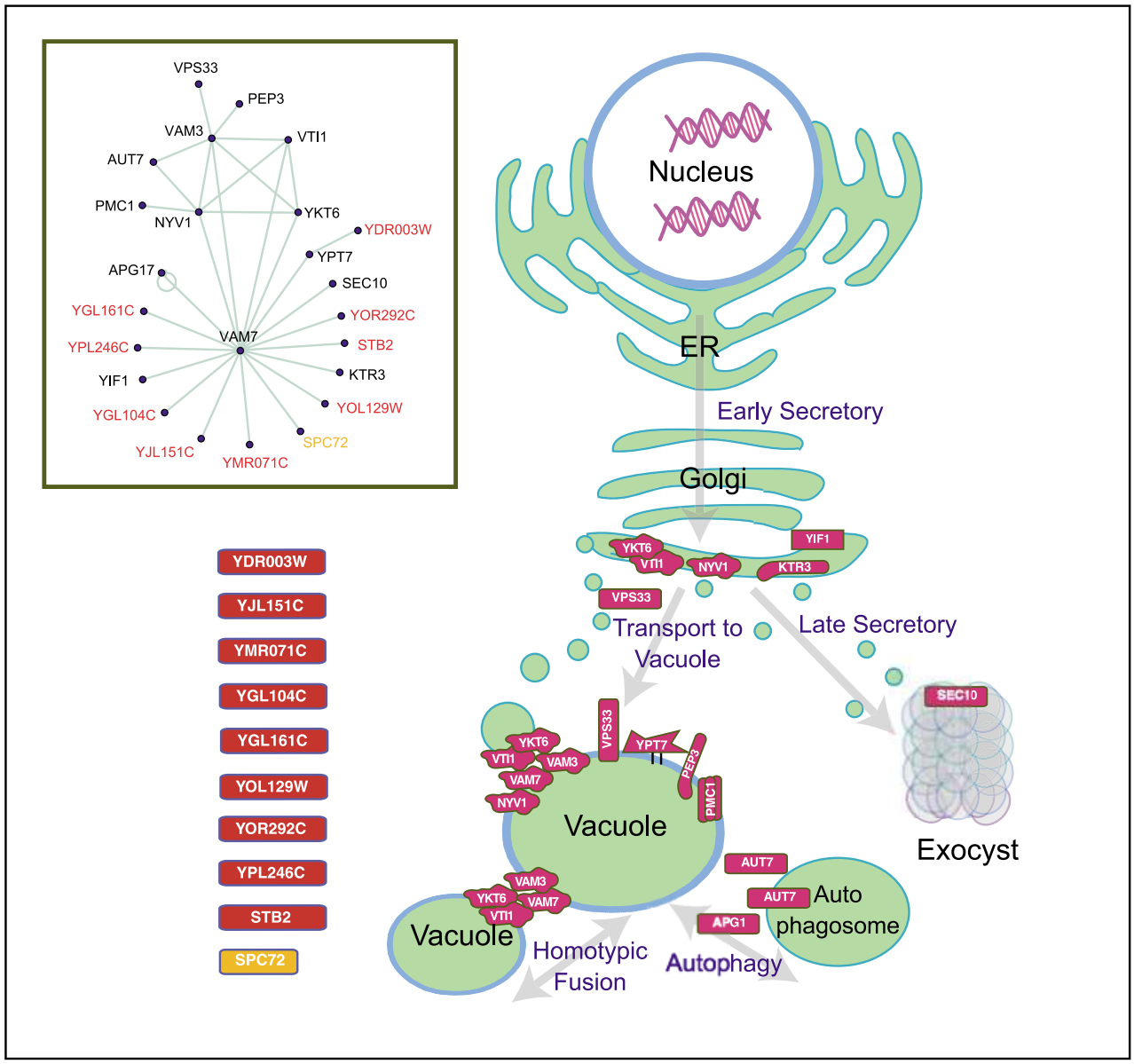

Fig. 3. Analysis of a functional module. Example of a functional module and its use for functional assignment. Here, cluster 55 recovers a set of protein interactions (**inset**) that are involved in the transport route from the endoplasmic reticulum to the vacuole via the prevacuolar compartment, as well as in homotypic vacuolar fusion[23] and vacuole-autophagosome fusion.[24] Ykt6p, Vti1p, Vam3p, Vam7p, and Nyv1p are SNARE proteins that have been shown to be involved in tethering and docking of membranes in different stages of these transport routes.[24] Ypt7p is vacuolar-specific "switch" (small GTPase from the Rab family), Pep3 is a vacuolar docking factor, and Ktr3p is an enzyme involved in protein glycosylation (mannosyltransferase).[23] Many proteins of unknown function (in red), associated directly or indirectly with Vam7p, can thus be predicted to localize to one of the compartments in these pathways, most likely the vacuole, and possibly be involved in homotypic vacuolar fusion, vacuole-autophagosome fusion, or transport from the Golgi to the vacuole. It is of interest that two of these proteins (YGL104C/Vps73p and YOL129W/Vps68p) have been found to be involved in vacuolar protein sorting in a large genomic screen for carboxypeptidase secretion mutants,[25] thus supporting the functional assignments made here. Furthermore, novel aspects of this pathway can be inferred from the analysis of the interactions that are recovered in this module. Vam7p, a vacuolar SNARE, interacts with Yif1p and Ktr3p. Yif1p is a Golgi-associated protein that interacts with an effector of the early secretory pathway switch Ypt1p and to Ktr3p, which is expected to be present in earlier stages of this pathway. This points to a previously undescribed role for Vam7p in earlier stages of this transport route. In addition, the inclusion of Sec10p, a component of the exocysts, suggests a connection between vacuolar transport and secretion.

Interaction Network Clustering

We first transform the protein interaction network into a weighted network, where the weights attributed to each interaction reflect the degree of confidence attributed to that interaction—where the confidence level represents the number of experiments that support the interactions as well as the number of different experimental methodologies. All experimental methodologies are considered equivalent (no subjective judgment is made on the reliability of each methodology and given the same absolute score, e.g.,

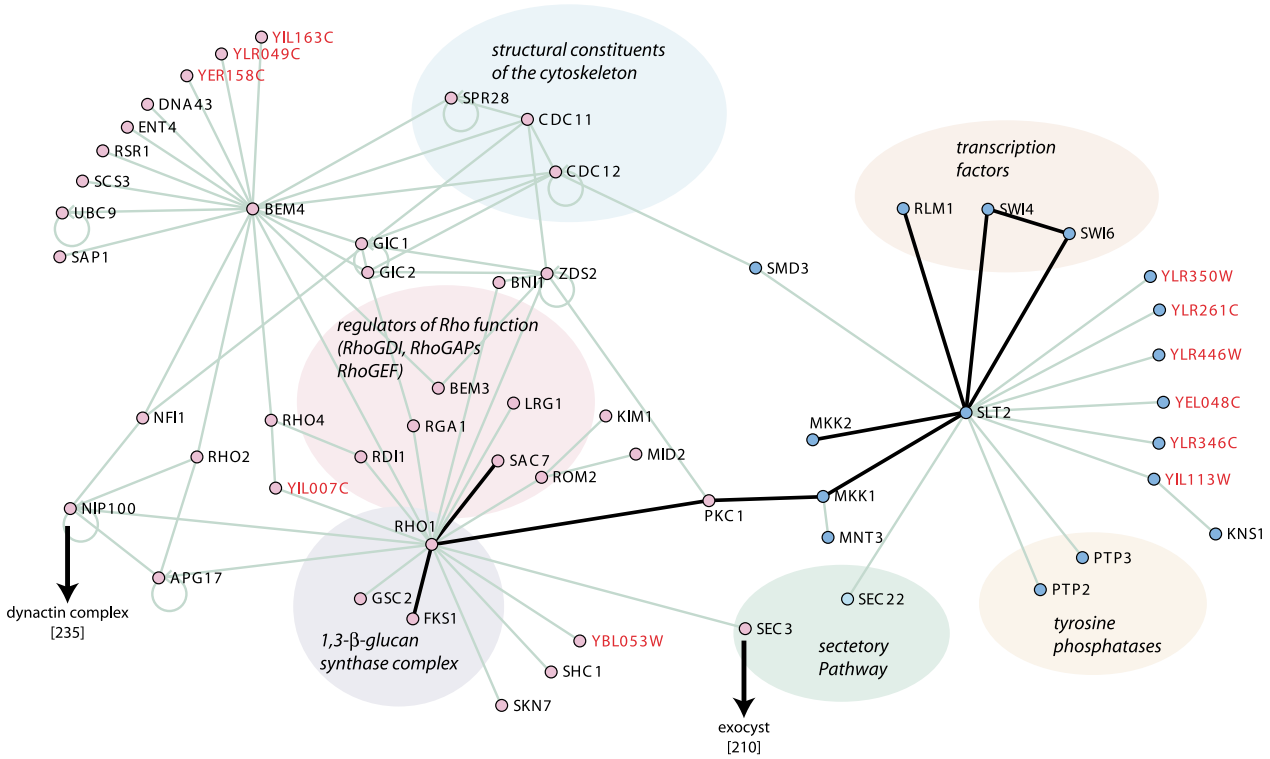

Fig. 4. Pathway discovery from connected modules. Protein interactions in clusters 32 and 86 (pink and blue nodes, respectively). The connections of components between these clusters permit the reconstruction of a signal transduction pathway central in the regulation of cell wall biogenesis in yeast (highlighted in black edges).[26,27] Shaded areas represent common functions in the clusters. Two proteins appear as central signal integration points in this pathway. Rho1p is a member of a cluster that includes the 1,3-β-glucan synthase complex and other proteins, essential for cell wall biogenesis.[26,27] In addition, regulators of Rho1p activity are included (e.g., Sac7p or Rom2p) as well as proteins that connect this cluster to other protein complexes (indicated by arrows connecting to the cluster number in square brackets and name). A second major signal integration point in this pathway is Slt2p, a MAP kinase that activates the transcription factors that are the "end point" of the pathway and connects to known regulators of kinase function (Ptp2p and Ptp3p protein phosphatases).

small- and large-scale experiments receive the same score). On the first instance of an interaction in the data set, we assign it an initial score of 3.0. For further instances of the interaction, the score is increased by 1.0 if the method is different from the other methods used to determine the interaction, or 0.25 if the interaction had already been observed by that method. In the end, all scores are normalized as a percentage of the highest score in the data set. These data represent a weighted network of proteins connected by interactions, where the weights qualitatively reflect the confidence we attribute to each interaction based on the amount of experimental evidence supporting it.

Next we express the network of proteins connected by interactions as a network of interactions; this is known in graph theory as a line graph. Each binary interaction is condensed into a node that includes the two interacting proteins. These nodes are then linked by shared protein content [Fig. 1(a) and (b)]. The scores for the original constituent binary interactions are then averaged and assigned to each edge. We then use TribeMCL,[2] an algorithm for clustering by graph flow simulation, at inflation value 3.0 to cluster the interaction network and recover clusters of associated interactions. These clusters of interactions are then transformed back from an interaction–interaction graph to a protein–protein graph for all subsequent validation and analysis. It is important to stress that this procedure is not specifically selecting for high-confidence interactions, nor is it biased to any experimental methodology.

Biological Significance of the Detected Modules

Clusters are validated by assessing the consistency of protein classifications within an individual cluster. We assign a score to each cluster reflecting the homogeneity of classifications within that cluster. This is measured, for each of the three classification schemes, by calculating the redundancy (R_j) of each cluster j

$$Rj = 1 - \left(\frac{-\left(\sum_{s=1}^{n} p_s \log_2 p_s \right)}{\log_2 n} \right)$$

In this equation, n represents the number of classes in the classification scheme, p_s represents the relative frequency of the class in cluster j, the numerator represents the information content in bits given by Shannon's entropy (H), and the denominator is a normalizing factor representing the maximum entropy for the cluster j ($H_{\max}$). All

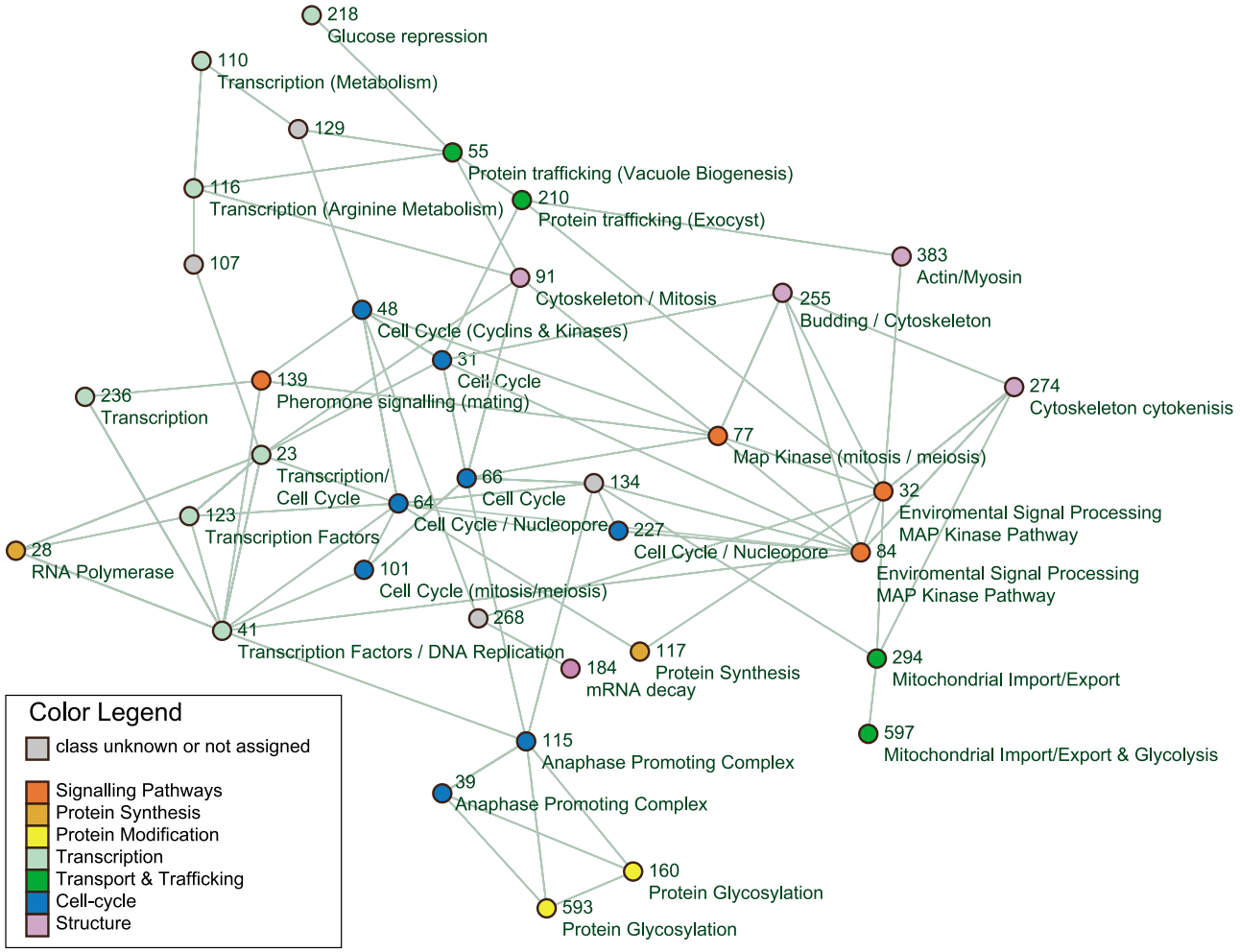

Fig. 5. Network of functional modules. Section of the network of functional modules connected by shared proteins. The nodes correspond to clusters of $R \geq 0.5$ and are colored according to general functional classes (see inset legend). Modules are assigned to functional classes by manual curation via analysis of the functions of their constituent proteins.

values of R lie between 0 and 1. With this scoring system, clusters containing many proteins with highly consistent classifications will receive high scores (R closer to 1), whereas those with disparate or conflicting classifications will receive low scores (R closer to 0).

Classification schemes include regulatory and metabolic pathway membership, functional classification, and cellular localization. Regulatory and metabolic information for each gene was obtained from KEGG[19] and automatic functional classification from GeneQuiz.[20] Information on cellular localization was obtained from MIPS/CYGD.[17] These three distinct classification schemes were chosen because they represent different types of classification: automatic versus manual, high granularity versus low granularity, sequence based on nonsequence based.

To test if detected clusters were enriched in functionally associated proteins, we obtained information on genetic interactions from MIPS/CYGD.[17] These genetic interactions were obtained from manual curation of the literature and from one large-scale study.[21] We then compare the number of pairs of genetic interactors that occur in the same cluster in the clustered network, with the same 1000 random networks (as described in Fig. 2).

Protein complexes were obtained from MIPS/CYGD,[17] representing 265 manually defined protein complexes. We define the protein complex size as the number of proteins in the complex that are available. We chose to use solely complexes with at least three proteins to avoid artificially increasing the recall values. We test for each complex if it corresponds to any of the clusters detected and which proportion of the complex is contained within a single cluster (recall).

ACKNOWLEDGMENTS

We thank Alex Bateman (Sanger Institute), Ildefonso Cases (EMBL-EBI), Chris Sander, Debbie Marks, Daniel Eisenbud (MSKCC, New York), and Ray Paton (University of Liverpool) for useful discussions. J.P-L is funded by Fundação para a Ciência e Tecnologia-Portugal. C.A.O. acknowledges additional support from the UK Medical Research Council and IBM Research.

REFERENCES

1. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 1998;95:14863–14868.
2. Enright AJ, Van Dongen S, Ouzounis CA. An efficient algorithm for large-scale detection of protein families. Nucleic Acids Res 2002;30:1575–1584.
3. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P, Qureshi-Emili A, Li Y, Godwin B, Conover D, Kalbfleisch T, Vijayadamodar G, Yang M, Johnston M, Fields S, Rothberg JM. A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 2000;403:623–627.
4. Gerstein M, Lan N, Jansen R. Proteomics. Integrating interactomes. Science 2002;295:284–287.
5. von Mering C, Krause R, Snel B, Cornell M, Oliver SG, Fields S, Bork P. Comparative assessment of large-scale data sets of protein-protein interactions. Nature 2002;417:399–403.
6. Whitney H. Congruent graphs and the connectivity of graphs. Am J Math 1932;54:150–168.
7. Harary F. Graph theory. Addison-Wesley: Reading, MA, 1969.
8. Watts DJ, Strogatz SH. Collective dynamics of "small-world" networks. Nature 1998;393:440–442.
9. Van Dongen S. Graph clustering by flow simulation. PhD thesis. Center for Mathematics and Computer Science (CWI); 2000.
10. Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex networks. Nature 2000;406:378–382.
11. Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL. The large-scale organization of metabolic networks. Nature 2000;407: 651–654.
12. Jeong H, Mason SP, Barabasi AL, Oltvai ZN. Lethality and centrality in protein networks. Nature 2001;411:41–42.
13. Nakamoto RK, Ketchum CJ, al-Shawi MK. Rotational coupling in the F0F1 ATP synthase. Annu Rev Biophys Biomol Struct 1999;28: 205–234.
14. Devenish RJ, Prescott M, Roucou X, Nagley P. Insights into ATP synthase assembly and function through the molecular genetic manipulation of subunits of the yeast mitochondrial enzyme complex. Biochim Biophys Acta 2000;1458:428–442.
15. Aparicio OM, Weinstein DM, Bell SP. Components and dynamics of DNA replication complexes in *S. cerevisiae*: redistribution of MCM proteins and Cdc45p during S phase. Cell 1997;91:59–69.
16. Knauer R, Lehle L. The oligosaccharyltransferase complex from *Saccharomyces cerevisiae*. Isolation of the OST6 gene, its synthetic interaction with OST3, and analysis of the native complex. J Biol Chem 1999;274:17249–17256.
17. Mewes HW, Frishman D, Guldener U, Mannhaupt G, Mayer K, Mokrejs M, Morgenstern B, Munsterkotter M, Rudd S, Weil B. MIPS: a database for genomes and protein sequences. Nucleic Acids Res 2002;30:31–34.
18. Xenarios I, Salwinski L, Duan XJ, Higney P, Kim SM, Eisenberg D. DIP, the Database of Interacting Proteins: a research tool for studying cellular networks of protein interactions. Nucleic Acids Res 2002;30:303–305.
19. Kanehisa M, Goto S, Kawashima S, Nakaya A. The KEGG databases at GenomeNet. Nucleic Acids Res 2002;30:42–46.
20. Andrade MA, et al. Automated genome sequence analysis and annotation. Bioinformatics 1999;15:391–412. s.html.
21. Andrade MA, Brown NP, Leroy C, Hoersch S, de Daruvar A, Reich C, Franchini A, Tamames J, Valencia A, Ouzounis C, Sander C. Systematic genetic analysis with ordered arrays of yeast deletion mutants. Science 2001;294:2364–2368.
22. Enright AJ, Ouzounis CA. BioLayout—an automatic graph layout algorithm for similarity visualization. Bioinformatics 2001;17:853–854.
23. Wickner W. Yeast vacuoles and membrane fusion pathways. EMBO J 2002;21:1241–1247.
24. Noda T, Suzuki K, Ohsumi Y. Yeast autophagosomes: de novo formation of a membrane structure. Trends Cell Biol 2002;12:231–235.
25. Bonangelino CJ, Chavez EM, Bonifacino JS. Genomic screen for vacuolar protein sorting genes in *Saccharomyces cerevisiae*. Mol Biol Cell 2002;13:2486–2501.
26. Cabib E, Drgonova J, Drgon T. Role of small G proteins in yeast cell polarization and wall biosynthesis. Annu Rev Biochem 1998; 67:307–333.
27. Smits GJ, van den Ende H, Klis F.M. Differential regulation of cell wall biogenesis during growth and development in yeast. Microbiology 2001;147:781–794.