METABOLIC ENGINEERING

The small world of metabolism

David A. Fell and Andreas Wagner

Genome sequencing is now advancing at a frenetic pace, which has the consequence that many organisms now being sequenced have not had their biochemistry extensively studied. Thus, the metabolic phenotype of these organisms has to be determined using annotated genome sequence data. Ideally, this determination should be automated, but that would require clear criteria and algorithms for identifying and classifying metabolism. Also, traditional textbook representations of metabolic pathways may neither capture the full number of potential network functions nor the network's resilience to disruption[1–3]. Whereas algorithmic approaches to these latter problems have been proposed, many aspects of metabolic network function remain to be clearly delineated. For example, in stoichiometric network analysis[4], it is convenient to define a subset of central metabolic intermediates that represent the products of catabolism that are used to initiate anabolism. However, even for *Escherichia coli*, there is no agreement on the identity of this subset[5,6].

In seeking to establish a firm basis for identifying a set of central metabolites defining the core of metabolism[7,8], we have taken advantage of analysis tools used by mathematicians to understand the structure of sociological networks. These include networks of personal and professional relations, such as collaboration networks of film actors or scientists. The aspect of such analyses germane to our endeavor is best illustrated with the example of the prolific Hungarian graph theorist Paul Erdös. He is the center of a graph of mathematical collaboration. Coauthors of a paper with Erdös are one step from Erdös himself, and have Erdös number 1. Coauthors of mathematicians with Erdös number 1 have Erdös number 2, and so on. Most mathematicians active this century can be connected to Erdös in a small number of steps. In this sense, he is the undisputed center of the mathematical world[9].

A similar principle underlies the Kevin Bacon game, which has the aim of connecting an arbitrarily chosen movie actor with the actor Kevin Bacon by the shortest sequence of actor-pairs who have appeared together in a film. The average Bacon number of a randomly chosen actor, representing the mean minimum number of actors connecting the actor to Kevin Bacon is only 2.87 (ref. 10). (However, Kevin Bacon is not even the center of this small world of film actor collaborations, defined as having on average the shortest distance to all the other stars. This center is Christopher Lee, with a mean path length of 2.60.)

We recently have analyzed the structure of the *E. coli* core metabolism with the following goal in mind: to identify metabolites central to metabolism in this sense, without relying on subjective criteria. To this end, we assembled a list of 317 stoichiometric equations involving 275 substrates that represent the central routes of energy metabolism and small-molecule building block synthesis in *E. coli*[11–15] under aerobic growth, with glucose as sole carbon source and O_2 as electron acceptor. From these reaction equations, we generated a connection matrix where two metabolites were regarded as connected if they appeared in the same reaction, whether as substrate or product. We did not include the common coenzymes, such as ATP, ADP, or NAD, because they are evidently ubiquitous. On this basis, the center of the *E. coli* metabolic map is glutamate, with a mean path length of 2.46, followed by pyruvate with a value of 2.59.

The analogy between metabolism and the collaboration networks of mathematicians and film stars does not end there. The question of how networks that are both large and sparse can nevertheless be traversed in very few steps (cf. "six degrees of separation"[16]) has been analyzed by Watts and Strogatz[17]. Uniform, latticelike networks tend to have long path lengths because pairs that are connected tend to be connected to the same other members of the set, that is, they are clustered. Randomly generated networks can have short path lengths, but also show little clustering because connected pairs show little similarity in their other connections. However, a number of the sparse, natural networks studied by Watts and Strogatz showed short path lengths but high clustering. They named such networks "small-world" networks. The friendship networks studied in sociology as well as mathematical and acting collaboration networks are of this type. We found that the *E. coli* metabolic network falls into the same category[7,8]. This implies that, in modeling the properties of metabolism, neither very regular structures nor completely random networks would be very faithful representations of metabolism in general.

What is the biological relevance of the "small-worldness" of the *E. coli* metabolism? One way to generate a small-world network is to take a regular network and randomly reassign some of the connections. Another way is by accretion, where new members are added by preferentially making connections to existing members that already have large numbers of connections. Barabási and Albert[18] have shown that the latter led to a small-world network where the number of connections of the members fell off in a power–law relationship (i.e., a small number of members have a large number of connections, and this falls off smoothly so that the larger number of members has few connections). Randomly reassigned regular networks in contrast have a notably peaked distribution of connections. Film collaborations, hyperlinks in the worldwide web, and the US power grid all show the power–law connectivity[18].

The *E. coli* network does so as well[7,8]. Here, glutamate followed by pyruvate were the most connected metabolites (again omitting common coenzymes). After that, the lists of metabolites ranked by their number of connections and by their minimum mean path length to other metabolites were not exactly congruent, although both lists featured tricarboxylic acid cycle intermediates, and associated amino acids, in highly ranked positions. Recently, Jeong et al.[19] have reached a similar conclusion for the metabolic networks of a number of microorganisms by a related though slightly different analysis.

If, early in the evolution of life, metabolic networks grew by adding new metabolites, then the most highly connected metabolites should also be the phylogenetically oldest. Glycolysis and the tricarboxylic acid cycle are perhaps the most ancient metabolic pathways, and various of their intermediates (e.g., 2-oxoglutarate, succinate, pyruvate, and 3-phosphoglycerate) occur near the top of our lists, along with the amino acids thought to be used earliest (glutamine, glutamate, aspartate, and serine). This potential link with evolutionary history is consistent with Morowitz's[20] claim that intermediary metabolism recapitulates the evolution of biochemistry.

David A. Fell is professor of biochemistry at the School of Biological and Molecular Sciences, Oxford Brookes University, Headington, Oxford, OX3 0BP, UK (daf@bms.brookes.ac.uk), and Andreas Wagner is assistant professor of biology at the Department of Biology, University of New Mexico, Albuquerque, NM 87131-1091.

Of course, metabolism might have evolved small-world characteristics to optimize metabolic function in some way. Watts and Strogatz[17] have studied how fast perturbations spread through small-world networks. They concluded that the time required for spreading of a perturbation in a small-world network is close to the theoretically possible minimum. The importance of minimizing the transition time between metabolic states is recognized[21,22], and small-worldness may be a factor in allowing a metabolism to react rapidly to perturbations, although this requires further investigation because metabolic dynamics are more complicated than the simple kinetics used by Watts and Strogatz.

In conclusion, a purely structural analysis of a metabolic network may be able to teach us about the network's evolutionary history and design principles. Some of our propositions are speculative at this stage, but our analysis has revealed aspects of the network structure that were not previously apparent. The wealth of metabolic information about to become available from the genomes of ecologically diverse microbes will undoubtedly help to test these propositions.

1. Schilling, C.H., Schuster, S., Palsson, B.O. & Heinrich, R. Metabolic pathway analysis: basic concepts and scientific applications in the post-genomic era. *Biotechnol. Prog.* **15**, 296–303 (1999).
2. Edwards, J.S. & Palsson, B.O. The *Escherichia coli* MG1655 *in silico* metabolic genotype: Its definition, characteristics, and capabilities. *Proc. Natl. Acad. Sci. USA* **97**, 5528–5533 (2000).
3. Schuster, S., Fell, D.A. & Dandekar, T. A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks. *Nat. Biotechnol.* **18**, 326–332 (2000).
4. Varma, A. & Palsson, B.O. Metabolic capabilities of Escherichia coli. 1. Synthesis of biosynthetic precursors and cofactors. *J. Theoret. Biol.* **165**, 477–502 (1993).
5. Ingraham, J.L., Maaloe, O.E. & Neidhardt, F.C. *Growth of the bacterial cell.* (Sinauer Associates Inc, Sunderland, MA; 1983).
6. Holmes, W.H. The central metabolic pathways of Escherichia coli: relationship between flux and control at a branch point, efficiency of conversion to biomass, and excretion of acetate. *Curr. Top. Cellul. Regul.* **28**, 69–105 (1986).
7. Fell, D.A. & Wagner, A. Structural properties of metabolic networks: Implications for evolution and modelling of metabolism. In *Animating the cellular map.* (eds Hofmeyr, J.-H.S., Rohwer, J.M. & Snoep, J.L) 79–85 (Stellenbosch University Press, Stellenbosch, South Africa; 2000).
8. Wagner, A. & Fell, D.A. The small world inside large metabolic networks. Working paper 00-07-041 The Santa Fe Institute (2000). http://www.santafe.edu/sfi/publications/00wplist.html.
9. The Erdös number project (2000). http://www.acs.oakland.edu/ grossman/erdoshp.html.
10. The Oracle of Bacon at Virginia (2000). http://www.cs.virginia.edu/oracle/.
11. Neidhardt, F.C (ed.). *Escherichia coli* and *Salmonella*: *Molecular and cellular biology.* (ASM Press, Washington DC; 1996).
12. Karp, P., Riley, M., Paley, S., Pellegrini-Toole, A. & Krummenacker, M. Eco Cyc: Encyclopedia of *Escherichia coli* genes and metabolism. *Nucleic Acids Res.* **27**, 55–58 (1999).
13. Pramanik, J. & Keasling, J.D. Stoichiometric model of Escherichia coli metabolism: incorporation of growth-rate dependent biomass composition and mechanistic energy requirements. *Biotechnol. Bioeng.* **56**, 398–421 (1997).
14. Selkov, E. *et al*. The metabolic pathway collection from EMP—the enzymes and metabolic pathways database. *Nucleic Acids Res.* **24**, 26–28 (1996).
15. Bairoch, A. The ENZYME data bank in 1999. *Nucleic Acids Res.* **27**, 310–311 (1999).
16. Migram, S. The small—world problem. *Psychology Today* **2**, 60–67 (1967).
17. Watts, D.J. & Strogatz, S.H. Collective dynamics of `small-world' networks. *Nature* **393**, 440–442 (1998).
18. Barabási, A.L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
19. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407**, 651–654 (2000).
20. Morowitz, H.J. *Beginnings of cellular life: metabolism recapitulates biogenesis.* (Yale University Press, New Haven, CT; 1992).
21. Easterby, J.S. The effect of feedback on pathway transient response. *Biochem. J.* **233**, 871–875 (1986).
22. Cascante, M., Meléndez-Hevia, E., Kholodenko, B.N., Sicilia, J. & Kacser, H. Control analysis of transit-time for free and enzyme-bound metabolites—physiological and evolutionary significance of metabolic response-times. *Biochem. J.* **308**, 895–899 (1995).