Properties of Metabolic Networks: Structure versus Function

R. Mahadevan*† and B. O. Palsson†
*Genomatica Inc., San Diego, California 92121; and †Department of Bioengineering, University of California, La Jolla, California 92093

ABSTRACT Biological data from high-throughput technologies describing the network components (genes, proteins, metabolites) and their associated interactions have driven the reconstruction and study of structural (topological) properties of large-scale biological networks. In this article, we address the relation of the functional and structural properties by using extensively experimentally validated genome-scale metabolic network models to compute observable functional states of a microorganism and compare the ''structure versus function'' attributes of metabolic networks. It is observed that, functionally speaking, the essentiality of reactions in a node is not correlated with node connectivity as structural analyses of other biological networks have suggested. These findings are illustrated with the analysis of the genome-scale biochemical networks of three species with distinct modes of metabolism. These results also suggest fundamental differences among different biological networks arising out of their representation and functional constraints.

Received for publication 4 November 2004 and in final form 18 November 2004.

Address reprint requests and inquiries to R. Mahadevan, E-mail: rmahadevan@genomatica.com.

The combination of various high-throughput technologies has enabled the coordinated study of cellular components at the level of genes, proteins, metabolites, and the associated interactions among them. These studies have resulted in the generation of large-scale data sets that have served as the foundation for the reconstruction of metabolic, regulatory, signaling, and protein-protein interaction networks. The availability of these interaction networks has spurred the analysis of the structural, i.e., topological, properties of these networks using quantitative approaches (1). Since functional states of metabolic networks (that correspond to phenotypic functions) can be assessed, we can now compare how important structural properties of networks are when it comes to interpreting their functional states.

Network analysis has suggested that biological networks have two important structural properties. First, it has been shown that several of these networks, including metabolic networks, are scale-free and possess a ''small world'' property (1). Second, scale-free networks are suggested to have high error tolerance (tolerance against random failure) and low attack tolerance (vulnerability to the failure of the highly connected nodes). However, biological networks can have differences in their functional states. In general, the network of interactions among biological entities (genes, proteins, metabolites, etc.) can be classified on the basis of nature of the interaction into two broad categories:

1. Influence networks, where the nature of the interactions are ''influence-based'' such as protein-protein interaction or signaling networks, and the links represent whether an interaction is present or not. This class might be extended to include cases where the type of interaction is important in addition to presence/absence of the interaction (for example, gene regulatory networks where transcription factors can either activate or repress genes expression).
2. Flow networks, where a specific variable such as mass or energy flow may be conserved at each node, such as metabolic networks. Thus, the fundamental properties of biological networks in these two classes can be significantly different.

In this report, we argue that metabolic networks have unique properties resulting from a), the conservation constraints that have to be satisfied at each node, and b), the way the metabolic networks are represented, where nodes are metabolites and the links are reactions that are catalyzed by specific gene products. This representation is different from protein-protein interaction networks, where the nodes are the gene products and the links correspond to interactions. The analysis of protein-protein interaction networks has suggested that the deletion of the most highly connected proteins correlates well with a lethal phenotype (2). In contrast, a node (i.e., a metabolite) in metabolic networks cannot be deleted by genetic techniques, but links can.

These topological properties are derived from network structure, but not from their functional, or phenotypic, states. Recent studies have indicated that metabolic networks have flux distributions with an average path length that is longer than the length obtained from consideration of network structure (3) and that their functional states may not have scale-free characteristics. The proposed error and attack tolerance properties of metabolic networks can be assessed in the context of functional states using flux balance analysis of genome-scale metabolic networks (4). Such in silico models are currently available for several organisms (4). The in silico models of *Escherichia coli* and *Saccharomyces cerevisiae* in particular, have been extensively validated with physiological

doi: 10.1529/biophysj.104.055723

data including data on knockout phenotypes (5, 6). These genome-scale models of metabolic networks have been found to compute observable functional network states (7, 8) and can thus be used to assess systematically the attack and error tolerance of nodes that have a high number of connections relative to those that have a low number of connections.

As other biological networks, metabolic networks have to satisfy many functions simultaneously. To support growth, a genome-scale metabolic network has to produce more than 35 different compounds (9) (Fig. 1, *right panel*). The genome-scale in silico models of *E. coli* and *S. cerevisiae* have been used to analyze the lethality of gene deletion and have been shown to correctly predict the experimental phenotype in 78.7% (6) and 82.6% (5) of the 13,750 and 4,154 cases, respectively. The computed lethal gene knockouts can be used to compare the connectivity, C_i, of node i (a metabolite) and the lethality of the links (reactions) to it. The number of lethal reactions (or connections, $C_{L,i}$) among all the reactions around every metabolite was calculated using data from the in silico deletion analysis (5, 6). The averaged fraction of the lethal reactions ($f_{L,I} = \langle C_{L,i}/C_i \rangle$) was then plotted as a function of the metabolite connectivity (C_i) for the metabolic networks of *E. coli*, *S. cerevisiae*, and *Geobacter sulfurreducens* (Fig. 1, *left panel*).

From these results, it is observed that the lethality fraction (f_L) of some of the less connected metabolites is higher than that of the highly connected metabolites irrespective of the size or the complexity of the metabolic network. In fact, surprisingly, for all of the networks studied, most of the points fall within a narrow range from 0.2–0.5. There are several metabolites with a connectivity of two that have a lethality fraction of 1. These metabolites often occur in a linear pathway, where the end metabolite is an essential biomass component (e.g., amino acid such as histidine). The relation of the lethality fraction to the connectivity is illustrated in Fig. 1, *right panel*, for two representative metabolites. Additionally, one can also investigate the behavior of the metabolic network when a specific metabolite is eliminated from the network by removing all the reactions that consume/produce the metabolite. To realize this experimentally, all the genes corresponding to the reactions would have to be simultaneously deleted. The results of this analysis are depicted in Fig. 2, and it can be immediately seen that these plots show a remarkably different trend compared to Fig. 1, *left panel*. Here, the connectivity correlates well with the average lethality fraction ($\langle f_L^m \rangle$) when all the reactions corresponding to the metabolite have been deleted. Interestingly, even in this case, there appear to be metabolites that are more connected but

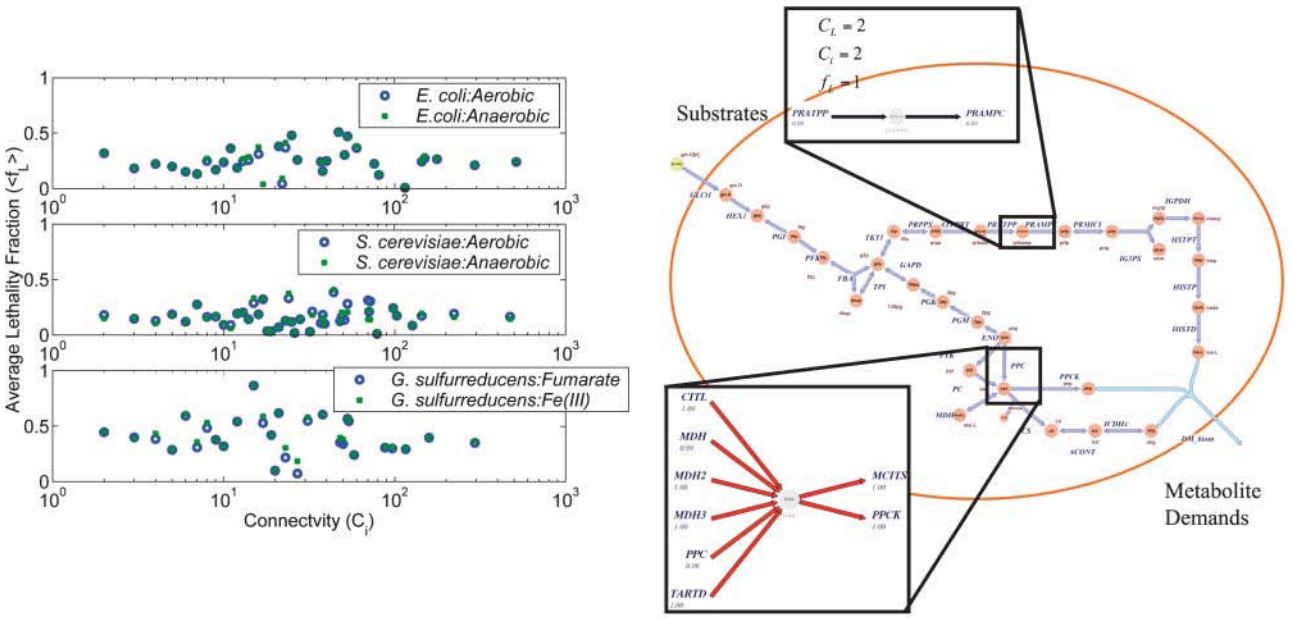

FIGURE 1 (***Left panel***) **Plot of the average lethality fraction ($\langle f_{L,i} \rangle$) as a function of the metabolite connectivity (C_i) for the metabolic networks of *E. coli, S. cerevisiae*, and *G. sulfurreducens* under different growth conditions (see supplementary information). (*Right panel*) The reactions consuming/producing oxaloacetate (oaa, a key metabolite in the TCA cycle) and phosphoribosyl-AMP (prbamp, an intermediate in the histidine biosynthetic pathway) are shown. The reactions predicted to be essential for growth based on the in silico analysis are shown in black, whereas the nonessential reactions are shown in red along with the normalized predicted growth rate, the connectivity (C_i), and the lethality fraction ($f_{L,i}$) for both the metabolites. The number of lethal reactions around a highly connected metabolite such as oxaloacetate (oaa, $C_i = 10$) is shown in comparison to a lowly connected metabolite such as phosphoribosyl-AMP (prbamp, $C_i = 2$). However, the number of lethal reactions for both metabolites is 2, as prbamp occurs in a linear pathway in histidine biosynthesis and the deletion of either one of the linked reactions leads to the loss of network function (growth).**

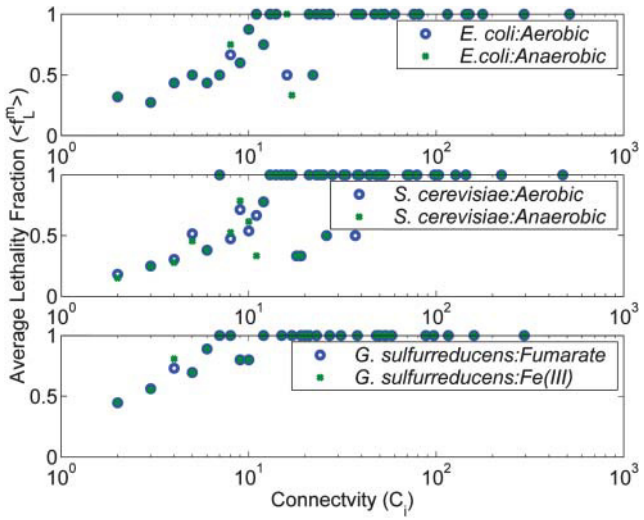

FIGURE 2 Plot of the average lethality fraction when all reactions corresponding to a metabolite are removed ($\langle f^{m}_{L,i} \rangle$) as a function of the metabolite connectivity (C_i).

with lower average lethality fraction. These findings illustrate some of the unique properties of metabolite networks in part due to their representation as a network and partly due to the conservation constraints these networks have to satisfy.

This analysis based on network functions indicates that even the least connected nodes in genome-scale metabolic networks are just as likely to be critical to the overall network functions as the most highly connected nodes. The inactivation of even lowly connected nodes that disrupts the specific function of the subsystem could lead to the failure of the overall network (lethality). These findings indicate that, in addition to network structure, the functional states of the metabolic pathways have to be considered for the study of network properties and how they relate to observable biological functions. Interestingly, studies of the transcriptional regulatory networks have shown similar properties where the highly connected nodes do not correlate well with essentiality (10). Taken together, these studies seem to suggest that even though networks of biological entities might have some similar properties, there appear to be fundamental differences in the nature of these networks, investigation of which can lead to valuable insights on their functions.

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.

ACKNOWLEDGMENTS

The authors acknowledge Professor Barabasi at the University of Notre Dame for comments on this article.

REFERENCES and FOOTNOTES

(1) Barabasi, A. L., and Z. N. Oltvai. 2004. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5: 101–113.

(2) Jeong, H., S. P. Mason, A. L. Barabasi, and Z. N. Oltvai. 2001. Lethality and centrality in protein networks. *Nature.* 411:41–42.

(3) Arita, M. 2004. The metabolic world of *Escherichia coli* is not small. *Proc. Natl. Acad. Sci. USA.* 101:1543–1547.

(4) Kauffman, K. J., P. Prakash, and J. S. Edwards. 2003. Advances in flux balance analysis. *Curr. Opin. Biotechnol.* 14:491–496.

(5) Duarte, N. C., M. J. Herrgard, and B. Palsson. Reconstruction and validation of *Saccharomyces cerevisiae* iND750, a fully compartmentalized genome-scale metabolic model. 2004. *Genome Res.* 14:1298–1309.

(6) Covert, M. W., E. M. Knight, J. L. Reed, M. J. Herrgard, and B. O. Palsson. Integrating high-throughput and computational data elucidates bacterial networks. 2004. *Nature.* 429:92–96.

(7) Edwards, J. S., R. U. Ibarra, and B. O. Palsson. 2001. In silico predictions of *Escherichia coli* metabolic capabilities are consistent with experimental data. *Nat. Biotechnol.* 19:125–130.

(8) Ibarra, R. U., J. S. Edwards, and B. O. Palsson. 2002. *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. *Nature.* 420:186–189.

(9) Neidhardt, F., J. L. Ingraham, and M. Shaechter. 1990 Physiology of the Bacterial Cell. Sinauer Associates, Sunderland, MA.

(10) Yu, H., D. Greenbaum, L. H. Xin, X. Zhu, and M. Gerstein. 2004. Genomic analysis of essentiality within protein networks. *Trends Genet.* 20:227–231.