

Minireview

Network-based approaches for linking metabolism with environment

Sarath Chandra Janga and M Madan Babu

Address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 0QH, UK.

Correspondence: Sarath Chandra Janga. Email: sarath@mrc-lmb.cam.ac.uk

Published: 24 November 2008

Genome Biology 2008, **9**:239 (doi:10.1186/gb-2008-9-11-239)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/11/239>

© 2008 BioMed Central Ltd

Abstract

Progress in the reconstruction of genome-wide metabolic maps has led to the development of network-based computational approaches for linking an organism with its biochemical habitat.

The sequential nature of the reactions in metabolic pathways means that they can be modeled in the form of a graph (network) of enzymes and chemical transformations, and network theory can be used to represent and understand metabolism [1,2]. The connected collection of metabolic pathways, describing the set of all enzymatic interconversions of one small molecule into another, is defined as the metabolic network of an organism (Figure 1a).

The most commonly used network representations are 'metabolite-centric'. They consider metabolites as the nodes of the graph and two metabolites are linked if one can be converted into the other by an enzymatic reaction (Figure 1b, left). An alternative network representation is 'enzyme-centric'. It considers the enzymes as nodes and links enzymes that catalyze successive reactions (Figure 1b, right). Although several studies have provided insights into the structure and evolution of a metabolic network, very few have addressed the influence of environment on metabolic network structure in species from diverse environmental conditions. The availability of many completely sequenced genomes means that metabolic-network analysis can now be extended from a few model organisms to species from different branches of the tree of life and living in very different environments. This should enable the elucidation of general principles underlying metabolic networks.

Two recent studies, published in the *Proceedings of the National Academy of Sciences* by Eytan Ruppin and colleagues (Kreimer *et al.* [3] and Borenstein *et al.* [4]), provide important insights into links between the environment of an organism and the structure of its metabolic network. Using

data from a large number of bacterial metabolic networks, Kreimer *et al.* address the question of how the topologies of the metabolic networks from different species reflect both genome size and the diversity of environmental conditions the species would encounter. Borenstein *et al.* set out to identify the 'seed set' - that set of small molecules that are absolutely needed from the external environment - of each species and how this seed set differs across species from different environments.

A network view of metabolism

Several studies have addressed a wide-range of questions using network representation of small-molecule metabolism [5-7]. For instance, at the structural level, the metabolic network of an organism has been shown to have a scale-free topology with few nodes (for example, pyruvate or coenzyme A) reacting with many other substrates [8,9]. A distinguishing feature of such scale-free networks is the existence of a few highly connected metabolites, which participate in a very large number of metabolic reactions. By definition, when a large number of links integrate several substrates into a single highly connected component, fully separated modules will not exist. This has led to the notion of hierarchical modular structures within the fully connected metabolic network, where a 'module' is defined as a group of nodes that are more connected to each other than to other nodes in the network [10].

Kreimer *et al.* [3] have carried out a comprehensive, large-scale characterization of metabolic-network modularity (defined as in [11]) using 325 prokaryotic species with

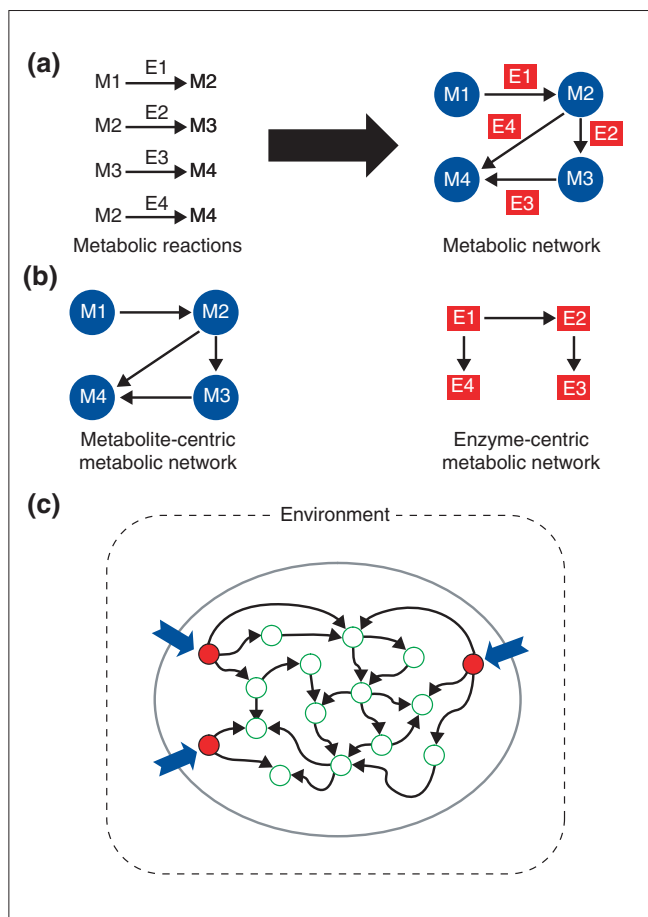


Figure 1
Metabolic networks. (a) A set of related metabolic reactions can be represented as a network. M1, M2, and so on are metabolites and E1, E2, and so on are the enzymes that catalyze the conversion of one metabolite into another. The arrows represent the direction of the reaction. (b) Different ways of representing a metabolic network: left, with the metabolites as nodes; right, with the enzymes as nodes. (c) Representation of seed compounds in a hypothetical metabolic network. The metabolic boundary of the organism is represented by the gray oval. Metabolites (the nodes in the network) are represented by colored circles. The set of compounds that cannot be internally synthesized but must be obtained from the environment is referred to as the seed set, and is represented here as red circles. Seed metabolites form the interface between the environment and the metabolic system and link the metabolic habitats of an organism with its core metabolic processes. In this hypothetical network, it is possible to reach any of the internal nodes (open green nodes) from any other node except those that have to be obtained from the environment (blue arrows).

sequenced genomes and metabolic networks in the KEGG pathway database [12]. They found that network size was an important topological determinant of modularity, with larger genomes exhibiting higher modularity scores (that is, a higher proportion of edges in the network forming part of modules than would be expected by chance). In addition, several environmental factors were shown to contribute to the variation in metabolic-network modularity across species. In particular, the authors found that endosymbionts and

mammal-specific pathogens have lower modularity scores than bacterial species that occupy a wider range of niches. Moreover, among the pathogens, those that alternate between two distinct niches, such as insect and mammal, were found to have relatively high metabolic-network modularity. This supports the notion previously put forward by Parter *et al.* [13] that variability in the natural habitat of an organism promotes modularity in its metabolic network. Kreimer *et al.* [4] also reconstructed likely ancestral states, and found that modularity tends to decrease from ancestors to descendants; they attribute this to niche specialization and incorporation of peripheral metabolic reactions.

In line with the above effects of environmental diversity on network structure, Pal *et al.* [14] observed that bacterial metabolic networks grow by retaining horizontally acquired genes (genes acquired from other species) involved in the transport and catalysis of external nutrients, and that evolutionary changes in networks are primarily driven by adaptation to changing environments. Accordingly, horizontally transferred genes were found to be integrated at the periphery of the network, whereas the central parts remain evolutionarily stable. Indeed, genes encoding physiologically coupled reactions were often found to be transferred together, frequently in operons. This suggests that bacterial metabolic networks evolve by direct uptake of peripheral reactions in response to changing environments [14].

In this regard, a recent genome-wide study in yeast found that central and highly connected enzymes evolve more slowly than less connected ones and that duplicates of highly connected enzymes tend to have a higher likelihood of retention [15]. Enzymes carrying high metabolic fluxes under natural biological conditions were also found to experience greater evolutionary constraints. Interestingly, however, it was shown that highly connected enzymes are no more likely to be essential to survival than the less connected ones [15].

The functional and evolutionary modularity of the *Homo sapiens* metabolic network has also been investigated from a topological point of view and was shown to be organized with a highly modular, ‘core and periphery’ topology [16]. In such a structure, the core modules are tightly linked together and perform basic metabolic functions, whereas the peripheral modules only interact with few other modules and accomplish relatively independent and specialized functions. Interestingly, as in bacteria and yeast, peripheral modules were found to evolve more cohesively and faster than core modules [16].

Linking external environment to the metabolic circuitry

Microorganisms constantly monitor their surroundings for the availability of nutrients and other chemicals, using both

Box 1. Models of metabolic pathway evolution

The most influential models of metabolic pathway evolution have been the 'retrograde model' proposed by Horowitz in 1945 [24] and the 'patchwork model' proposed by Ycas in 1974 [25] and later improved by Jensen in 1976 [26].

The retrograde model

In the retrograde model, pathways evolve bottom-up from a key metabolite, which is assumed to be initially abundant in the ancestral condition. The model presupposes the existence of a chemical environment in which both the key metabolite and potential intermediates are available. An organism primarily dependent on molecule Z will use up environmental reserves of the metabolite to the point at which its growth is restricted; in such an environment, an organism capable of synthesizing molecule Z from environmental precursors X and Y will have a selective advantage. Any natural variant evolving an enzyme that catalyzes this synthesis will have a fitness advantage in such an environment. As a result, with the drop in environmental concentration of X or Y, the process will be repeated, with the similar recruitment of further enzymes.

The retrograde model also proposes that the simultaneous unavailability of two intermediates (say X and Y) would favor symbiotic association between two mutants, one capable of synthesizing X and the other of synthesizing Y from other environmental precursors. One of the major assumptions of this model is that the evolution of metabolic pathways occurs in an environment rich in metabolic intermediates, and it therefore cannot explain their evolution during major environmental transitions in the history of life such as, for example, the depletion of organic molecules from the environment [24,27]. The retrograde model also fails to explain the development of pathways that include labile metabolites, which could not have accumulated in the environment for long enough for retrograde recruitment to take place.

The patchwork model

In light of these limitations, Ycas [25] and Jensen [26] proposed the patchwork model of metabolic pathway evolution, in which pathway evolution depends on the initial existence of broad-specificity enzymes. In its original formulation [25], such enzymes catalyze whole classes of reactions, forming a large network of possible pathways. The broad specificities would mean that many metabolic chains, synthesizing key metabolites, may have existed, although short and incomplete compared with the pathways observed today. The duplication of genes in such pathways (advantageous because increased levels of the enzyme would generate more of the key metabolites), followed by their specialization, would account for extant pathways. Jensen [26] subsequently pointed out that the fortuitous evolution of a novel chemistry, together with the biological leakiness of such a system, could allow the production of a key metabolite from a novel intermediate, even if it is several enzymatic steps away from the original product.

external and internal sensors to respond dynamically to environmental changes [17]. Integration of the external environment with metabolism occurs through the import of compounds from the environment and results, for example, in a transcriptional response or an allosteric interaction with an enzyme [18-20]. In the second of the recent studies from Ruppin and co-workers, Borenstein *et al.* [4] propose a graph-theoretical approach to define these exogenously acquired compounds - the seed set of an organism - and have identified their repertoire across the tree of life (Figure 1b). This is one of the most comprehensive studies so far that links organisms' metabolic circuitry with their environment.

The authors represent the metabolic network of a given species as a directed graph with nodes representing metabolites and edges corresponding to the linking reactions converting substrates to products. Using this, they identify

the maximal set of metabolites that can be synthesized from a particular precursor metabolite. This graph-based representation of the metabolic network then enabled them to discover the seed-set compounds for each of the 478 prokaryotic species with available metabolic networks in the KEGG database [12]. On the whole, they found that about 8-11% of the compounds in the metabolic network of an organism correspond to the seed set. Their predictive ability to correctly identify seed compounds reached a precision of 95% when benchmarked against a set of compounds experimentally characterized as being taken up from the environment by the rickettsia that cause the disease ehrlichiosis in humans and animals. Recall values (defined as the percentage of correctly identified seeds of all exogenously acquired compounds) based on the same dataset were low, suggesting that other factors might have a role in the identification of seed compounds of an organism, such as

the incompleteness of the metabolic network or ways of acquiring an exogenous compound that cannot be captured by currently available metabolic maps. The resulting compilation, which represents the overall static metabolic interface of each organism characterizing its biochemical habitat, enabled Borenstein *et al.* to trace the evolutionary history of both metabolic networks and growth environments.

When the seed sets identified in each organism were analyzed in detail, species living in variable environments were found to have more versatile seed sets, in terms of variability of size and diversity of composition. On the other hand, obligate parasites like *Buchnera aphidicola* and those microorganisms, such as archaea, that live in extreme and narrowly defined environments, were found to have much smaller seed set sizes. These results suggest that although organisms surviving in predictable environments can take up many compounds from their surroundings, this capability is still significantly smaller than in organisms that have to survive in a wide range of niches.

Borenstein *et al.* [4] carried out a phylogenetic analysis of the seed sets across different taxa, which suggested not only that an accurate tree of life can be reconstructed from them but that such a tree can provide insights into the evolutionary dynamics of seed compounds. In particular, the study revealed that novel compounds can be integrated into the metabolic network of an organism as either non-seeds or seeds, and that seed compounds are more likely to be lost during evolution than non-seed compounds. From the comparison with ancestral metabolic networks, Borenstein *et al.* [4] suggest that the transition from seed to non-seed compound occurs 2.5 times more often than the reverse. This suggested that, of the two main current hypotheses of metabolic network evolution - the 'patchwork' and 'retrograde' models (see Box 1) - the retrograde model, in which pathways evolve in a direction opposite to the metabolic flow, might best explain the observed events. However, the observations of Borenstein *et al.* [4] on the high overall rate of integration of non-seed compounds and the relatively high rate of transition of non-seed compounds into seed metabolites, suggest that some aspects of network evolution could be explained by the patchwork and other models. The results highlight the fact that these models are not mutually exclusive, but complementary, and might have contributed to pathway evolution to different extents [21,22].

It should be noted that there are limitations to studies such as those reported here, in that the incompleteness of metabolic maps, the reversibility of reactions, possible alternative mechanisms controlling metabolic import, and the ignoring of the distinction between catabolic and anabolic pathways can all potentially result in false positives in the identified seed sets. Nevertheless, it is exciting to note that seed sets obtained using the approach developed in these studies not only reflect the metabolic environments of the species

themselves but also provide insight into their natural biochemical habitats - the union of all the metabolic environments an organism encounters.

Hence, such approaches can be exploited to study the interaction and association of microbes with other species thriving in similar habitats. This may help in the identification of host-parasite and symbiotic relationships between organisms and also enable the prediction and design of drugs that can precisely target an organism of interest without adversely affecting the host. With the availability of metagenomic data ranging from viromes to biomes [23], we anticipate that similar approaches can be applied to study metagenomic environments to decipher species relationships and dependencies occurring in large ecological niches, thereby providing insights into ecological imbalances or tradeoffs.

Acknowledgements

SCJ and MMB acknowledge financial support from the MRC Laboratory of Molecular Biology. SCJ acknowledges financial support from Cambridge Commonwealth Trust. MMB thanks Darwin College and Schlumberger Ltd for generous support. We thank A Wuster, R Janky, K Weber, V Espinosa-Angarica and JJ Diaz-Mejia for critically reading the manuscript and providing helpful comments.

References

- Papin JA, Price ND, Wiback SJ, Fell DA, Palsson BO: **Metabolic pathways in the post-genome era.** *Trends Biochem Sci* 2003, **28**:250-258.
- Feist AM, Palsson BO: **The growing scope of applications of genome-scale metabolic reconstructions using *Escherichia coli*.** *Nat Biotechnol* 2008, **26**:659-667.
- Kreimer A, Borenstein E, Gophna U, Ruppin E: **The evolution of modularity in bacterial metabolic networks.** *Proc Natl Acad Sci USA* 2008, **105**:6976-6981.
- Borenstein E, Kupiec M, Feldman MW, Ruppin E: **Large-scale reconstruction and phylogenetic analysis of metabolic environments.** *Proc Natl Acad Sci USA* 2008, **105**:14482-14487.
- von Mering C, Zdobnov EM, Tsoka S, Ciccarelli FD, Pereira-Leal JB, Ouzounis CA, Bork P: **Genome evolution reveals biochemical networks and functional modules.** *Proc Natl Acad Sci USA* 2003, **100**:15428-15433.
- Spirin V, Gelfand MS, Mironov AA, Mirny LA: **A metabolic network in the evolutionary context: multiscale structure and modularity.** *Proc Natl Acad Sci USA* 2006, **103**:8774-8779.
- Guimera R, Nunes Amaral LA: **Functional cartography of complex metabolic networks.** *Nature* 2005, **433**:895-900.
- Wagner A, Fell DA: **The small world inside large metabolic networks.** *Proc Biol Sci* 2001, **268**:1803-1810.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL: **The large-scale organization of metabolic networks.** *Nature* 2000, **407**:651-654.
- Ravasz E, Somera AL, Mongru DA, Oltvai ZN, Barabasi AL: **Hierarchical organization of modularity in metabolic networks.** *Science* 2002, **297**:1551-1555.
- Newman ME: **Modularity and community structure in networks.** *Proc Natl Acad Sci USA* 2006, **103**:8577-8582.
- Okuda S, Yamada T, Hamajima M, Itoh M, Katayama T, Bork P, Goto S, Kanehisa M: **KEGG Atlas mapping for global analysis of metabolic pathways.** *Nucleic Acids Res* 2008, **36**(Web Server issue):W423-W426.
- Parter M, Kashtan N, Alon U: **Environmental variability and modularity of bacterial metabolic networks.** *BMC Evol Biol* 2007, **7**:169.
- Pal C, Papp B, Lercher MJ: **Adaptive evolution of bacterial metabolic networks by horizontal gene transfer.** *Nat Genet* 2005, **37**:1372-1375.
- Zhao J, Ding GH, Tao L, Yu H, Yu ZH, Luo JH, Cao ZW, Li YX: **Modular co-evolution of metabolic networks.** *BMC Bioinformatics* 2007, **8**:311.

16. Vitkup D, Kharchenko P, Wagner A: **Influence of metabolic network structure and function on enzyme evolution.** *Genome Biol* 2006, **7**:R39.
17. Martinez-Antonio A, Janga SC, Salgado H, Collado-Vides J: **Internal-sensing machinery directs the activity of the regulatory network in *Escherichia coli*.** *Trends Microbiol* 2006, **14**:22-27.
18. Seshasayee AS, Fraser GM, Babu MM, Luscombe NM: **Principles of transcriptional regulation and evolution of the metabolic system in *E. coli*.** *Genome Res* 2008. doi: 10.1101/gr.079715.108.
19. Balaji S, Babu MM, Aravind L: **Interplay between network structures, regulatory modes and sensing mechanisms of transcription factors in the transcriptional regulatory network of *E. coli*.** *J Mol Biol* 2007, **372**:1108-1122.
20. Janga SC, Salgado H, Martinez-Antonio A, Collado-Vides J: **Coordination logic of the sensing machinery in the transcriptional regulatory network of *Escherichia coli*.** *Nucleic Acids Res* 2007, **35**:6963-6972.
21. Diaz-Mejia JJ, Perez-Rueda E, Segovia L: **A network perspective on the evolution of metabolism by gene duplication.** *Genome Biol* 2007, **8**:R26.
22. Teichmann SA, Rison SC, Thornton JM, Riley M, Gough J, Chothia C: **Small-molecule metabolism: an enzyme mosaic.** *Trends Biotechnol* 2001, **19**:482-486.
23. Dinsdale EA, Edwards RA, Hall D, Angly F, Breitbart M, Brulc JM, Furlan M, Desnues C, Haynes M, Li L, McDaniel L, Moran MA, Nelson KE, Nilsson C, Olson R, Paul J, Brito BR, Ruan Y, Swan BK, Stevens R, Valentine DL, Thurber RV, Wegley L, White BA, Rohwer F: **Functional metagenomic profiling of nine biomes.** *Nature* 2008, **452**:629-632.
24. Horowitz NH: **On the evolution of biochemical syntheses.** *Proc Natl Acad Sci USA* 1945, **31**:153-157.
25. Ycas M: **On earlier states of the biochemical system.** *J Theor Biol* 1974, **44**:145-160.
26. Jensen RA: **Enzyme recruitment in evolution of new function.** *Annu Rev Microbiol* 1976, **30**:409-425.
27. Lazcano A, Miller SL: **On the origin of metabolic pathways.** *J Mol Evol* 1999, **49**:424-431.