

Minireview

Untangling the web of functional and physical interactions in yeast

Markus J Herrgård and Bernhard Ø Palsson

Address: Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0412, USA.

Correspondence: Bernhard Ø Palsson. E-mail: palsson@ucsd.edu

Published: 8 June 2005

Journal of Biology 2005, **4**:5

The electronic version of this article is the complete one and can be found online at <http://jbiol.com/content/4/2/5>

© 2005 BioMed Central Ltd

Abstract

An analysis of an integrated network of over 150,000 functional and physical interactions in yeast suggests that the network can be hierarchically decomposed into themes and thematic maps. This decomposition can be used to explore the organizational principles of integrated biological networks within cells.

Reconstructing the structures of the regulatory and signaling networks operating within and between cells is a key component of the 'systems biology' approach to understanding biological processes [1]. The multitude of high-throughput datasets now available for an increasing number of model organisms has enabled the development of systematic approaches for this reconstruction task for specific types of networks. A number of approaches have been developed for identifying co-regulated gene modules from gene expression and protein-DNA binding data [2]. Similarly, protein-protein interaction data have been used to reconstruct signaling networks and protein complexes [3]. These reconstructions form a blueprint of the networks operating within the cell and can be used as a starting point for further studies on network function.

One of the key uses of network reconstructions is deciphering the general organizational principles of biological networks; these principles that are commonly studied include topological and functional modularity [4,5], network redundancy and robustness [6,7], and pathway cross-talk [8,9]. For certain types of network, including those involved

in metabolism or well-characterized signaling pathways, such organizational principles can be discovered through a set of well-established computational methods in an unbiased fashion [5,10]. There is, however, a need to develop approaches for discovering fundamental organizational principles of integrated networks that combine, for example, signaling and transcriptional pathways.

High-throughput data about a variety of interactions between genes (and gene products) can be used both to reconstruct networks and to elucidate the organizational principles of these networks. Information about features such as protein-protein [11] and protein-DNA [12] interactions can be directly interpreted as physical associations between macromolecules. In principle, such physical interaction data types can be used directly to reconstruct signaling and regulatory networks, but the noisiness of high-throughput datasets makes the reconstruction task challenging. Other types of interaction information, such as the list of genetic interactions obtained from synthetic lethality screens [13] - showing that combination of mutations in two genes causes lethality while either mutation

alone does not - do not necessarily indicate direct physical interactions between the gene products. Links between two genes can also be derived from computational analysis of datasets that include information about both genes - for example, on the basis of high gene expression or sequence similarity. These types of links are also likely to be indirect in that, in many cases, a direct physical interaction does not underlie each functional link [14].

The study by Zhang *et al.* [15] in *Journal of Biology* presents a novel approach for integrating multiple types of biological interactions so as to reconstruct simultaneously the modular structures of cellular networks and to identify their organizational principles. The authors use five different yeast datasets at the same time in order to allow integrated analysis of multiple data types and to study the interplay of different biological network types. The interaction data types include both direct physical interactions - protein complexes and protein-DNA interactions - and indirect functional interactions - genetic interactions and interactions based on gene-expression correlation and sequence homology. The overall approach used in this work (see Figure 1) first decomposes the combined interaction network into significantly enriched multi-color 'network motifs', in which each color corresponds to one type of interaction data, and then assembles these motifs into 'network themes' consisting of overlapping motifs. The network themes are further assembled into 'thematic maps' that represent a bird's-eye view of functional relationships between different subsystems of the overall biochemical network.

The network decomposition approach used by Zhang *et al.* [15] to untangle the complex network of 154,759 functional or physical links between a total of 5,831 yeast genes is based on the concept of network motifs. This concept, first introduced in the context of transcriptional regulatory networks [16,17], refers to small subnetworks that occur in the overall network more commonly than is expected by chance. Zhang *et al.* [15] extend the concept of a network motif to networks with multiple types of interaction by considering structurally similar motifs with different color interactions. The authors identify enriched three-gene motifs out of the total set of 50 possible three-gene patterns with different types of interaction connecting the genes. Because of the large number of potential multi-color four-gene motifs, a similar search was performed for only a small subset of these.

Although network motifs can be considered to be the building blocks of a graphical representation of biological networks, they may not necessarily correspond to functional building blocks of the actual networks inside a cell. This observation, originally made in the context of regulatory networks [18], led Zhang *et al.* [15] to introduce a new

concept, the network theme, to describe a collection of overlapping motifs of the same type. For example, multiple motifs containing a transcription factor that regulates two physically interacting proteins can be combined into a single theme corresponding to a protein complex whose component proteins are controlled by the same transcription factor (Figure 1a,b). The same motif-aggregation process can be applied to other types of multi-color motifs to identify a range of different types of theme. Many of the resulting themes correspond directly to modules that can be identified on the basis of physical interaction data alone, such as co-regulated gene groups or protein complexes. The inclusion of functional interaction data, however, allows the identification of network themes that bridge multiple types of physical or functional interaction.

The most interesting themes identified by Zhang and coworkers [15] are ones involving genetic interactions. These include the 'alternative subunits' theme, in which two genes are connected to each other by synthetic lethal interactions and to other members of the complex by protein-protein interactions. The second genetic-interaction-based theme consists of two protein complexes internally connected by protein-protein interactions that are bridged by a large number of genetic interactions (Figure 1d,e). This theme indicates a structure in which either of the complexes is needed to perform an essential function, but the complexes can compensate for one another. Because the same complex can compensate for the function of more than one other complex, all the 'compensatory complex' themes can be further assembled into a thematic map (Figure 1f). In this map, nodes correspond to individual protein complexes and edges to bundles of synthetic lethal interactions between complexes. The 'compensatory complex' map (Figure 1f) provides a global view of the built-in redundancies in the yeast biochemical network and allows the establishment of novel links between diverse functional processes.

Both of the network themes discussed above have genetic interactions as their core components. The interpretation of these interactions has attracted increasing attention with the development of experimental methods for the systematic discovery of genetic interactions in yeast [13,19-22]. A recent paper by Kelley and Ideker [23] approaches this interpretation task from a different angle from that of Zhang *et al.* [15], but arrives at qualitatively similar results. In their work, Kelley and Ideker [23] seek to interpret genetic interactions in yeast through the physical interaction network between this organism's genes that can be generated from protein-protein and protein-DNA interaction data, as well as gene-gene interactions derived from the metabolic network (that was extracted from the KEGG database [24]).

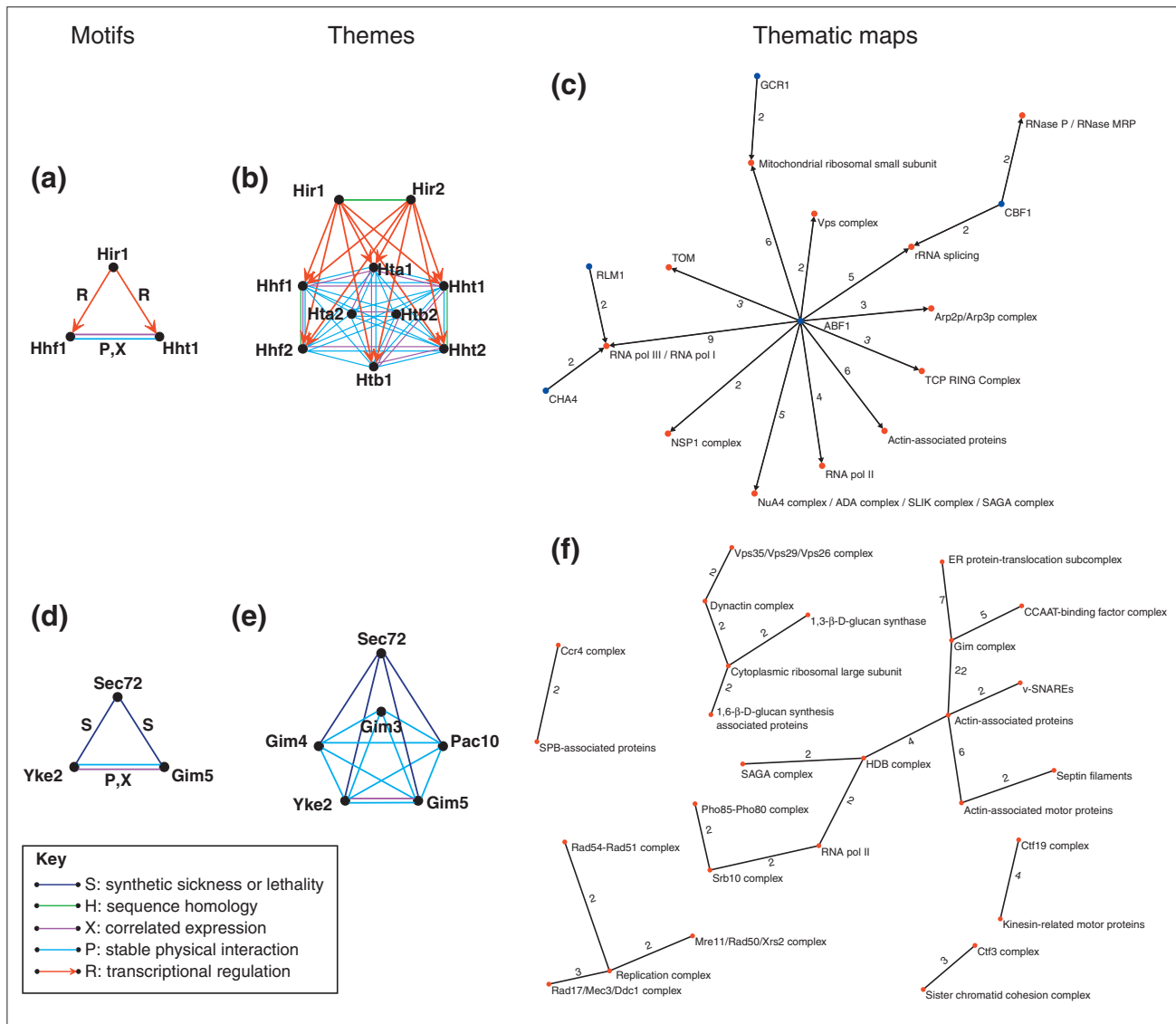


Figure 1

Examples of motifs, themes, and thematic maps discovered in Zhang *et al.* [15]. **(a-c)** A theme capturing the co-regulation of members of a protein complex by a pair of transcription factors, and a corresponding thematic map of complex co-regulation relationships. **(d-f)** The ‘compensatory complex’ theme and the corresponding thematic map. See text for further details; images reproduced from [15].

They focus on interpreting genetic interaction networks by ‘between-pathways’ or ‘within-pathway’ models encoded in the network of physical interactions. Within-pathway models refer to cases where genetic interactions occur between genes whose gene products participate in the same pathway or complex, whereas between-pathway models refer to cases where genetic interactions connect genes whose gene products operate in two distinct pathways or complexes. The ‘compensatory complex’ theme described by Zhang *et al.* [15] corresponds to the between-pathway

interpretation of a dense bundle of genetic interactions. On the other hand, the network theme in which components of the same complex have genetic interactions with each other corresponds to the within-pathway interpretation of Kelley and Ideker [23].

Kelley and Ideker [23] constructed a network of all the between-pathway explanations, identified using their method, which corresponds to the thematic map of compensatory complexes constructed by Zhang *et al.* [15]. While

these two maps share components and interactions, there also appear to be some differences; these are most likely to be due to the use of different underlying physical interaction datasets and the different computational methods used in the two studies. The approach of Zhang *et al.* [15] integrates a more diverse set of interaction data, including expression correlations and homology-based links, and hence can discover network themes that have weaker support from only one type of data such as protein-protein interactions. On the other hand, Kelley and Ideker [23] use an explicit probabilistic model of genetic and physical interactions that allows prediction of new protein functions and genetic interactions in addition to studying the organizational principles of the integrated network.

There are a number of future directions that can be taken to extend the work by Zhang *et al.* [15] on identifying themes and constructing thematic maps. Their current work does not account for metabolic networks that play a key role in the overall cellular function of yeast and that have connections to many other types of network. Methods have been developed for determining functional interactions between genes in metabolic networks based on the mass-balance stoichiometric structure of the network [10,25]; thus the motif-based approach could be directly extended to these networks. Zhang *et al.* [15] also apply their approach to a static interaction network, whereas in reality only subsets of these interactions are active under any particular biological condition. Extending their analysis to condition-dependent network structures identified by, for example, combining gene expression and physical interaction data [26,27], would allow the identification of condition-dependent thematic maps.

Given the generality of the approach introduced by Zhang *et al.* [15], it can readily be extended to different types of cellular networks so as to decipher the interplay of these networks as relevant experimental data become available. In conjunction with complementary methods, such as the one described by Kelley and Ideker [23], the thousands of physical and functional interactions that exist within all cells can begin to be untangled to provide a basis for detailed network reconstruction and to help elucidate the fundamental organizational principles of biological networks.

References

1. Palsson B: **Two-dimensional annotation of genomes.** *Nat Biotechnol* 2004, **22**:1218-1219.
2. Siggia ED: **Computational methods for transcriptional regulation.** *Curr Opin Genet Dev* 2005, **15**:214-221.
3. Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, Marcotte EM: **Protein interaction networks from yeast to human.** *Curr Opin Struct Biol* 2004, **14**:292-299.
4. Hartwell LH, Hopfield JJ, Leibler S, Murray AW: **From molecular to modular cell biology.** *Nature* 1999, **402** Suppl:C47-C52.
5. Papin JA, Reed JL, Palsson BO: **Hierarchical thinking in network biology: the unbiased modularization of biochemical networks.** *Trends Biochem Sci* 2004, **29**:641-647.
6. Stelling J, Sauer U, Szallasi Z, Doyle FJ 3rd, Doyle J: **Robustness of cellular functions.** *Cell* 2004, **118**:675-685.
7. Kitano H: **Biological robustness.** *Nat Rev Genet* 2004, **5**:826-837.
8. Kolch W, Calder M, Gilbert D: **When kinases meet mathematics: the systems biology of MAPK signalling.** *FEBS Lett* 2005, **579**:1891-1895.
9. Papin JA, Palsson BO: **The JAK-STAT signaling network in the human B-cell: an extreme signaling pathway analysis.** *Biophys J* 2004, **87**:37-46.
10. Price ND, Reed JL, Palsson BO: **Genome-scale models of microbial cells: evaluating the consequences of constraints.** *Nat Rev Microbiol* 2004, **2**:886-897.
11. Uetz P, Giot L, Cagney G, Mansfield TA, Judson RS, Knight JR, Lockshon D, Narayan V, Srinivasan M, Pochart P *et al.*: **A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*.** *Nature* 2000, **403**:623-627.
12. Harbison CT, Gordon DB, Lee TI, Rinaldi NJ, Macisaac KD, Danford TW, Hannett NM, Tagne JB, Reynolds DB, Yoo J *et al.*: **Transcriptional regulatory code of a eukaryotic genome.** *Nature* 2004, **431**:99-104.
13. Tong AH, Lesage G, Bader GD, Ding H, Xu H, Xin X, Young J, Berriz GF, Brost RL, Chang M *et al.*: **Global mapping of the yeast genetic interaction network.** *Science* 2004, **303**:808-813.
14. Herrgård MJ, Covert MW, Palsson BO: **Reconciling gene expression data with known genome-scale regulatory network structures.** *Genome Res* 2003, **13**:2423-2434.
15. Zhang LV, King OD, Wong SL, Goldberg DS, Tong AH, Lesage G, Andrews B, Bussey H, Boone C, Roth FP: **Motifs, themes and thematic maps of an integrated *S. cerevisiae* interaction network.** *J Biol* 2005, **4**:6.
16. Shen-Orr SS, Milo R, Mangan S, Alon U: **Network motifs in the transcriptional regulation network of *Escherichia coli*.** *Nat Genet* 2002, **31**:64-68.
17. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U: **Network motifs: simple building blocks of complex networks.** *Science* 2002, **298**:824-827.
18. Dobrin R, Beg QK, Barabasi AL, Oltvai ZN: **Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network.** *BMC Bioinformatics* 2004, **5**:10.
19. Tong AH, Evangelista M, Parsons AB, Xu H, Bader GD, Page N, Robinson M, Raghibizadeh S, Hogue CW, Bussey H *et al.*: **Systematic genetic analysis with ordered arrays of yeast deletion mutants.** *Science* 2001, **294**:2364-2368.
20. Wong SL, Zhang LV, Tong AH, Li Z, Goldberg DS, King OD, Lesage G, Vidal M, Andrews B, Bussey H *et al.*: **Combining biological networks to predict genetic interactions.** *Proc Natl Acad Sci USA* 2004, **101**:15682-15687.
21. Ooi SL, Shoemaker DD, Boeke JD: **DNA helicase gene interaction network defined using synthetic lethality analyzed by microarray.** *Nat Genet* 2003, **35**:277-286.
22. Segre D, DeLuna A, Church GM, Kishony R: **Modular epistasis in yeast metabolism.** *Nat Genet* 2005, **37**:77-83.
23. Kelley R, Ideker T: **Systematic interpretation of genetic interactions using protein networks.** *Nat Biotechnol* 2005, **23**:561-566.
24. Kanehisa M, Goto S, Kawashima S, Okuno Y, Hattori M: **The KEGG resource for deciphering the genome.** *Nucleic Acids Res* 2004, **32(Database):D277-D280.**
25. Burgard AP, Nikolaev EV, Schilling CH, Maranas CD: **Flux coupling analysis of genome-scale metabolic network reconstructions.** *Genome Res* 2004, **14**:301-312.
26. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M: **Genomic analysis of regulatory network dynamics reveals large topological changes.** *Nature* 2004, **431**:308-312.
27. de Lichtenberg U, Jensen LJ, Brunak S, Bork P: **Dynamic complex formation during the yeast cell cycle.** *Science* 2005, **307**:724-727.