Yeast Protein Interactome topology provides framework for coordinated-functionality

André X. C. N. Valente* and Michael E. Cusick[1,2]

Biometry Research Group, Department of Health and Human Services, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892, USA, [1]Center for Cancer Systems Biology and Deptartment of Cancer Biology, Dana-Farber Cancer Institute and and [2]Deptartment of Genetics, Harvard Medical School, Boston, MA 02115, USA

ABSTRACT

The architecture of the network of protein–protein physical interactions in *Saccharomyces cerevisiae* is exposed through the combination of two complementary theoretical network measures, betweenness centrality and 'Q-modularity'. The yeast interactome is characterized by well-defined topological modules connected via a small number of inter-module protein interactions. Should such topological inter-module connections turn out to constitute a form of functional coordination between the modules, we speculate that this coordination is occurring typically in a pairwise fashion, rather than by way of high-degree hub proteins responsible for coordinating multiple modules. The unique non-hub-centric hierarchical organization of the interactome is not reproduced by gene duplication-and-divergence stochastic growth models that disregard global selective pressures.

INTRODUCTION

The set of all physical protein–protein interactions in a cell—the interactome—presents a foundational picture for biology, sitting at the lowest level of description at which it is possible to have an holistic view of a cell rather then just an isolated study of its individual components. In this article, we make a small contribution to the ongoing effort to understand the global architecture of this fundamental physical network (1–10). The interactome can be represented in an abstract way as a network of nodes connected by links, where nodes stand for proteins and links for direct physical interactions between proteins. In recent years, there has been much interest in applying statistical mechanics to the study of such complex networks (11,12). However, the validity of such an approach is always conditional on the fundamental assumptions of statistical mechanics being satisfied. For instance, many statistical measures will not be able to distinguish between a network with an intrinsic hierarchical topology and one without it (Figure 1).

RESULTS AND DISCUSSION

Our approach sidesteps the limitations alluded to in the preceding paragraph and focuses directly on how the interactome topology relates to two broad biological concepts. The first of these, hierarchical organization, is in this context the notion that there may exist a hierarchy in the role of proteins (10). On one hand, there are proteins that perform very specific, local functions, relevant only within the context of a particular biological process. On the other hand, some proteins may possess a global, high level role, perhaps acting as mediators of distinct biological processes. To study the topological hierarchy in the interactome, we use the graph theoretical betweenness centrality measure (13–15) (Supplementary Data). Betweenness centrality (denoted 'traffic', henceforth) for a node is the total number of shortest paths (between any two other nodes) in the network that pass through that node. A high traffic value for a protein therefore correlates with that protein being topologically central in the interactome. The second of these is the concept of biological functional modularity (16). In the context of proteins, at the extreme, this takes the form of protein machines performing specific functions in a cell (7). More generally, it consists of an expectation that the density of protein–protein interactions will rise as we zoom into an increasingly functionally related set of proteins. To assess modularity in the interactome topology, we use the 'Q-modularity' measure of Newman (17), which assigns a modularity score, Q, to any given partition of the network into modules. The modularity Q is defined as the difference between the ratio (intra-module edges)/(total edges) for the network in question, and the expected value for this ratio if edges in the network were randomized, subject to every node maintaining its original degree. We use the algorithm of Clauset *et al.* (18) to find an interactome partition into modules that corresponds to a large Q value.

---

*To whom correspondence should be addressed. Tel: +1 301 496 8550; Fax: +1 301 402 0816; Email: andre@deas.harvard.edu
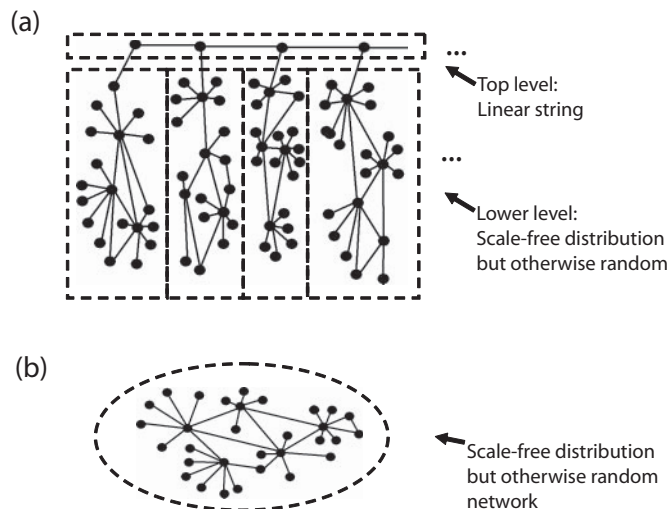
Figure 1. Hidden intrinsic hierarchy in networks. In this example sketch, network (**a**) has two hierarchical levels: it consists of a linear string of nodes at the top level that connect to a lower level set of subgraphs possessing scale-free degree distributions, but that are otherwise random. Network (**b**) has only one hierarchical level, also with a scale-free degree distribution and random in other respects. Arguably, for many applications the difference between topology (a) and topology (b) is of relevance. Yet, an analysis based on common measures such as degree distribution (11,12), clustering coefficient as a function of degree (11,44), or degree correlation measures (61), to name a few, would indicate networks (a) and (b) to be topologically identical. This is due to the much larger number of nodes at the lower hierarchical level statistically overwhelming, and therefore hiding, the top level structure.

We now explain how to produce an interactome polar map (Figure 2) by combining the information contained in the modularity and traffic analyses. The position of every protein in the map is specified in terms of its radial and angular coordinates. The radial coordinate is a function of its traffic (19). More precisely, it is proportional to log(max traffic/ protein's traffic), where 'max traffic' is the maximum node traffic in the network (Supplementary Data). A logarithmic scale is used due to the long tail of the traffic distribution (20). The protein angular coordinates are assigned such that all proteins in the same module fall within the same angular range (Supplementary Data). This way, an interactome map is created where topologically increasingly central proteins are radially increasingly closer to the center of the map, while angular sectors correspond to topological modules. To determine the circular ordering of the modules in the map, we introduce a Ring Ordering Algorithm that, based on the interactome inter-module connectivity, attempts to place closer to each other, to the extent that it is possible, modules that are more topologically related (Supplementary Data).

We apply this analysis and discuss its implications in the context of the *Saccharomyces cerevisiae* interactome. The interactome dataset we use is the higher confidence 'filtered yeast interactome' (FYI) (10), consisting of interactions supported either by small-scale screens as reported in the MIPS database (21) or by at least two distinct methods from amongst (i) high-throughput yeast two-hybrid experiments (22,23), (ii) computational predictions based on gene co-occurrence (24,25), gene neighborhood (24,26) or gene

fusion (24,25), (iii) high-throughput affinity purification/ mass spectrometric protein complex identification experiments (27,28) and (iv) small-scale or module-scale experimental identification of protein complexes as reported in MIPS (21) (Supplementary Data). We consider only the giant connected component produced by this dataset, a network containing 1741 interactions amongst 741 proteins. Our dataset choice reflects a desire to bias the data towards a thorough and accurate coverage of a limited region of the interactome as opposed to a wider, but likely shallower and more error prone, sampling. Note that when defining whether two proteins interact, a binary description is being imposed on what ideally would be characterized in terms of an affinity constant. Experimentally, effectively the aim has been for a cut-off that is high enough to exclude indiscriminate low-affinity interactions, such as those that occur between a general protein and proteasomal, ribosomal, or heat shock proteins, since in principle these are less informative interactions (29,30). Hence, such interactions are largely absent from datasets. With the vast majority of the proteins at the periphery of the map and well-defined modules connected through a handful of more central proteins, the yeast interactome polar map (Figure 2) presents what we term a coordinated-functionality architecture. Next, we discuss how this interactome architecture fits in with the biology of the cell.

An examination of the MIPS database functional (biological process) annotation of the proteins (21) demonstrates that the topological modules make very good sense as biological functional modules. On average 88.8% of the proteins in a module share a similar function based on their MIPS classification (21) (Supplementary Data), confirming earlier studies connecting topology and function (31–40). Part of the mismatch between the topological modules and current protein functional annotations will likely vanish, once more complete and accurate interactome datasets become available. However, it is the disparities not due to dataset limitations that are truly interesting, for those are instances where the functional modules based on the interactome topology are not the ones we currently assign in functional classification schema. In view of (i) the good overall match found and (ii) the foundational role of the interactome in the cell, we propose that the interactome topology represents a fundamental source for the division of proteins into functional groups. As such, its modularity analysis provides a rigorous alternative to the currently subjective functional annotation present in protein databases. In accordance, we name the topological modules found so as to reflect their perceived biological roles (Figure 2).

The average degree of essential (41,42) proteins in the dataset is 5.7, while that of non-essential proteins is 3.9, a difference that is too large to be attributable to chance alone (43) (Supplementary Data). This difference may indeed be biologically meaningful. Although physically knocking out a gene associated with a high or a low degree protein, say of degree 2 or 10 respectively, may be considered equivalent, from a mathematical network perspective it is not. In one case, it involves deleting 1 node and 2 links, in the other it involves deleting 1 node and 10 links. Note that such an explanation would not involve ascribing any out of the ordinary, higher-level role to hubs, the large degree proteins. Alternatively, the observed higher average degree
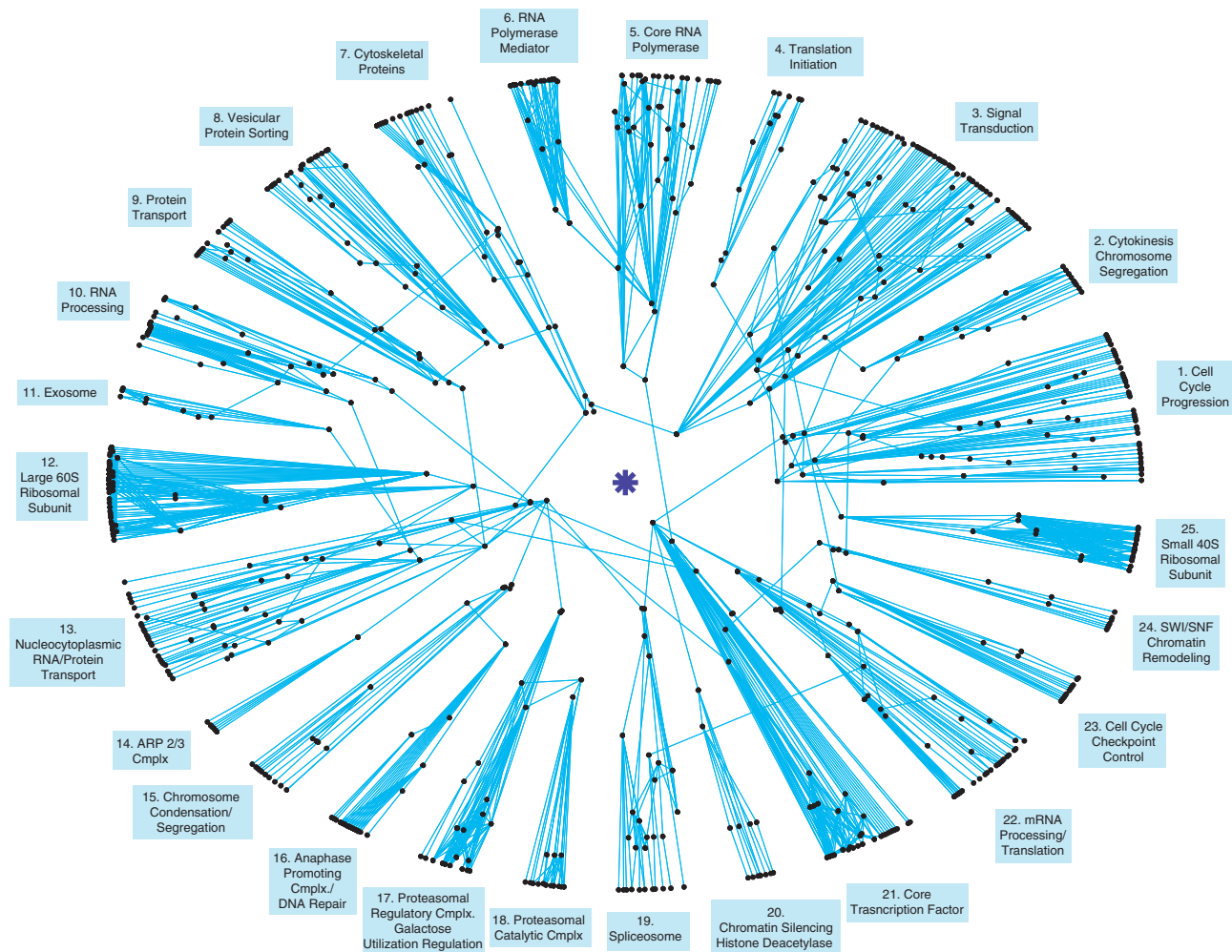
Figure 2. *S.cerevisiae* interactome polar map. The map is constructed, in an unsupervised manner, based solely on protein–protein interaction data. The module captions (blue text boxes) were manually chosen, a posteriori, to reflect the biological role of each module. The map suggests a 'coordinated-functionality' architecture for the interactome, arguably an ideal framework for the cell to physically implement the concept of distinct, yet coordinated, biological functional modules. This would be a pairwise-coordination, as inter-module physical interactions occur in a pairwise fashion: of the 76 proteins that possess inter-modular connections, only four connect their module to more than a single other module (TAF25 in module #21 and SRP1 in module #13 have links to four other modules, while NUP1 in module #13 and CLB2 in module #1 have links to two other modules). The map is based on the higher confidence FYI protein–protein interaction dataset (10), consisting of interactions validated either through small-scale experiments or through at least two distinct procedures. The giant connected component, shown here, consists of 741 proteins and 1752 protein–protein interactions.

of essential proteins could still stem from lingering systematic biases in the FYI dataset.

While the translation of the interactome modular topology into biological functional modularity is straightforward, this is not the case for the interactome topological hierarchical organization. A fundamental question that the traffic analysis gives rise to is how the interactome topological hierarchical organization phenomenologically expresses itself. For instance, do more hierarchically central proteins in fact perform a higher-level, coordinating role in the cell? At present, these are open biological questions.

Comparing proteins of equal degree, we find no significant correlation between a protein's traffic and its essentiality (41,42) (Supplementary Data), something not too surprising, as knocking out a key protein that renders an essential functional module inoperant is plausibly more damaging than knocking out a protein that mediates two distinct

processes that nonetheless can still function independently. An intriguing, though at the moment still unsupported hypothesis is that, if a protein is disrupted, its traffic level correlates with the likelihood of causing non-lethal side effects in multiple areas of the biology of the cell. Speculating further, perhaps our representation of the interactome can provide clues as to where those side effects may arise—a matter of critical importance in drug development. A different possibility is that, for some of these module-connecting proteins, interacting with multiple modules is not a sign of a role in coordinating the functionality of the modules, but rather just a result of the protein being independently used in those modules. In opposition to a true functional 'connector', we would call such a protein a 'bolt' (alternatively, 'widget'), in reference to how, analogously, a mechanical bolt can be used in multiple functional modules of a human engineered machine, while playing no role in coordinating their

functionality. Finally, note that a priori the observed 47 inter-module interactions are particularly susceptible to be false-positives, because a false-positive interaction between two random unrelated proteins is likely to result in an inter-module, high traffic interaction. However, significantly, 45 out of these 47 inter-module interactions belong to the set of interactions supported by small-scale targeted experiments and arguably it is not very likely that a false-positive interaction of the type just described would go unnoticed in a targeted experiment and further make it into the peer-reviewed literature. Of the remaining 2 inter-module interactions, one is reported in the Ito *et al.* (23) and Uetz *et al.* (22) high-throughput yeast two-hybrid datasets as well as in the MIPS dataset of protein complexes identified via small- or module-scale experiments (21), while the other is reported in the Gavin *et al.* (27) high-throughput protein complex identification study and again in the Uetz *et al.* (22) yeast two-hybrid dataset. Out of the 45 supported by small-scale experiments, four are also reported in a high-throughput yeast two-hybrid study (22,23), three in a high-throughput protein complex identification study (27,28) and two in the MIPS dataset of protein complexes identified via small or module-scale experiments (21). Whatever biological role central proteins turn out to play, we submit that they call for further experimental investigation, given their unique topological placement in the interactome.

Having hierarchically classified the proteins with the traffic measure, we are now in a position to consider network degree distribution related questions without falling prey to previously noted statistical problems (Figure 1). Of particular interest is how degree changes as one moves hierarchically across the interactome (14) (Figure 3a). Surprisingly, nodes of different degree are rather homogeneously hierarchically spread across the interactome: note the large spread between the green, red and blue curves relative to their small positive slopes; or, for a more quantifiable attribute, how the 10% of nodes with the largest degree in the periphery of the inter-actome have a significantly larger degree than the average degree at the center of the interactome. Thus, the interactome is not hierarchically stratified by degree. In particular, the interactome has a non-hub-centric hierarchical organization. Further, note that should the inter-module protein interactions indeed represent a form of functional coordination, then this coordination is apparently occurring overwhelmingly in a pair-wise fashion: out of the 76 proteins that possess links to modules outside their own, only 4 connect their home module to more than one other module (TAF25 and SRP1 have links to four other modules, NUP1 and CLB2 to two other modules). The other 72 proteins connect their home module to a single other module (Supplementary Data). It is also noteworthy that, amongst these 76 connecting proteins, the higher degree ones in general belong clearly in their assigned home modules (specifically, for connecting proteins of degree 4 or higher, let us exclude the two interactions that each of these proteins must have by default to connect it to its home module and to one linked module; then, 95.3% of the remaining interactions of connecting proteins are with the protein's respective home module. Supplementary Data). This pairwise-coordination is in sharp contrast with the picture of a hub protein connecting and mediating multiple modules (11,43,44).
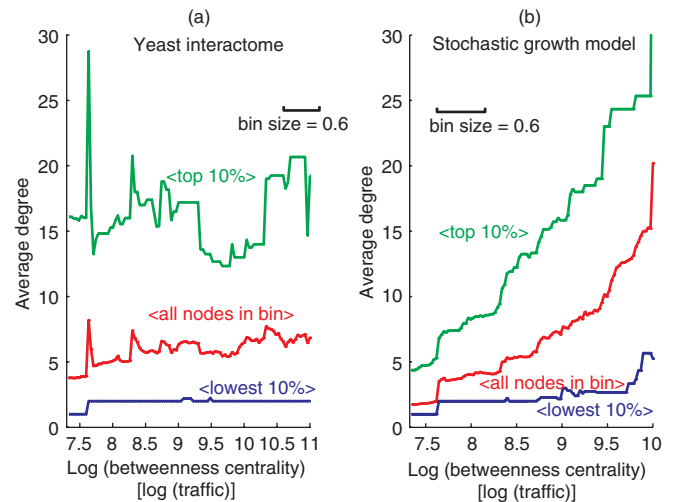

Figure 3. A non-hub-centric hierarchical organization. In the yeast inter-actome (**a**), the degree distribution does not change greatly as one moves from the periphery to the center of the Figure 2 polar map (i.e, as one moves from a low to a high traffic region). In other words, hubs are not hierarchically central in the yeast interactome. In contrast, gene duplication and divergence inter-actome stochastic growth models [45–47] produce hub-centric interactomes, where the average degree markedly increases with traffic and the hierarchical center of the interactome is therefore dominated by hubs. (**a**) Analysis for the yeast interactome giant connected component, based on the FYI data set (741 proteins, 1752 interactions) [10]. Red curve—the average degree for the set of nodes whose log (traffic) value falls within 0.3 of the log (traffic) value indicated in the *x*-axis. That is, a log (traffic) bin of size 0.6 is continuously slid along the log (traffic) axis and the average degree for all the nodes that fall within the bin is calculated. The last bin also includes all nodes with a log (traffic) value larger than the range shown in the figure. The bins cover the entire data set, with every bin containing at least 26 nodes. The first bin contains 417 nodes. The last bin contains 37 nodes. Green curve—similar to the red curve, except this time the degree average is done only over the 10% largest degree nodes in the bin. Blue curve—similar to green curve, but this time averaging over the 10% lowest degree nodes in the bin. The average degree in the highest traffic bin (rightmost data point in the red curve) is only 0.42 times the average degree of the 10% largest degree nodes in the lowest traffic bin (leftmost data point in the green curve). (**b**) Corresponding plots for the giant component of an interactome evolved under the gene duplication and divergence stochastic growth model of Pastor-Satorras *et al.* [45]. In this case, the giant component contains 759 nodes and 1542 interactions. Every bin contains at least 26 nodes. The first and last bins contain 432 and 40 nodes, respectively. The average degree in the highest traffic bin is now 4.6 times larger than the average degree of the 10% largest degree nodes in the lowest traffic bin. Similar results were achieved under multiple trials, model parameters and gene dupli-cation growth models (Supplementary data.).

The non-hub-centric organization runs contrary to a number of network growth models that have been proposed to explain the topology of the interactome (11,45–48). The models are based on evolution by stochastic gene duplication and divergence (49,50). Amongst other reproduced statistics, the models are able to generate the power-law degree distribution observed in early interactome datasets (51). Since these models do not make an appeal to evolutionary selection pressures, a major conclusion taken from their success was that natural selection is not required to reproduce the global structure of the interactome; instead, stochastic gene duplication and divergence suffices to give rise to that topology (45,47). However, here we report that these gene duplication models lead to hierarchically hub-centric networks. In Figure 3b, we show data pertaining to an interactome built using the model
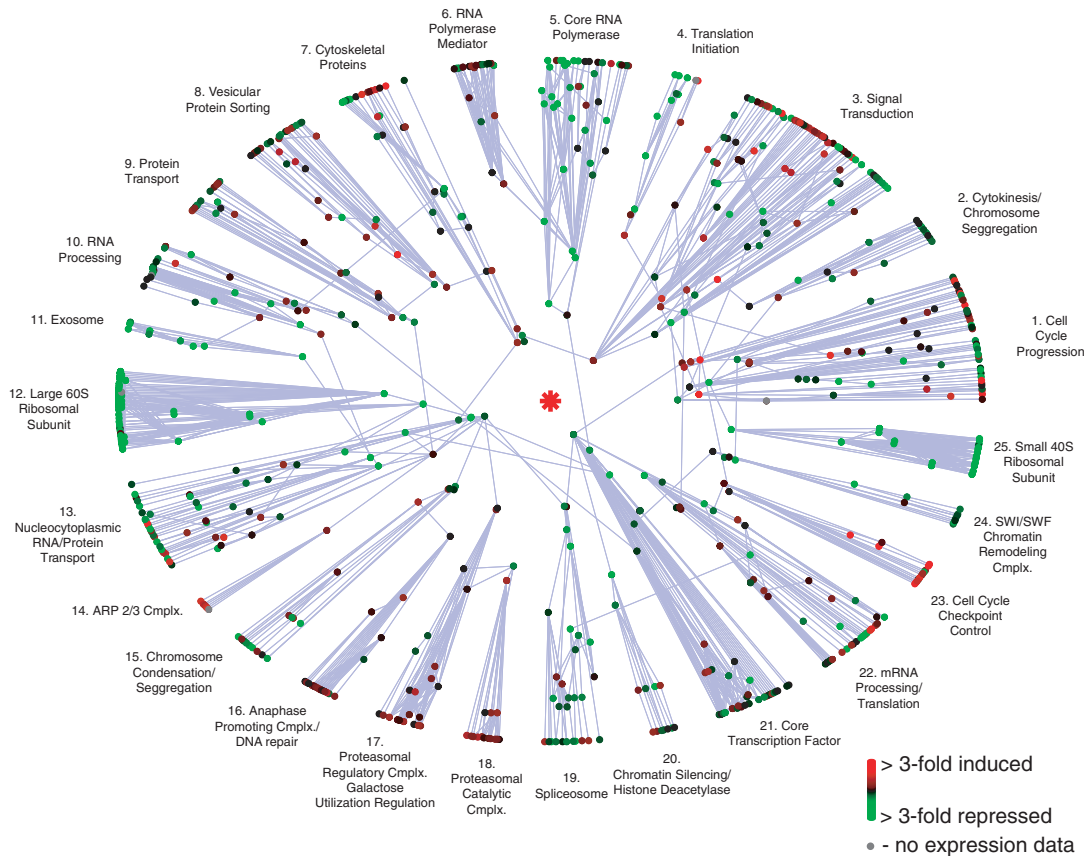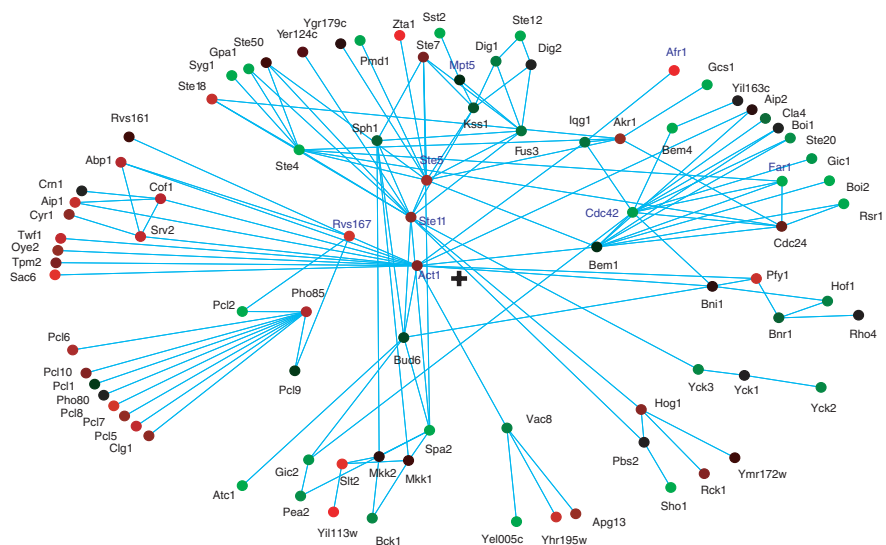
Figure 4. Overlay of 20 min post heat shock mRNA expression data (62) upon the interactome polar map. Maps are constructed as in Figure 2 and Supplementary Fig. 8 (proteins connected to other modules in the global map have names in blue). In addition, node color indicates mRNA expression level. As previously reported (62), there is a sharp decline in expression of mRNA for ribosomal proteins. Other modules with a significant proportion of constituent proteins showing a decline are the Translation Initiation and Core RNA Polymerase modules. These observations are consistent with the repression of genes involved in RNA and protein synthesis upon heat shock (62). Modules with a significant proportion of constituent proteins showing an increase in expression are the small Cell Cycle Checkpoint Control module and parts of the Signal Transduction and Cell Cycle Progression modules. Overlay of mRNA expression data upon the single module maps of the latter two modules shows clear over expression in the protein chaperone submodule of CellCycle Progression (Supplementary Fig. 9), and into the phosph-cyclin and cell growth/morphogenesis submodules of Signal Transduction (above). The increased expression of genes in these submodules supports the notion that cell growth is checked upon heat shock.

of Pastor-Satorras *et al.* (45). By comparison with the same plot for the yeast interactome, this time the network is clearly stratified by degree, with the larger degree nodes concentrated at the hierarchical center of the network. Now it is the average degree at the center of the network that is significantly larger than the average degree of the 10% of nodes with largest degree at the periphery. The models of Vázquez (46) (slightly different implementation of the gene duplication and divergence process) and of Wagner (47) (emphasizing a continuous divergence in the form of gain and loss of interactions amongst existing proteins) produce similar hub-centric networks (Supplementary Data). In summary, there are at least three possible explanations for the non-hub-centric hierarchy we observe in yeast: (i) it is a spurious effect associated with the limitations of existing data and the interactome is in fact hub-centric; (ii) there is some crucial feature of the gene duplication and divergence mechanics that is not understood and/or is not captured by current models or (iii) the non-hub-centric hierarchy is in fact shaped by natural selection pressures on the global interactome structure.

Our study is based on the present-day knowledge of the yeast interactome, which is still rather deficient (52–58). We minimized false-positives using the higher confidence yeast FYI dataset as the source for our study. The good correspondence between the topological modular breakdown of the interactome and the known functional annotation of proteins corroborates that false-positives are not an overriding problem in this dataset. Nonetheless, we repeated the interactome analysis using a dataset of interactions reported in the MIPS database that are validated through small-scale screenings, the most reliable source of data (55) (giant component: 392 proteins, 675 interactions. Supplementary Data). The correspondence between topological and functional modules (now on average 92% of the proteins in a module shared a similar function based on their MIPS classification), as well as the reported non-hub-centric hierarchy were again supported by this small-scale dataset (Supplementary Data). Regarding the limited coverage of our datasets [FYI ≈12% coverage, small-scale ≈7% coverage, assuming ≈6000 proteins in yeast (23)], it is of note that the doubling in size of the network, going from the small-scale to the FYI dataset, did not dilute the observed non-hub-centric topology nor the pair-wise inter-module connectivity pattern. In fact, the FYI dataset produces an even slightly less hub-centric interactome than the small-scale dataset (Supplementary Data). Likewise, the pairwise inter-module connectivity pattern is no less present in the FYI than in the small-scale dataset (where out of the 47 proteins with inter-module links, four connect their home module to more than one other module. Supplementary Data). The observation that higher degree connecting proteins are in general strongly attached to their home modules is equally confirmed in both datasets (the earlier mentioned 95.3% of home module interactions for the FYI higher degree connecting proteins, now becomes 95.7% in the small-scale dataset). Still, it is important to bear in mind the limitations of current datasets. For instance, regarding the typical pairwise coordination, the possibility that this observation is only the result of a high-number of false-negatives for inter-module protein interactions cannot be ruled out. Note, for example, how in the

map there are no interactions between the translation initiation module and the ribosomal subunit modules, even though such interactions must certainly exist. Ultimately, only the generation of more accurate and comprehensive interactome datasets can unequivocally confirm or disprove some of the results and hypotheses put forward in this article (52,53).

So far our analysis has focused on the global interactome topology. Now we would also like to highlight its potential as a framework for exploiting the wealth of interactome data. The interactome can form a valuable platform for crystalizing biological thought. We briefly introduce two relevant extensions to our work. First, one may zoom into a module of interest in the interactome and locally repeat the analysis, producing a single module polar map (Figures in Supplementary Data). Such a local map provides one with a starting point to discuss the biology of the process under study, interpret and design experiments, and generate new biological hypotheses. Second, we note that the entire proteome is rarely, if ever, evenly expressed by the cell (49). Therefore, perhaps the interactome is best viewed as a potential network at the cell's disposal, with different parts of it being turned on and off to different degrees, as biologically required. Integrating mRNA expression data with the interactome polarmap (59) (Figure 4), permits a proper, unified analysis of this dynamical network.

The interactome represents an elementary abstraction of the multitude of complex biochemical interactions taking place in the rich physiological environment of the cell. In the trade-off between simplicity and realism, arguably some facts may be beneficially incorporated in future interactome models: For instance, protein interactions vary within a continuum of binding affinities and post-translational modifications, as well as allosteric interactions effectively change the possible binding partners of a protein, to name a few of the more prominent omissions at present (49). We end by noting that the organization of a network through a procedure akin to the one used in this paper may also be of relevance to the problem of network motif finding (60), as certain motifs may turn out to occur sparsely overall and yet be statistically significant in specific regions of the network (e.g. only in the high traffic central area, or only in a particular module).

SUPPLEMENTARY DATA

Supplementary data are available at NAR online

ACKNOWLEDGEMENTS

Conflict of interest statement. None declared.

REFERENCES

1. Eisenberg,D., Marcotte,E.M., Xenarios,I. and Yeates,T.O. (2000) Protein function in the post-genomic era. *Nature*, **405**, 823–826.
2. Lu,H., Zhu,X., Liu,H., Skogerbo,G., Zhang,J., Zhang,Y., Cai,L., Zhao,Y., Sun,S., Xu,J. *et al.* (2004) The interactome as a tree—an attempt to visualize the protein-protein interaction network in yeast. *Nucleic Acids Res.*, **32**, 4804–4811.
3. Vidal,M. (2001) A biological atlas of functional maps. *Cell*, **104**, 333–339.
4. Bader,G.D., Heilbut,A., Andrews,B., Tyers,M., Hughes,T. and Boone,C. (2003) Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends Cell Biol.*, **13**, 344–356.
5. Sharom,J.R., Bellows,D.S. and Tyers,M. (2004) From large networks to small molecules. *Curr. Opin. Chem. Biol.*, **8**, 81–90.
6. Tucker,C.L., Gera,J.F. and Uetz,P. (2001) Towards an understanding of complex protein networks. *Trends Cell Biol.*, **11**, 102–106.
7. Alberts,B. (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell,*, **92**, 291–294.
8. Bork,P., Jensen,L.J., von Mering,C., Ramani,A.K., Lee,I. and Marcotte,E.M. (2004) Protein interaction networks from yeast to human. *Curr. Opin. Struc. Biol.*, **14**, 292–299.
9. Gavin,A.-C. and Superti-Fuga,G. (2003) Protein complexes and proteome organization from yeast to man. *Curr. Opin. Chem. Biol.*, **7**, 21–27.
10. Han,J.D., Bertin,N., Hao,T., Goldberg,D.S., Berriz,G.F., Zhang,L.V., Dupuy,D., Walhout,A.J., Cusick,M.E., Roth,F.P. *et al.* (2004) Evidence for dynamically organized modularity in the yeast protein–protein interaction network. *Nature*, **430**, 88–93.
11. Barabási,A.-L. and Oltvai,Z.N. (2004) Network biology: understanding the cell's functional organization. *Nature Rev. Genet.*, **112**, 101–114.
12. Newman,M.E.J. (2003) The structure and function of complex networks. *SIAM Rev.*, **45**, 167–256.
13. Freeman,L. (1977) A set of measures of centrality based upon betweenness. *Sociometry*, **40**, 35–41.
14. Joy,M.P., Brock,A., Ingber,D.E. and Huang,S. (2005) High betweenness proteins in the yeast protein interaction network. *J. Biomed. Biotechnol.*, **2**, 96–103.
15. Hahn,M.W. and Kern,A.D. (2005) Comparative genomics of centrality and essentiality in three eukaryotic protein interaction networks. *Mol. Biol. Evol.*, **22**, 803–806.
16. Hartwell,L.H., Hopfield,J.J., Leibler,S. and Murray,A.W. (1999) From molecular to modular cell biology. *Nature*, **402**, C47–C51.
17. Newman,M.E.J. (2004) Fast algorithm for detecting community structure in networks. *Phys. Rev.*, **E69**, Art. No. 066133.
18. Clauset,A., Newman,M.E.J. and Moore,C. (2004) Finding community structure in very large networks. *Phys. Rev.*, **E70**, art. no. 066111.
19. Brandes,U. (2003) Communicating centrality in policy network drawings. *IEEE Trans Vis. Comput. Graph.*, **9**, 241–253.
20. Goh,K.-I., Kahng,B. and Kim,D. (2001) Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.*, **87**, art. no. 278701.
21. Mewes,H.W., Frishman,D., Guldener,U., Mannhaupt,G., Mayer,K., Mokrejs,M., Morgenstern,D., Munsterkotter,M., Rudd,S. and Weil,B. (2002) MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.*, **30**, 31–34.
22. Uetz,P., Giota,L, Cagney,G., Mansfield,T.A., Judson,R.S., Knight,J.R., Lockshon,D., Narayan,V., Srinivasan,M. and Pochart,P. (2000) A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature*, **403**, 623–627.
23. Ito,T., Chiba,T., Ozawa,R., Yoshida,M., Hattori,M. and Sakaki,Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Nature*, **403**, 623–627.
24. von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, **417**, 399–403.
25. Marcotte,E.M., Pellegrini,M., Thompson,M.J., Yeates,T.O. and Eisenberg,D. (1999) A combined algorithm for genome-wide prediction of protein function. *Nature*, **402**, 83–86.
26. Dandekar,T., Snel,B., Huynen,M. and Bork,P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem. Sci.*, **23**, 324–328.
27. Gavin,A.-C., Bosche,M., Krause,R., Grandi,P., Marzioch,M., Bauer,A., Schultz,J., Rick,J.M., Michon,A.M., Cruciat,C.M. *et al.* (2002)
Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature*, **415**, 141–146.
28. Ho,Y., Gruhler,A., Heilbut,A., Bader,G.D., Moore,L., Adams,S.L., Millar,A., Taylor,P., Bennett,K., Yang,L. *et al.* (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature*, **415**, 180–183.
29. Krause,K., von Mering,C., Bork,P. and Dandekar,T. (2004) Shared components of protein complexes—versatile building blocks or biochemical artefacts? *BioEssays*, **26**, 1333–1343.
30. Soti,C., Pal,C., Papp,B. and Csermely,P. (2005) Molecular chaperones as regulatory elements of cellular networks. *Curr. Opin. Cell. Biol.*, **17**, 210–215.
31. Schwikowski,B., Uetz,P. and Fields,S. (2000) A network of protein–protein interactions in yeast. *Nat. Biotechnol.*, **18**, 1257–1261.
32. Rives,A.W. and Galitski,T. (2003) Modular organization of cellular networks. *Proc. Natl Acad. Sci. USA*, **3**, 1128–1133.
33. Bu,D., Zhao,Y., Cai,L., Xue,H., Zhu,X., Lu,H., Zhang,J., Sun,S., Ling,L., Zohang,N. *et al.* (2003) Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic Acids Res.*, **31**, 2443–2450.
34. Vázquez,A., Flammini,A., Maritan,A. and Vespignani,A. (2003) Global protein function prediction from protein–protein interaction networks. *Nat. Biotechnol.*, **21**, 697–700.
35. Spirin,V. and Mirny,A. (2003) Protein complexes and functional modules in molecular networks. *Proc. Natl Acad. Sci. USA,*, **21**, 12123–12128.
36. Zhou,H. (2003) Distance, dissimilarity and network community structure. *Phys. Rev.*, **E67**, art. no. 06101.
37. Pereira-Leal,J.B., Enright,J.A. and Ouzounis,C.A. (2004) Detection of functional modules from protein interaction networks. *Proteins*, **54**, 49–57.
38. Brun,C., Herrmann,C. and Guénoche,A. (2004) Clustering proteins from interaction networks for the prediction of cellular functions. *BMC Bioinformatics*, **5**, Art. No. 95.
39. Dunn,R., Dudbridge,F. and Sanderson,C.M. (2005) The use of edge-betweenness clustering to investigate biological function in protein interaction networks. *BMC Bioinformatics*, **6**, Art. no. 39.
40. Nabieva,E., Jim,K., Agarwal,A., Chazelle,B. and Singh,M. (2005) Whole-proteome prediction of protein function via graph-theoretic analysis of interaction maps. *Bioinformatics*, **21**, I302–I310.
41. Winzeler,E.A., Shoemaker,D.D., Astromoff,A., Liang,H., Anderson,K., Andre,B., Baugham,R., Boeke,J.D., Bussey,H., Chu,A.M., Connelly,C. *et al.* (1999) Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
42. Giaever,G., Chu,A.M., Ni,L., Connelly,C., Riles,L., Veronneau,S., Dow,S., Lacau-Danila,A., Anderson,K., Andre,B., Arkin,A.P. *et al.* (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature*, **418**, 387–391.
43. Jeong,H., Mason,S.P., Barabási,A.-L. and Oltvai,Z.N. (2001) Lethality and centrality in protein networks. *Nature*, **411**, 41–42.
44. Ravasz,E. and Barabási,A.-L. (2003) Hierarchical organization in complex networks. *Phys. Rev.*, **E67**, Art. no. 026112.
45. Pastor-Satorras,R., Smith,B. and Solé,R.V. (2003) Evolving protein interaction networks through gene duplication. *J. Theor. Biol.*, **222**, 199–210.
46. Vázquez,A., Flammini,A., Maritan,A. and Vespignani,A. (2003) Modeling protein in teraction networks. *Complexus*, **1**, 38–44.
47. Wagner,A. (2003) How the global structure of protein interaction networks evolves. *Proc. Biol. Sci.*, **270**, 457–466.
48. Middendorf,M., Ziv,E. and Wiggins,C.H. (2005) Inferring network mechanisms: the *Drosophila melanogaster* protein interaction network. *Proc. Natl Acad. Sci. USA,*, **102**, 3192–3197.
49. Alberts,B., Johnson,A., Lewis,J., Raff,M., Roberts,K. and Walter,P. (2002) *Molecular Biology of the Cell,* 4th edn. Garland Science, NY.
50. Ohno,S. (1970) *Evolution by Gene Duplication.* Springer, Berlin.
51. Tanaka,R., Yi,T.-M. and Doyle,J. (2005) Some protein interaction data do not exhibit power law statistics. *FEBS Lett.*, **579**, 5140–5144.
52. Cornell,M., Paton,N.W. and Oliver,S.G. (2004) A critical and integrated view of the yeast interactome. *Comp. Funct. Genom.*, **5**, 382–402.
53. Hakes,L., Robertson,D.L. and Oliver,S.G. (2005) Effect of dataset selection on the topological interpretation of protein interaction networks. *BMC Genomics*, **6**, 131.
54. Ito,T., Ota,K., Kubota,H., Yamayuchi,Y., Chiba,T., Sakuraba,K. and Yoshida,M. (2002) Roles for the two-hybrid system in exploration of

the yeast protein interactome. *Mol. Cell Proteomics*, **1**, 561–566.

55. von Mering,C., Krause,R., Snel,B., Cornell,M., Oliver,S.G., Fields,S. and Bork,P. (2002) Comparative assessment of large-scale data sets of protein–protein interactions. *Nature*, **417**, 399–403.

56. Bader,G.D. and Hogue,C.W. (2002) Analyzing yest protein–protein interaction data obtained from different sources. *Nat. Biotechnol.*, **20**, 991–997.

57. Aloy,P. and Russell,R.B. (2002) The third dimension for protein interactions and complexes. *Trends Biochem. Sci.*, **27**, 633–638.

58. Edwards,A.M., Kus,B., Jansen,R., Greenbaum,D., Greenblatl,J. and Gerstein,M. (2002) Bridging structural biology and genomics: assessing protein interaction data with known complexes. *Trends Genet.*, **18**, 529–536.

59. Lichtenberg,U., Jensen,L.J., Brunak,S. and Bork,P. (2005) Dynamic complex formation during the yeast cell cycle. *Science*, **307**, 724–727.

60. Milo,R., Shen-Orr,S., Itzkovitz,S., Kashtan,N., Chklovskii,D. and Alon,U. (2002) Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.

61. Maslov,S. and Sneppen,K. (2002) Specificity and stability in topology of protein networks. *Science*, **296**, 910–913.

62. Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D. and Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to envoironmental changes. *Mol. Biol. Cell.*, **11**, 4241–4257.