FEBS Letters 579 (2005) 4642–4646 FEBS 29829

Functional essentiality from topology features in metabolic networks: A case study in yeast

Maria Concetta Palumbo[a], Alfredo Colosimo[a], Alessandro Giuliani[b,*], Lorenzo Farina[c]

[a] *Physiology and Pharmacology Department, University of Rome, 'La Sapienza', Rome, Italy*
[b] *Istituto Superiore di Sanità, Environment and Health Department, Viale Regina Elena 299, 00161 Rome, Italy*
[c] *Department of Computer and Systems Science "A. Ruberti", University of Rome, 'La Sapienza', Rome, Italy*

Received 30 May 2005; revised 16 July 2005; accepted 18 July 2005

Available online 10 August 2005

Edited by Robert B. Russell

Abstract The relation between the position of mutations in *Saccharomyces cerevisiae* metabolic network and their lethality is the subject of this work. We represent the topology of the network by a directed graph: nodes are metabolites and arcs represent the reactions; a mutation corresponds to the removal of all the arcs referring to the deleted enzyme. Using publicly available knock-out data, we show that lethality corresponds to the lack of alternative paths in the perturbed network linking the nodes affected by the enzyme deletion. Such feature is at the basis of the recently recognized importance of 'marginal' arcs of metabolic networks.

Keywords: Systems biology; Lethality; Phenotype/genotype relations

1. Introduction

Information regarding genomes and metabolic interactions are stored in many databases and scientists have the urgent need for new tools in order to find effective representations and models for such interaction networks. Only a multidisciplinary approach may lead to significant results, which implies the merging of biological knowledge and methodologies coming from many and diverse fields such as mathematics, statistics, control theory, computer science, etc.

In the present work, we adopted a simple heuristic principle in use for the study of failures in power supply networks: the effects of a damage (in the form of the elimination of a pre-existent link) on the entire network are related to the fact that elimination of the arc isolates one or more node from the rest of the network. That is to say that failures with greater consequences on the entire system are those for which it is not possible to devise 'alternative pathways' for the nodes affected by the arc(s) elimination. This simple principle is a very old 'topic' of biochemists when considering the effect of a step elimination in biochemical networks represented as directed graph having metabolites as nodes and reactions as arcs. The principle implies that essential mutations, corresponding to the enzymes whose elimination ends up in the death of the organism, are seldom located in highly connected portions of the graph, since these positions will probably have a lot of alternative pathways for reaching the same nodes. On the contrary, the essential mutations will tend to have a peripheral position in the network so to maximize the possibility to provoke the isolation of a previously connected portion of the graph.

This conjecture was proved by the analysis of the *Saccharomyces Cerevisiae* mutation databases and its implications for both systems biology and the development of new targets for drugs are investigated.

2. Materials and methods

An enzyme is defined as "essential" if the deletion of the gene coding for that enzyme has lethal effects for the organism under a given experimental condition. In the case of the yeast, data about the knock-outs are available for the entire genome in the databases of the Stanford' University Saccharomyces Genome Deletion Project, publicly available in the web site http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3. In order to obtain the data regarding the surviving ability of the yeast after a gene deletion, we referred to both the Stanford' collection and Jeong and colleagues [1]. The yeast metabolic network was built according to Ma and Zeng [2,3]: the data come from the KEGG database (http://www.genome.jp/kegg/), possible irreversibility of reactions is taken into account and recurrent metabolites and cofactors (H_2O, NADH, ATP, ADP, etc.) are deleted so to make topological analysis more realistic. Each node (metabolite) is labeled by a number (see Supplementary Material 1) according to the authors terminology [2,3]. The obtained global yeast metabolic network having metabolites as nodes and reactions as arcs was then transformed into an adjacency matrix. This is a binary square matrix (Fig. 1) having as rows and columns the intervening metabolites and taking 1/0 values in row i and column j depending on the presence/absence of an arc connecting node i to node j. In the case of metabolic networks, an arc corresponds to an enzyme catalysing a chemical reaction transforming a metabolite into another. The adjacency matrix formalism is very useful in describing many network structures and the literature adopting this notation for describing organic molecules is particularly rich. In this case an arc represents a chemical bond between two atoms and the matrix is isomorphic to the structural formula of the molecule [4,5].

As a second step, the KEGG nomenclature for enzymes was transformed into the corresponding gene nomenclature so to look for the essential character of the relative knock-out in the yeast genome database.

When the arcs corresponding to the reactions catalyzed by the knocked-out enzyme are deleted, the topology of the network was checked as for the isolation of a previously connected node, i.e. for the loosing of the possibility to reach one or more nodes previously connected by a pathway.

It is worth noting that the specific enzyme is considered as essential, and consequently the correspondent arcs are deleted, if at least the 75% of genes coding for the particular enzyme were demonstrated to be

*Corresponding author. Fax: +39 064 9902355.
E-mail address: alessandro.giuliani@iss.it (A. Giuliani).

0014-5793/$30.00
doi:10.1016/j.febslet.2005.07.033

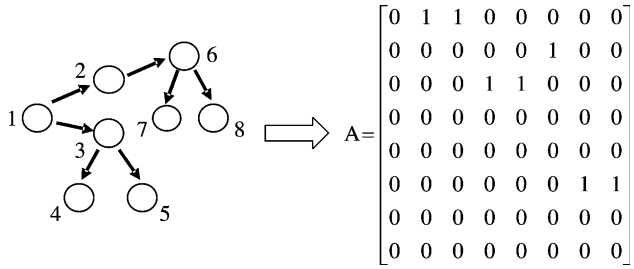

Fig. 1. Networks and Matrices. The figure reports pictorially the isomorphism between network structures and adjacency matrices.

lethal in the knocked-out organism. If a specific reaction is catalysed by multiple enzymes, the corresponding arc in the graph is deleted if the above condition is fulfilled for all the enzymes involved. These very stringent conditions allowed us to select only 37 essential enzymes over a total of 412 (reported in Supplementary Materials 2, 3).

3. Results

A metabolic network can be imagined as composed by four distinct classes of nodes (Fig. 2). The first one is the strong connected components (SCC), made of nodes each other linked by directed paths; the 'reactant subset' consists of metabolites entering the system as reactants and called 'sources' because they can reach the SCC, but cannot be reached from it. The third class is represented by the 'product subset', made of metabolites that are accessible from the SCC, but don't have any connections to it. Such metabolites are positioned at the end of a given pathway and thus exiting the system as products (sinks). The metabolites in the last class, called 'isolated subset', are inserted in autonomous pathways with respect to SCCs.

This connectivity structure can be found even in the architecture of scale-free networks such as Word Wide Web, the electric power transmission grid or airport connections [6–10].

In our formalization a path going from i to j node is intended as a sequence of vertices connected by arcs corresponding to the order of the vertices in the sequence, in which no repetition of vertices is allowed. In the case of yeast we counted 4 distinct SCCs with more than 6 nodes (as a matter of fact any single isolated node represents a 'trivial' SCC): the biggest one is made up of 110 metabolites and is called the 'giant set'; the second most populated SCC has 8 metabolites; the third, 7, and the last SCC is made of 6 nodes.

In Fig. 3 the most populated set of nodes in the yeast metabolic network is shown: nodes coloured in blue, light blue, violet and green belong to the SCCs, while red nodes belong to the reactant, product and isolated subsets.

The pathways redundancy of the SCCs, especially of the giant set, makes these components very robust to failure or removal of reactions by gene knock-out or mutations.

The main effects of the deletion of a gene coding for an essential enzyme can be briefly grouped into the following categories:

1. Preventing the connection between the nodes at the ends of the eliminated arc: no other path is available to restore the previous connection (37/37 essential enzymes).
2. Isolating and separating a pathway from the rest of the network (29/37 essential enzymes).
3. Interrupting the full connectivity in a strong connected component (7/37 essential enzymes, corresponding to 4/37 in the giant set and 3/37 in the minor SCCs).
4. Creating a new smaller SCC from another one (1/37 essential enzymes).
5. Disrupting of a cluster of nodes in the IS compartment (16/37 essential enzymes).

As it is evident from the above points, the main tenet of our work, corresponding to the 'biochemist common sense' is fulfilled: the whole set of lethal mutations refer to paths that cannot be restored by the remaining arcs (category 1). At the website http://www.dis.uniroma1.it/~farina/Yeast/, the different portions of the network where the essential enzyme are located are enlarged and reported as eps figures.

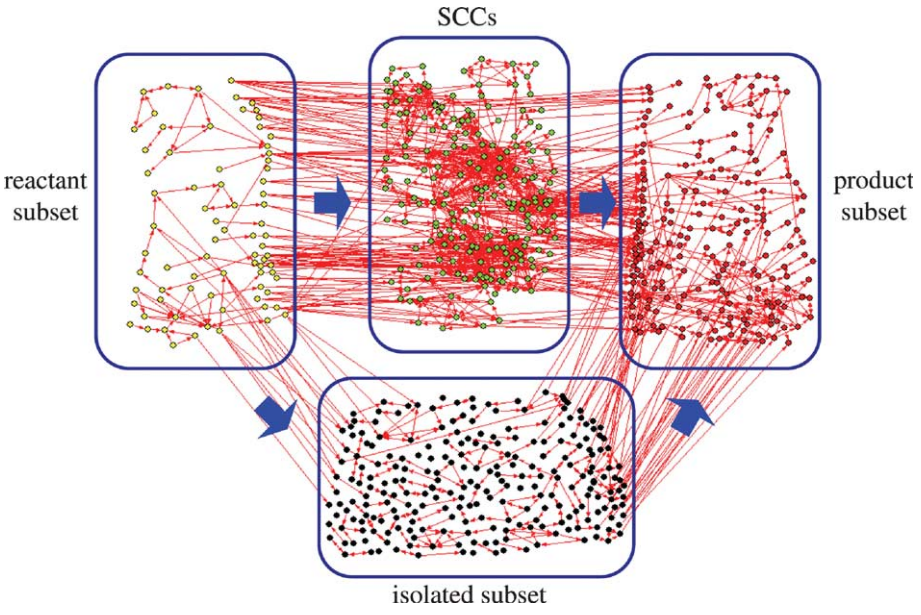

Fig. 2. General representation of a metabolic network, adapted from [18]. On the top, the SCC module corresponds to strongly connected components; on the left there is the 'reactant' subset (sources); on the right, the 'product' subset (sinks); the 'isolated' subset is on the bottom.

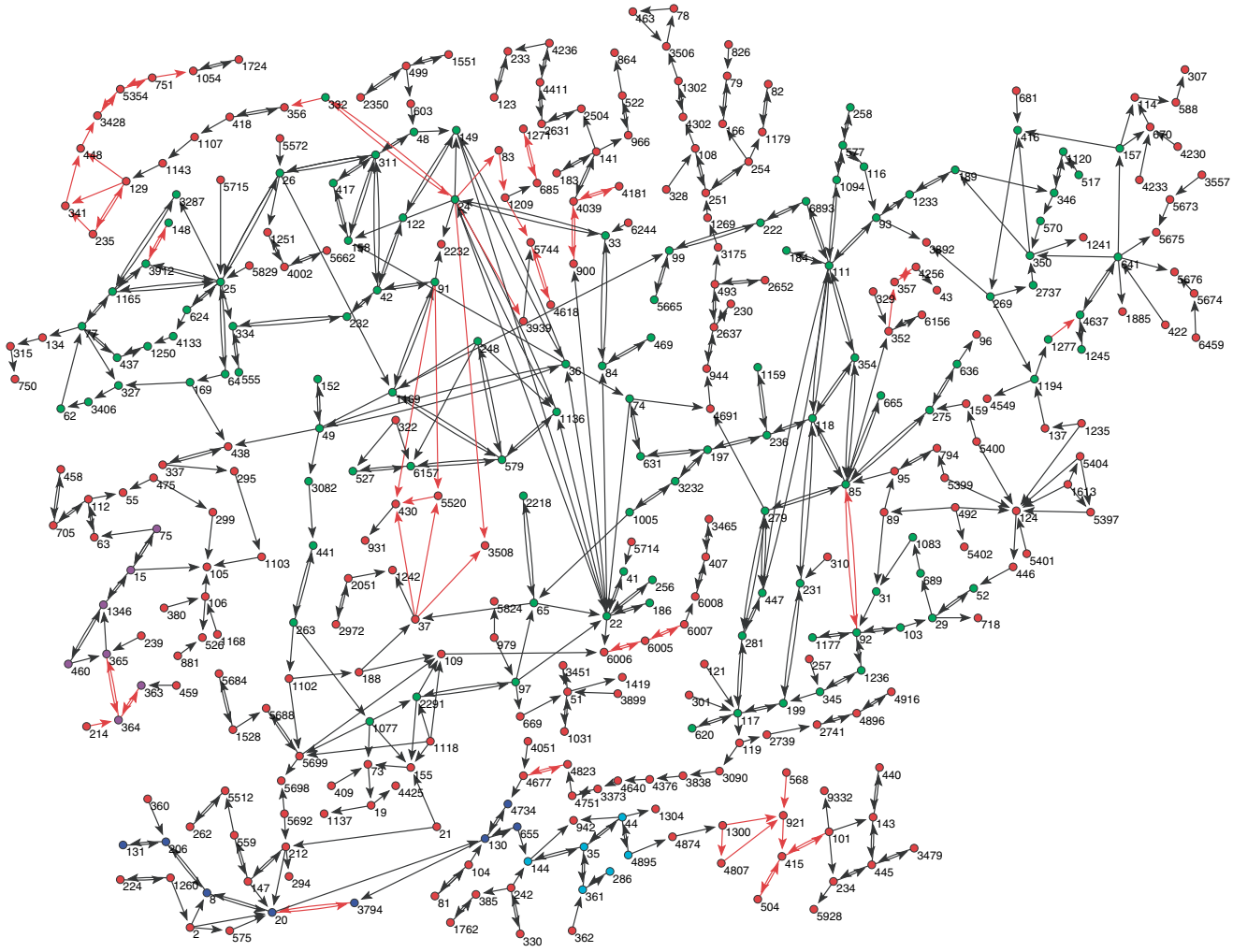

Fig. 3. Nodes coloured in blue, light blue, violet and green belong to the SCCs while red nodes pertain to 'reactant', 'product' and 'isolated' subsets. Red arrows show the location of essential enzymes. For a better view, see the eps figure in the website http://www.dis.uniroma1.it/~farina/Yeast/, the excel files reporting the information regarding the figure are enclosed.

In Fig. 4 a very well known metabolic pathway (fatty acid biosynthesis) is reported in both a biochemical meaningful and in the synthetic graph formalization. The location of essential enzymes is marked by a red arrow in the graph diagram (which in turn also reports some nearby elements connecting the pathway to other portions of the network). It is evident that the elimination of the essential arcs isolates some nodes. The correspondence between the two representations can be inferred from Table 1.

Going more in detail, we note how only a minority of the essential mutations refer to arcs inside SCCs (see Fig. 3), thus confirming the conjecture of the greater robustness of such components. Moreover the SCCs related essential mutations are located at the periphery of the respective components and, in any case, isolate the affected nodes from the SCC.

The general message we can derive from our results is the most crucial points in the metabolic networks corresponds to enzymes at the periphery of the biological system, given the elimination of the corresponding arcs are more prone to disconnect some metabolites from the general balance without the buffering provided by the presence of alternative paths.

4. Discussion

The results of our work are consistent with the recently discovered relevance of the so called non-hub-connectors [11]. These enzymes are located at the periphery of the metabolic modules and connect different functionalities into a coordinated whole. This is in somehow reminding the structure of power supply networks, where the most crucial electrical lines are those transferring electricity from one strongly connected cluster to another. It is important to stress the fact that our representation of the metabolic network is the classical one having the metabolites as nodes and the enzymes as arcs. Thus the modifications of the network correspond to the elimination of a connection and not of a specific element (node). At odds with this formalization, the systems biology approaches based on protein networks are based on the elimination of specific nodes: in these representations [12–14] the proteins are the nodes and the edges correspond to the presence of an interaction between the two intervening proteins.

In the protein–protein interaction case, the central position of a network element is not a signature of 'failure resistance',

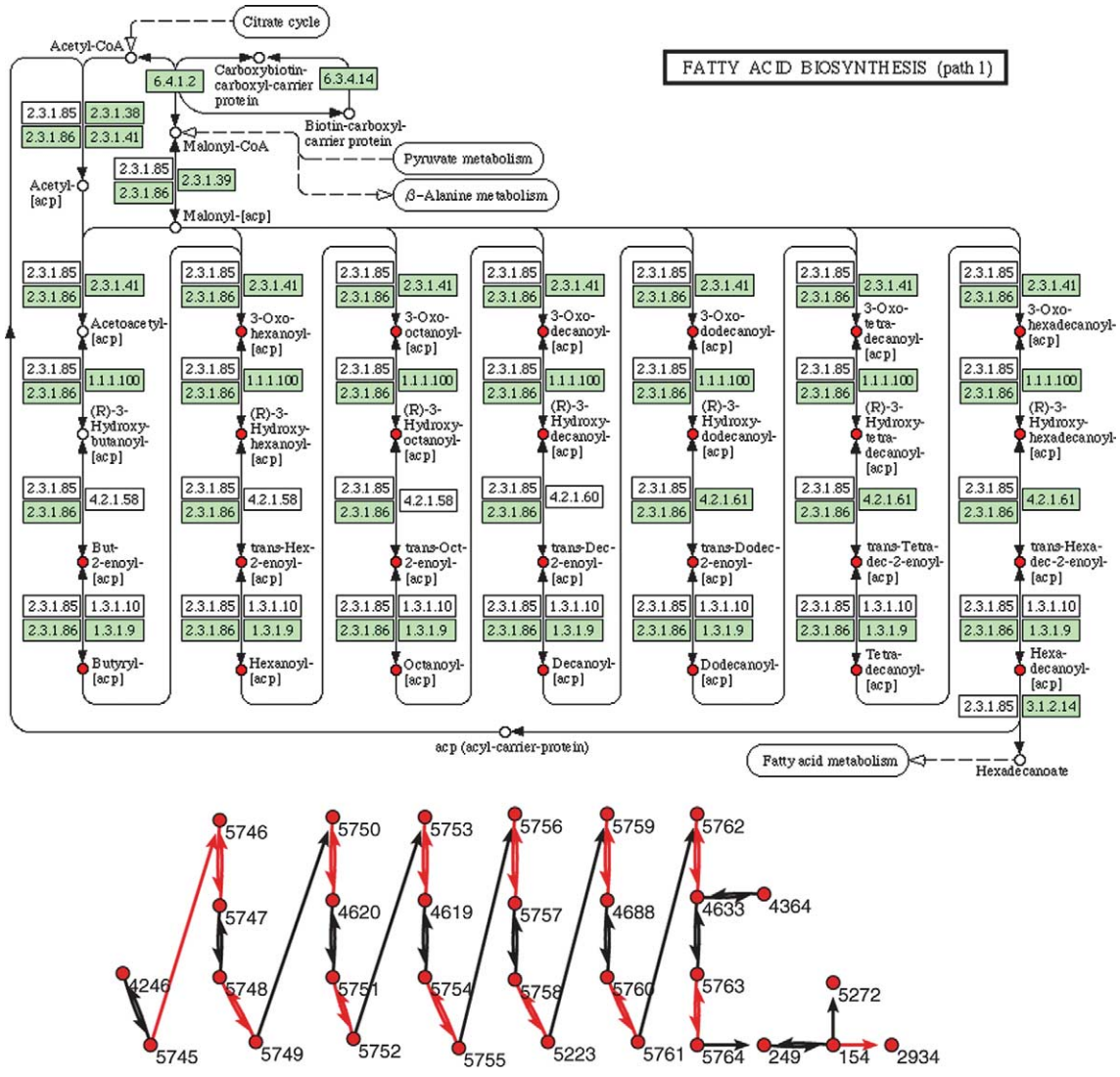

Fig. 4. The figure reports the fatty acid biosynthesis in both biochemical (upper panel) and graph (lower panel) ways of representations. The first one comes from KEGG database, the second one is obtained using Pajek software on Ma and Zeng [2] database. Table 1 allows for the conversion of the upper to the lower panel. It is worth noting only green labelled enzymes are present in yeast. The lower panel includes some other peripherical arcs and nodes linking fatty acid biosynthesis to the whole network.

like in our metabolic networks, but it is related to the amount of damage provoked by its elimination. Intuitively this comes from the fact that eliminating a strongly connected element interferes with a greater number of relations with respect to the elimination of a periphery element.

The possibility to define the essential character of a reaction on the pure basis of its topological position in the metabolic network points toward the presence of a high level of integration in the system. The proof of the presence of a specific 'signature' of essential mutations not referring to the specific biochemical role played by the corresponding enzyme but only to the presence of a specific connected path points to the presence of an 'holistic' level of the metabolic network functioning as a whole. Recently Samal and colleagues [15] obtained very similar results in the case of *E. coli* metabolic network. The authors demonstrated that the essential enzymes are preferentially located in very small (2 or 3 reactions) modules of metabolic network, thus confirming the hypothesis that elements outside giant component and with a low degree of connectivity are the most crucial ones for the entire system functioning. They explain this behaviour with the fact low degree metabolites are a factor of stability of the network because, for example, a metabolite that can be produced in only one reaction and consumed in only one causes both reactions to have correlated fluxes in any steady state. We enlarged this perspective by the demonstration that even metabolites involved in more complex pathways with more than one or few reactions can be inserted in the same picture by the notion of essentiality corresponding to the non-reachability of one or more metabolites of the network. Moreover, we based our analysis on an eukariotic system, while Samal et al. [15] worked on a bacterium: the relative importance of operon-like regulatory scheme is much more prominent in prokariots than in eukariots, thus the obtaining of similar results in yeast, enlarges the validity of the results well beyond the 'operon-like' regulation hypothesis.

Table 1
The table reports the conversion key from the biochemical to graph representations and viceversa

Node number	Metabolite name
4246	**But-2-enoyl-[acp]**
5745	**Butyryl-[acp]**
5746	**3-Oxohexanoyl-[acp]**
5747	**(R)-3-Hydroxyhexanoyl-[acp]**
5748	**trans-Hex-2-enoyl-[acp]**
5749	**Hexanoyl-[acp]**
5750	**3-Oxooctanoyl-[acp]**
4620	**(3R)-3-Hydroxyoctanoyl-[acyl-carrier protein]**
5751	**trans-Oct-2-enoyl-[acp]**
5752	**Octanoyl-[acp]**
5753	**3-Oxodecanoyl-[acp]**
4619	**(3R)-3-Hydroxydecanoyl-[acyl-carrier protein]**
5754	**trans-Dec-2-enoyl-[acp]**
5755	**Decanoyl-[acp]**
5756	**3-Oxododecanoyl-[acp]**
5757	**(R)-3-Hydroxydodecanoyl-[acp]**
5758	**trans-Dodec-2-enoyl-[acp]**
5223	**Dodecanoyl-[acyl-carrier protein]**
5759	**3-Oxotetradecanoyl-[acp]**
4688	**(3R)-3-Hydroxytetradecanoyl-[acyl-carrier protein]**
5760	**trans-Tetradec-2-enoyl-[acp]**
5761	**Tetradecanoyl-[acp]**
5762	**3-Oxohexadecanoyl-[acp]**
4633	**(3R)-3-Hydroxypalmitoyl-[acyl-carrier protein]**
4364	2-Hexadecenoyl-[acyl-carrier protein]
5763	**trans-Hexadec-2-enoyl-[acp]**
5764	**Hexadecanoyl-[acp]**
249	Palmitate
154	Palmitoyl-CoA
5272	trans-Hexadec-2-enoyl-CoA
2934	3-Dehydrosphinganine

The bolded elements are the ones included in the biochemical pathway.

Obviously we only referred to the static representation of the yeast metabolic network, without specific reference to kinetic data (not available at the whole network level). This allows us only a quite limited picture of the system. Nevertheless even this forcedly limited, purely topological view was demonstrated to be able to derive general biological predictions.

The above results allow for trying and develop a rational strategy for drug development [16] on the basis of the recognition of the possibly most crucial elements of a metabolic system and trying and derive quantitative genotype/phenotype correlations on rational bases. On a more theoretical ground, the above analysis can be intended as a step toward a quantitative approach to the so called pleiotropy effect [17], i.e. the phenotypic effect of the mutation of a specific gene on a multiplicity of different traits.

Acknowledgments: We wish to thank an anonymous reviewer for having driven our attention to the Samal et al. pre-print.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2005.07.033.

References

[1] Jeong, H., Oltvai, Z.N. and Barabasi, A.-L. (2003) Prediction of protein essentiality based on genomic data. ComPlexUs 1, 19–28.
[2] Ma, H.W. and Zeng, A.P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. Bioinformatics 19, 270–277.
[3] Ma, H.W. and Zeng, A.P. (2003) The connectivity structure, giant strong component and centrality of metabolic networks. Bioinformatics 19, 1423–1430.
[4] Lukovits, I. (2000) A compact form of the adjacency matrix. J. Chem. Inf. Comput. Sci. 40, 1147–1150.
[5] Benigni, R., Passerini, L., Pino, A. and Giuliani, A. (1999) The information content of the eigenvalues from modified adjacency matrices: large scale and small scale correlations. QSAR 15, 449–455.
[6] Albert, R., Jeong, H. and Barabasi, A.-L. (2000) Error and attack tolerance of complex networks. Nature 406, 378–382.
[7] Crucitti, P., Latora, V. and Marchiori, M.A. (2004) Topological of the Italian electric power grid. Physica A 338, 92–97.
[8] Crucitti, P., Latora, V., Marchiori, M.A. and Rapisarda, A. (2004) Error and attack tolerance of complex networks. Physica A 340, 388–394.
[9] Crucitti, P., Latora, V., Marchiori, M.A. and Rapisarda, A. (2003) Efficiency of scale-free networks: error and attack tolerance. Physica A 320, 622–642.
[10] Roy, S., Asavathiratham, C., Lesieutre, B.C., and Verghese, G.C. (2001) Network models: growth, dynamics and failure. Proceedings of the 34th Hawaii International Conference in System Sciences.
[11] Guimera, R. and Amaral, L.N. (2005) Functional cartography of complex metabolic networks. Nature 433, 895–900.
[12] Jeong, H., Mason, S.P., Barabasi, A.-L. and Oltvai, Z.N. (2001) Lethality and centrality in protein networks. Nature 411, 41–42.
[13] Fraser, H.B., Hirsh, A.E., Steinmetz, L.M., Scharfe, C. and Feldman, M.W. (2002) Evolutionary rate in the protein interaction network. Science 296, 750–752.
[14] Yu, H., Greenbaum, D., Lu, H.X., Zhu, X. and Gerstein, M. (2004) Genomic analysis of essentiality within protein networks. Trends Genet. 20, 227–231.
[15] Samal, A., Shalini, G., Giri, V., Krishna, S., Raghuram, N., and Jain, S. (2005) Low degree metabolites enhance modularity in biological networks and explain essential metabolic reactions. Available from: http://arXiv.org/abs/q-bio.MN/0504026.
[16] Csermely, P., Agoston, V. and Pongor, S. (2005) The efficiency of multi-target drugs: the network approach might help drug design. Trends Pharmacol. Sci. 26, 178–182.
[17] Foster, K., Shaulsky, G., Strassmann, J.E., Queller, D.C. and Thompson, C.R. (2004) Pleiotropy as a mechanism to stabilize cooperation. Nature 43, 693–696.
[18] Ma, H.W., Zhao, X.M., Yuan, Y.J. and Zeng, A.P. (2004) Decomposition of metabolic network into functional modules based on the global connectivity structure of reaction graph. Bioinformatics 20, 1870–1876.