Appl Biochem Biotechnol (2010) 160:222–231
DOI 10.1007/s12010-009-8572-5

Inferring Minimal Feasible Metabolic Networks of *Escherichia coli*

Da Jiang · Shuigeng Zhou · Hui Liu · Yi-Ping Phoebe Chen

Received: 23 October 2008 / Accepted: 16 February 2009 /
Published online: 27 May 2009

Abstract Since the organism contains many redundant reactions, the minimal feasible metabolic network that contains the basic growth function is not the collection of reactions that associate the essential genes. To identify minimal metabolic reaction set is a challenging work in theoretical approach. A new method is presented here to identify the smallest required reaction set of growth-sustaining metabolic networks. The content and number of the minimal reactions for growth are variable in different random processes. Though the different carbon sources also vary the content of the reactions in the minimal metabolic networks, most essential reactions locate in the same metabolic subsystems, such as cofactor and prosthetic group biosynthesis, cell envelope biosynthesis, and membrane lipid metabolism.

Keywords Metabolic networks · Minimal reaction sets · Structure analysis

Introduction

The process that generates mass, energy, information transfer, and cell fate specification is the result of integration of various reactions in a cell or microorganism [1], which is named as the metabolic network. This network is one of the most challenging biological networks [2] and is the most potential one for immediate applicability [3]. One fundamental problem regarding the capabilities of metabolic network is whether or not it contains the reactions necessary to form a product from a given set of media substrates [4].

D. Jiang · S. Zhou · H. Liu
Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China

D. Jiang · S. Zhou · H. Liu
School of Computer Science, Fudan University, Shanghai, China

Y.-P. P. Chen (✉)
Faculty of Science and Technology, Deakin University, Melbourne, Australia
e-mail: phoebe@deakin.edu.au

Recently conducted research focus on finding the minimal metabolic network by flux balance analysis (FBA) and its extensions [5, 6]. FBA is a milestone to analyze genome-scale wild-type metabolic networks, which considers the prior metabolites and reactions and analyzes the flux of each reaction [7, 8]. The MILP-FBA model [5] analyzed the *Escherichia coli* metabolic network iJE660 [9] which incorporates 454 metabolites and 720 reactions. It has determined that at least 224 reactions out of 720 are required for growth on glucose, which made use of the discrete modeling approach to identify the smallest required sets of growth-sustaining metabolic reactions. The number is at least 122 reactions under the rich medium condition. The metabolic network iJE660 of *E. coli* is simple, and hence, the scientists added or modified some important reactions [10]. Then, a regenerated metabolic network iJR904 was presented, which contains 762 metabolites and 932 reactions. By using the FBA method to repeatedly simulate the successive loss of reactions (genes) under the rich carbon source, the minimal reaction networks contain around 245 reactions [6]. These researches are based on the hypothesis that the organism grows optimally to transform more biomass from the carbon source and some cofactor. Though the use of this simple optimization criterion has wide applications [11], other researches pointed that the organism is not always an optimization of growth [12–15].

In this paper, we propose a topology analysis method to identify the smallest required reaction sets of growth-sustaining metabolic networks under given nutrition. We also investigate the functional classification of essential reactions among all the minimal metabolic networks.

Materials and Methods

Metabolic Reactions The metabolic reactions are obtained from published *E. coli* model iJR904 which contains 762 metabolites and 932 reactions [10]. This model is more complete and a chemically accurate description of *E. coli* than the iJE660a [9]. It is more important that iJR904 gives the gene-to-reaction relationship, which can infer gene essentiality by reaction essentiality. To validate the essential genes predicted by our method, we can refer the profiling of the *E. coli* chromosome (PEC database, http://www.shigen.nig.ac.jp/ecoli/pec/) maintained by the Genetic Resource Committee of Japan.

Original Seed It is well known that the organism needs some inorganic and organic substances to survive. The reconstructed metabolic networks of *E. coli* just give the nutrition environment [10]. One hundred twenty-nine organic compounds taken as rich nutrition environment are also taken from the *E. coli* model iJR904 [10]. Because most substrates with high degree of incoming and outgoing links are complex compounds but clear reaction process to synthesize is absent, we select these substrates as preexisted inner metabolites according to the previous analysis results [1]. These assistant substrates which are important for the whole metabolic networks are as follows: H_2O, ADP, Pi, ATP, NADP, NADPH, PP, NAD, NADH, COA, and ACP. Because the current metabolic networks are still not complete, some metabolites are interdependent, which bring a “closed cycle.” In this cycle, two compounds cannot be produced by other reactions but can be produced by each other. However, their by-products are important for other reactions. Absence of the by-products may influence other reactions, which leads ultimately to no biomass production. In order to avoid this case, we add these two-cycle interdependent metabolites to the inner metabolites set, considering that these metabolites do not impact on the whole networks. We identify eight pairs of two-metabolite cycles which are interdependent and cannot be

broken by other reactions. These are as follows: mqn8↔mql8, no2↔no3, gdp↔gtp, q8h2↔q8, 2dmmq8↔2dmmql8, trdox↔trdrd, fadh2↔fad, and nadp↔nadph. Most of these in the interdependent cycles are complex metabolites. Table 1 gives the detail of the inner metabolites. Table 2 gives the detail of the inorganic environment.

The Target Reaction The target reaction is important for simulating the minimal reaction set. Here, we set the biomass reaction as the target, which presents that the organism needs lots of compounds to grow. Here, we focus on the qualitative analysis of topology structure which does not need the coefficients of the metabolites. The biomass equation is also obtained from the *E. coli* model iJR904 [10].

Method Our method is to find the minimal reaction sets for sustaining organism growth. The whole process separates into three steps: random growth, reductive evolution, and repeating the above two steps for summarizing the three sets. Firstly, we take the inorganic compounds, the organic compounds, and the inner metabolites as the seed set. Randomly select a reaction to test if it can exist based on the seed set. If the substrates of the selected reaction have existed in the seed set, then this reaction should be added to the random growth reaction set and the products of this reaction should be added to the seed set. Repeat this process until the seed set contains the target compounds (the components of the biomass). However, the reactions in the random growth set are not the minimal reactions to

Table 1 Inner metabolites.

Abbreviation	Metabolite name
Mqn8	Menaquinone 8
Mql8	Menaquinol 8
no2	Nitrite
no3	Nitrate
gdp	GDP
gtp	GTP
q8h2	Ubiquinol-8
q8	Ubiquinone-8
2dmmq8	2-Demethylmenaquinone 8
2dmmql8	2-Demethylmenaquinol 8
trdox	Oxidized thioredoxin
trdrd	Reduced thioredoxin
fadh2	FADH2
fad	FAD
nadp	NADP
nadph	NADPH
nad	NAD
nadh	NADH
adp	ADP
atp	ATP
coa	Coenzyme A
acp	Acyl carrier protein
ppi	Diphosphate

Table 2 Inorganic environment.

Abbreviation	Metabolite name
CO_2	Carbon dioxide
Fe^{2+}	Iron(II)
H^+	Protons
H_2O	Water
K^+	Potassium
Na^+	Sodium
NH_4	Ammonia
O_2	Oxygen
PI	Phosphate
SO_4	Sulfate

produce biomass. Secondly, we further test whether every reaction in random growth set is necessary for producing the biomass. Randomly delete a reaction from the random growth set to test if the remains can still produce the biomass. If not, restore the reaction to the random growth set. If yes, select another reaction to test. Repeat these processes until reactions can be deleted. Due to the alternate pathways, there are a lot of minimal reaction sets. Thirdly, summarize three sets of the whole metabolic networks: (1) the fundamental reaction set that all reactions are always necessary for producing biomass; (2) the alternative reaction set that these reactions are necessary sometimes; (3) the dispensable reaction set that these reactions are not necessary. We repeat the above two steps with enough number of times until the sizes of the three sets are not changed at all.

Results

The Minimal Metabolic Reaction Sets

The minimal networks are variable in reaction content and number. However, the minimal lethal reaction set is invariable on the condition of the same organic environment. When

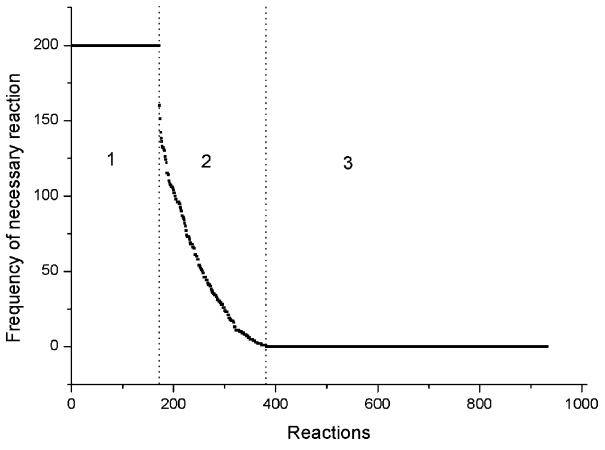

Fig. 1 The frequency of necessary reactions for the minimal reaction sets under the condition of glucose-only environment. It shows that the whole reaction set can be broken into three subsets, which are (1) the fundamental reaction set, (2) the alternative reaction set, and (3) the dispensable reaction set

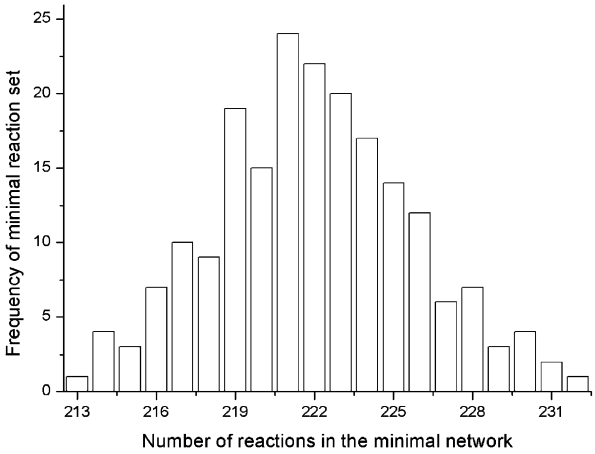

Fig. 2 The frequency of minimal metabolic reaction sets for minimal environment

setting the single organic compound D-glucose as the minimal nutrition, 172 reactions constitute the fundamental reaction set in which every reaction is essential for sustaining growth; 189 reactions constitute the alternative reaction set, and the residual belongs to the dispensable reaction set (Fig. 1). The fundamental reaction set cannot completely constitute the minimal reaction set for sustaining organism growth because some reactions in the alternative reaction set are also important under some conditions. There are alternative pathways to the same product that remove either of the reactions which makes the other essential for survival [16]. And if all of these alternative pathways to the same product are deleted, it will jeopardize the organism surviving. The number of growth-sustained reactions is between 213 and 232 (Fig. 2). Because the simulation of the minimal reaction set is based on a random approach, we need to repeatedly process a lot to investigate the range of the sizes of the minimal metabolic networks sets. Some reactions always exist in the minimal metabolic networks sets, but others do not. In order to testify that repeating 200 times is enough to give the accurate three sets, we record the number of common reactions and the number of all reactions used in all the minimal metabolic networks. Figure 3 shows that the number of common reactions used in each minimal metabolic network descends quickly before the first ten iterative times, and it does not descend no matter how many the

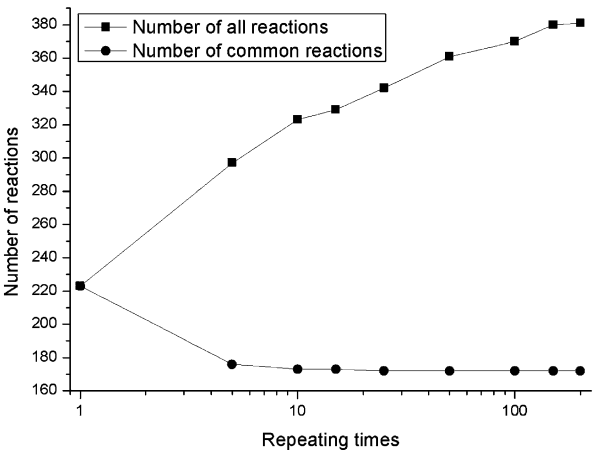

Fig. 3 The common reactions and the total reactions used in all minimal reaction sets

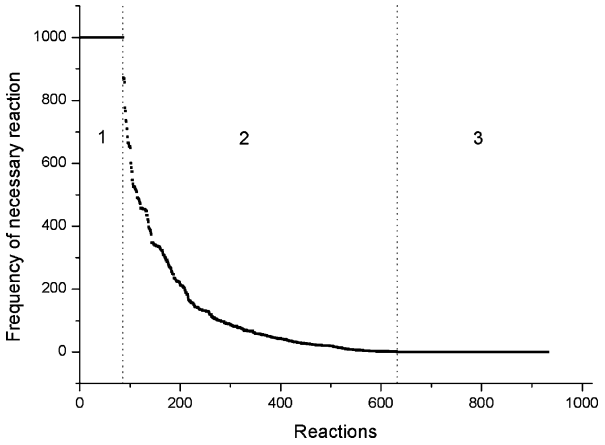

Fig. 4 The frequency of necessary reactions for the minimal reaction sets under the condition of rich nutrition environment. The whole reaction set also can be divided into three subsets. which are (1) the fundamental reaction set, (2) the alternative reaction set, and (3) the dispensable reaction set. The alternative reaction set is bigger than other sets

iterative times are. Figure 3 also shows that the number of reactions used in all minimal reaction sets ascends fast in the first 20 iterative times and then increases slowly, and when the iterative time exceeds 150, the number of reactions will not change. This figure shows that the number and the content of the lethal reactions are unvaried as long as it is given enough iterative times. Due to the particularity of connected properties, we just iterate 200 times to get the stabile state (Fig. 3). Therefore, the metabolic networks are not the random network. Otherwise, they need enormous iterative times to reach unvaried state for random networks.

To investigate the minimal reaction networks changing with the rich organic nutrition, we add all the 129 organic compounds to the networks. In order to get a stabile state, we simulate 1,000 times to identify the minimal metabolic networks, and we find that the number of the lethal reactions is 86 (Fig. 4). It is very different from the result of the single organic D-glucose that needs 25 iterative times to stabilize the number of reactions used in every minimal metabolic network. Obviously, due to lots of alternate organic substances, the organism can have many candidate substrates to survive. Hence, it needs more iterative times to obtain the lethal reaction set. Also, we find that the viable minimal networks are reduced in which the number arrives at between 140 and 182 (Fig. 5).

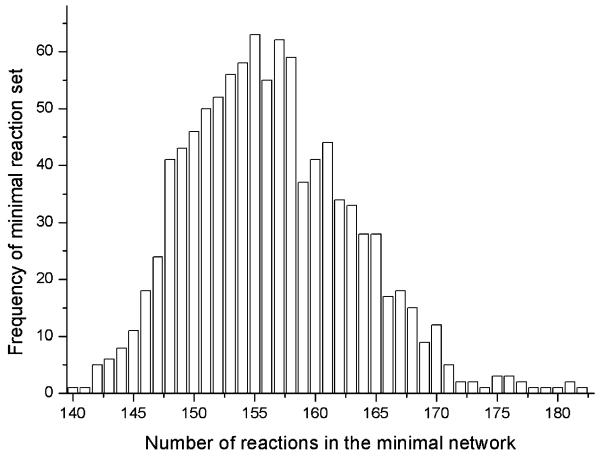

Fig. 5 The frequency of minimal metabolic reaction sets for rich nutrition environment

Table 3 Functional classification of lethal reactions for growth.

Metabolic Function	Glucose	Rich
Alternate carbon metabolism	3	1
Arginine and proline metabolism	12	0
Cell envelope biosynthesis	36	35
Citrate cycle (TCA)	3	0
Cofactor and prosthetic group biosynthesis	20	20
Cysteine metabolism	6	0
Folate metabolism	1	1
Glutamate metabolism	2	0
Glycine and serine metabolism	3	0
Glycolysis/gluconeogenesis	2	2
Histidine metabolism	10	0
Membrane lipid metabolism	16	14
Methionine metabolism	5	0
Nucleotide salvage pathways	5	4
Purine and pyrimidine Biosynthesis	5	0
Threonine and lysine metabolism	10	0
Transport, extracellular	1	0
Tyrosine, tryptophan, phenylalanine metabolism	16	7
Unassigned	1	1
Valine, leucine, and isoleucine metabolism	14	0

Classification of Lethal Reactions

The lethal reaction set for the single glucose organic nutrition consists of 172 reactions. Table 3 shows that most of these lethal reactions exist in the cell envelope, cofactor and prosthetic group biosynthesis, membrane lipid biosynthetic reactions, and amino acid metabolism. The frequencies allocated here are roughly the same with the previous results

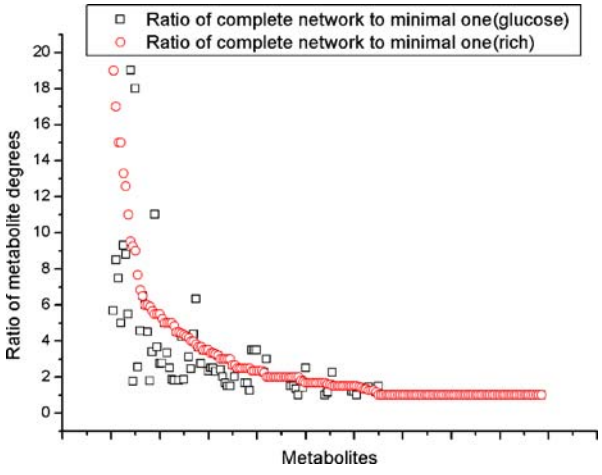

Fig. 6 The ratio of compound degree between the complete and the minimal metabolic networks

of single lethal reactions in FBA scheme [17]. Though the lethal reactions for the rich case are smaller than the ones for the glucose-only study, many of the lethal reactions still locate in the cell envelope, cofactor and prosthetic group biosynthesis, and membrane lipid biosynthetic reactions, as shown in Table 3. Our analytical results are compatible with the fact that a widespread strategy of anti-microbes is acting against cell wall synthesis [17].

The Node Degree Distributions

The metabolic network can be taken as a graph where nodes represent metabolites and the edges between nodes are enzymatic reactions. We randomly select a minimal metabolic network and count the node degree distributions of all the metabolites. Table 4 shows 21 key metabolites in *E. coli* ranked by degree connectivity. We find that most of the top-hub metabolites in minimal metabolic networks are also in current complete metabolic networks. For example, in the minimal metabolic networks of *E. coli*, the degree of "atp" is 48 and 50 in rich nutrition and glucose-only nutrition, respectively. This degree is high, which implies that the "atp" is important for the minimal metabolic networks. The degree in the complete metabolic network is high too. Hence, we validate that the hubs are functionally important and phylogenetically oldest in biological networks [18–20].

We compare the degrees of the compound existing both in the complete metabolic networks and the minimal metabolic networks (Fig. 6). Some compounds' degrees increase greatly. For example, the metabolite pyruvate's degree increases greatly from the minimal metabolic networks to the complete metabolic networks. The increased ratios are 19 and 5.7 for rich environment and glucose-only environment, respectively. However, not all compound's degrees increase from the minimal metabolic networks to the complete metabolic networks (Fig. 6). This phenomenon implies that the functions of a lot of metabolites do not change during the evolution of *E. coli*.

Discussion and Conclusion

We give a new method to extract the feasible minimal reactions to sustain growth in the complex metabolic network. It is seemingly difficult to enumerate all kinds of reactions deleting sequence to obtain feasible minimal reactions in a short computing time. However, due to the particularity of metabolic network, we can obtain it so long as we use enough iterative times in a short time. Among all these iterative processes, we obtain the three sets of reactions, which are (1) the fundamental reaction set, (2) the alternative reaction set, and (3) the dispensable reaction set. Of these three reaction sets, the fundamental reaction set is the most important for the organism in which all the reactions are essential. We also find that the minimal reaction set is determined by environment nutrition. Certainly, the content of the assistant substrates which are taken as the preexistent original seeds also varies the content of the minimal metabolic networks. The next work is to find the minimal number of assistant substrates so that the obtained minimal metabolic network will be more meaningful.

To get the minimal network which can sustain organism growth is still an open question. Though a few works (such as FBA method) have proposed to calculate the minimal metabolic network to sustain the growth, our method gives an alternative method to find the minimal metabolic network by the structure analysis. Each has its strong point. Recently, a similar work to detect the minimal metabolic networks is presented by traditional FBA [6]. There is one main difference between the FBA method and our method for simulating the

minimal metabolic networks. The method presented here is qualitative rather than quantitative, which is different from the constraints of FBA model. We do not constrain the balance of the metabolites in metabolic networks. But the balance of the metabolites is the foundation of traditional FBA method. The FBA method with maximizing the biomass yield may overestimate the number of essential reactions or the number of minimal reaction network to sustain the growth. In order to maximize the biomass, many by-products, which should excrete outside the organism with the growth of the organism, may return to the cellular. Then, a serial of additional reactions which convert these by-products to the biomass are taken as the essential ones. It needs more reactions in an organism to consume the redundant metabolites and transport them outside the organism lastly. Hence, the minimal reactions obtained by FBA are greater than the ones by the method used here, and it is certain that the content of the minimal metabolic network given in our simulation is the subset of what is given by the traditional FBA.

Acknowledgments Da Jiang, Hui Liu and Shuigeng Zhou are supported by National Natural Science Foundation of China under grant no. 60496327 and Shanghai Leading Academic Discipline Project (Project number B114).

References

1. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N., & Barabási, A. L. (2000). The large-scale organization of metabolic networks. *Nature*, *407*, 651–654. doi:10.1038/35036627.
2. Sharan, R., & Ideker, T. (2006). Modeling cellular machinery through biological network comparison. *Nature Biotechnology*, *24*, 427–433. doi:10.1038/nbt1196.
3. Becker, D., Selbach, M., Rollenhagen, C., Ballmaier, M., Meyer, T. F., Mann, M., et al. (2006). Robust Salmonella metabolism limits possibilities for new antimicrobials. *Nature*, *440*, 303–307. doi:10.1038/nature04616.
4. Schuster, S., Fell, D. A., & Dandekar, T. (2000). A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nature Biotechnology*, *18*, 326–332. doi:10.1038/73786.
5. Burgard, A. P., Vaidyaraman, S., & Maranas, C. D. (2001). Minimal reaction sets for *Escherichia coli* metabolism under different growth requirements and uptake environments. *Biotechnology Progress*, *17*, 791–797. doi:10.1021/bp0100880.
6. Pal, C., Papp, B., Lercher, M. J., Csermely, P., Oliver, S. G., & Hurst, L. D. (2006). Chance and necessity in the evolution of minimal metabolic networks. *Nature*, *440*, 667–670. doi:10.1038/nature04568.
7. Kauffman, K. J., Prakash, P., & Edwards, J. S. (2003). Advances in flux balance analysis. *Current Opinion in Biotechnology*, *14*, 491–496. doi:10.1016/j.copbio.2003.08.001.
8. Lee, J. M., Gianchandani, E. P., & Papin, J. A. (2006). Flux balance analysis in the era of metabolomics. *Briefings in Bioinformatics*, *7*, 140–150. doi:10.1093/bib/bbl007.
9. Edwards, J. S., & Palsson, B. O. (2000). The Escherichia coli MG1655 in silico metabolic genotype: Its definition, characteristics, and capabilities. *Proceedings of the National Academy of Sciences of the United States of America*, *97*, 5528–5533. doi:10.1073/pnas.97.10.5528.
10. Reed, J. L., Vo, T. D., Schilling, C. H., & Palsson, B. O. (2003). An expanded genome-scale model of *Escherichia coli* K-12 (iJR904 GSM/GPR). *Genome Biology*, *4*, R54. doi:10.1186/gb-2003–4–9-r54.
11. Feist, A. M., & Palsson, B. O. (2008). The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*. *Nature Biotechnology*, *26*, 659–667. doi:10.1038/nbt1401.
12. Fong, S. S., & Palsson, B. O. (2004). Metabolic gene-deletion strains of *Escherichia coli* evolve to computationally predicted growth phenotypes. *Nature Genetics*, *36*, 1056–1058. doi:10.1038/ng1432.
13. Fischer, E., & Sauer, U. (2005). Large-scale in vivo flux analysis shows rigidity and suboptimal performance of *Bacillus subtilis* metabolism. *Nature Genetics*, *37*, 636–640. doi:10.1038/ng1555.
14. Schuetz, R., Kuepfer, L., & Sauer, U. (2007). Systematic evaluation of objective functions for predicting intracellular fluxes in *Escherichia coli*. *Molecular Systems Biology*, *3*, 119. doi:10.1038/msb4100162.

15. Nielsen, J. (2007). Principles of optimal metabolic network operation. *Molecular Systems Biology*, *3*, 126. doi:10.1038/msb4100169.
16. Papp, B., Pal, C., & Hurst, L. D. (2004). Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature*, *429*, 661–664. doi:10.1038/nature02636.
17. Ghim, C. M., Goh, K. I., & Kahng, B. (2005). Lethality and synthetic lethality in the genome-wide metabolic network of *Escherichia coli*. *Journal of Theoretical Biology*, *237*, 401–411. doi:10.1016/j.jtbi.2005.04.025.
18. Fell, D. A., & Wagner, A. (2000). The small world of metabolism. *Nature Biotechnology*, *18*, 1121–1122. doi:10.1038/81025.
19. Wagner, A., & Fell, D. A. (2001). The small world inside large metabolic networks. *Proceedings of the Royal Society of London. Series B. Biological Sciences*, *268*, 1803–1810. doi:10.1098/rspb.2001.1711.
20. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N., & Barabási, A. L. (2002). Hierarchical organization of modularity in metabolic networks. *Science*, *297*, 1551–1555. doi:10.1126/science.1073374.