Scale-Rich Metabolic Networks

Reiko Tanaka*

Bio-Mimetic Control Research Center, RIKEN, Moriyama-ku, Nagoya 463-0003, Japan

(Received 1 October 2004; published 25 April 2005)

Recent research in biology has clarified many features of the global organization of metabolic networks, including the biochemical mechanisms responsible for power laws in metabolite degrees. The primary aim of this Letter is to give the simplest possible biochemical explanations and minimal toy models based on a highly optimized tolerance perspective, which show where and why metabolic networks have power laws. A second aim is to further explain why scale-free explanations fail in this case to correctly describe metabolism.

DOI: 10.1103/PhysRevLett.94.168101 PACS numbers: 87.17.Aa

Scale-free [1,2] (SF) and highly optimized (organized) tolerance (trade-offs) (HOT) [3] theories both predict that power-law node degree distributions should naturally arise in complex networks, but for opposite and largely incompatible reasons. Here we focus on the stoichiometry of metabolism, the simplest and most unambiguously known aspect of biological networks. Along with the Internet [4,5], metabolism has been proposed as the canonical SF network [1], making it an attractive basis for comparing these very different approaches. A stoichiometry matrix (S-matrix) has rows of metabolites and columns of reactions, as shown for a simple pair of reactions in Fig. 1. Figure 2 shows a color coding of the S-matrix for *H. Pylori* core metabolism [6], with both metabolites (rows) and reactions (columns) decomposed into modules.

This categorization of metabolites is compatible with the standard schematic "bow tie" structure of metabolism [7], where a large "fan in" of nutrient inputs is catabolized to produce a small handful of activated carriers and precursor metabolites, which then "fan out" to the biosynthesis of a large number of primary building blocks. The biologically natural modular decomposition in metabolites is thus into "knot" (carriers and precursors) and nonknot (others) parts of the bow tie. While this is largely a network-level interpretation of standard textbook biochemistry, statistical studies [8] of 80 fully sequenced organisms produce similar conclusions about the universal bow tie structure of metabolism. The information conveyed in the S-matrix can be visualized as a color-coded bipartite graph, which we call an S-graph, as shown in Fig. 1. Models which further project bipartite S-graphs, as is standard in the physics literature [1,2], with only either metabolites or reactions (by elimination of the other) necessarily destroy their biochemical meaning. An S-graph for the amino acids biosynthesis module in *H. Pylori* is shown in Fig. 3.

The central claim motivating this Letter is that metabolite degrees obey a power law, although degrees (number of edges from a node) for both types of nodes, reaction and metabolite, are biologically important (and equivalent to degrees of columns and rows of the S-matrix). The full network metabolite degrees (black +) in Figs. 4(d)–4(f) does show an approximate power-law distribution in a log-log Figs. 4(e) and 4(f) rank plot and has clearly higher variability than an exponential as seen in a semilog Fig. 4(d) plot. What is more fundamental, however, than power laws is high variability. For low variability processes, Gaussians arise naturally because of the well-known central limit theorem (CLT) and thus require no additional "special" explanations. Exponentials have other important invariance properties and are also thus quite common. All degrees of each module in Fig. 4 are closer to exponentials and have low variability. Even more important is that relaxing finite variance in the CLT yields

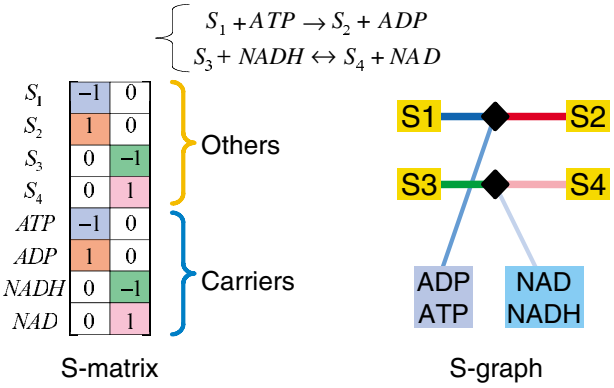

FIG. 1 (color). S-matrix and S-graphs of the two reactions shown among four carriers ATP, ADP, NADH, and NAD and four other metabolites S_1, S_2, S_3, and S_4, with enzymes hidden. The S-matrix has a color-code that helps visualize larger S-matrices. Red and blue correspond to positive and negative elements, respectively, for irreversible reactions, and pink and green correspond to positive and negative elements, respectively, for reversible reactions. An S-graph consists of reaction nodes (black diamond), noncarrier metabolite nodes (orange square), and carrier metabolite nodes (light blue square). Edges are color coded as in the S-matrix, so all the information in the S-matrix appears schematically in the S-graph. Carriers which always occur in pairs (ATP/ADP, NAD/NADH, etc.) are grouped into a single node.

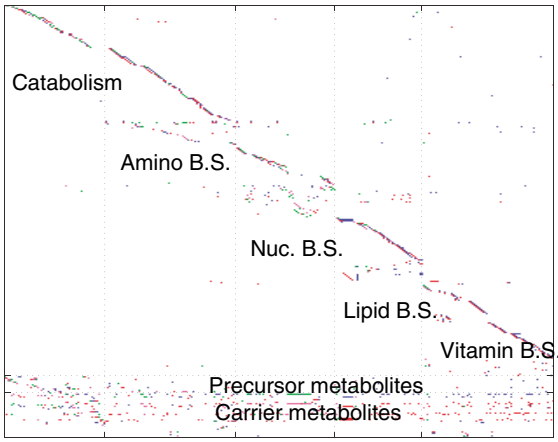

FIG. 2 (color). S-matrix for *H. Pylori* metabolism with 325 metabolites and 315 reactions. Reactions (columns) are decomposed into standard functional modules of catabolism, amino acid biosynthesis, nucleotide biosynthesis, lipid biosynthesis, and vitamin biosynthesis. The rows (metabolites) are arranged by their role in reaction modules to clarify the sparsity pattern of long chains of successive reactions from inputs to outputs in each module. The bottom rows are precursor metabolites and carrier metabolites, which appear throughout different reaction modules. The 12 precursor metabolites are outputs of catabolism and are the starting points for biosynthesis. Carrier metabolites correspond to conserved quantities, are activated in catabolism, and act as carriers to transfer energy and phosphate groups (ATP/ADP/AMP), hydrogen/redox (NADH/NAD), amino groups (AKG/GLU), acetyl groups (ACCOA/COA), and one carbon units (THF/METHF) throughout all modules. The other (than carriers and precursors) metabolites occur primarily in separate reaction modules.

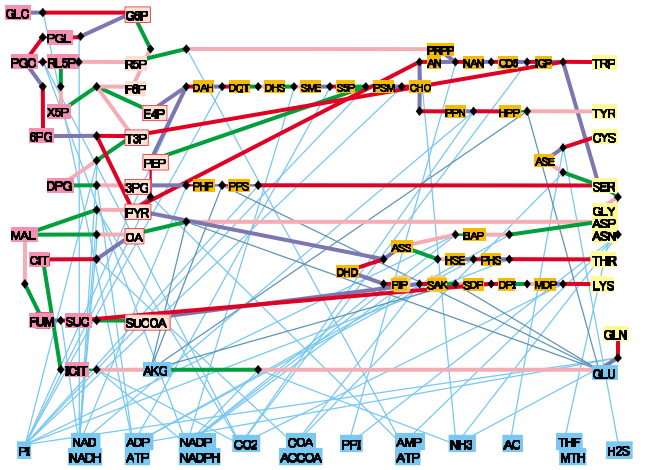

FIG. 3 (color). An S-graph for the amino acid biosynthesis module of the *H. Pylori* S-matrix. The conventions are the same as those in Fig. 1. This illustrates that long biosynthetic pathways build complex building blocks (in yellow on the right) from precursors (in orange on the left) in a series of simple reactions (in the middle). Each biosynthetic module has a qualitatively similar structure.

power laws, which are further invariant under marginalization, mixtures, and maximization [9]. Given the abundance of high variability phenomena, power laws are an obvious null hypothesis and should properly be viewed as "more *normal* than normal" [10]. Thus we focus on the mechanism responsible for high variability in total metabolite degrees despite low variability in all other degrees, including reactions and module metabolites.

Table I shows the coefficient of variation (CV $= \sigma/\mu$, where μ and σ are the sample mean and the standard deviation) for the horizontal and vertical decomposition of the S-matrix in Fig. 2. The CV is a standard measure of variability with low variability exponentials having CV $= 1$, and power laws having divergent CV for large data sets. The only high variability in Table I appears for all metabolites in the full network (all modules). It is obvious from Fig. 4(d), which shows the decomposition of metabolites into carrier ($\bigcirc$), precursor ($\diamond$), and other metabolites (*), that the high variability in the whole network is created by high σ from carrier metabolites mixed with low μ from others. Figure 4(a) shows the decomposition of carrier degrees into reaction modules. The larger marker corresponds to the degree in the whole network, whereas the smaller ones correspond to those in each reaction module. The sum of shared carrier metabolites across different reaction modules pushes the total degree of carriers much higher. In contrast, the degrees for other metabolites (*) stays smaller with many low degrees in total [Fig. 4(d)]. Its decomposition into reaction modules is shown in Fig. 4(c). As they appear almost uniquely in each reaction module, the sum across different modules increases the number and thus ranks, but not greatly the degrees. The node degrees for precursor metabolites have properties between those of carriers and others [Fig. 4(b)]. Figure 4(f) and the bottom row of Table I show another decomposition of all the metabolites in the full network (+) into reaction modules, each of which has relatively low variability. The reaction node degrees, the number of metabolites involved in each reaction, are shown in Fig. 5. The number of carriers involved in a reaction is also an important statistic. The typical reaction has four metabolites of which two are carriers, and no reactions differ greatly from this. The overall reaction degrees has very low variability (CV $=$ 0.30), since the enzymes of core metabolism are highly efficient and specialized for high fluxes of small metabolites and thus necessarily have few metabolites and involve simple reactions. This is not trivial, since the general purpose polymerases, chaparones, and proteases involved elsewhere in the cell have an almost unlimited number of distinct substrates.

In summary, the overall high variability and thus apparent power law in total metabolite degrees is created by a *mixture* of a few high-degree carriers with the many (high rank and) low degree of other metabolites unique to each reaction module, with the precursors filling in between

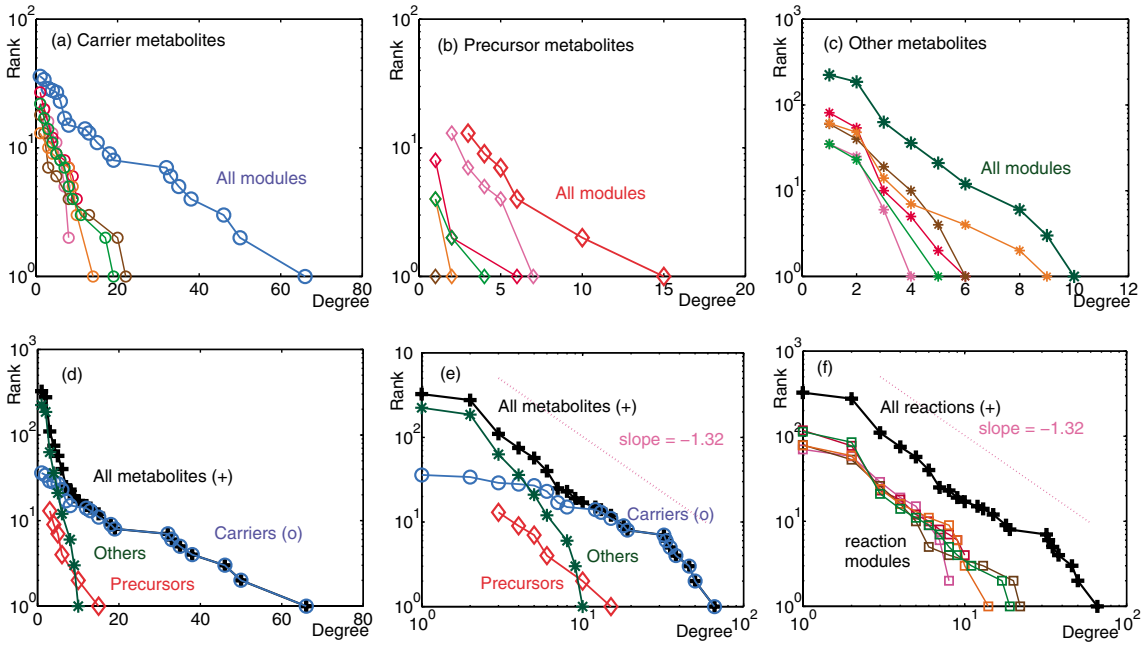

FIG. 4 (color). Rank (cumulative distribution) of metabolite node degree (= number of reactions = number of links) for metabolic networks of *H. Pylori*. Degrees of (a) carrier, (b) precursor, and (c) other metabolites in the whole network [large marker with (a) blue, (b) red, and (c) dark green] and in each reaction module (small markers with pink, dark red, brown, orange, and light green colors). Each module shows exponential distribution. (d) Metabolite node degrees of the whole network (black +) resulting from the mixture of carrier (○), precursor (◇), and other metabolites (*), for which the plot is the same as for (a), (b), and (c), respectively. (e) Log-log plot of (d) indicates total degrees are approximately power laws. (f) The total metabolites in each reaction module with exponential distribution sums up to create the power-law distribution in the whole network.

[Fig. 4(d)]. It is verified by the simplest HOT toy stoichiometry model possible that still yields this high metabolite variability. For its construction, we must abstract both the biological functionality that metabolism provides the cell, and the constraints on its components. A sufficient model assumes that each reaction has exactly one global carrier and one other metabolite, that there is just one carrier which appears in every reaction, and that each other metabolite is in just one reaction. With these assumptions, in r reactions, the σ and therefore the CV of both the carrier and other metabolites is exactly 0, the lowest CV value possible. The mixture of carriers and others has one degree-r carrier and r degree-1 others. For large r this gives $\mu \approx 2$ and $\sigma \approx \sqrt{r}$, so the total CV $\approx \sqrt{r}/2$. This is the highest possible CV value that the metabolites in a nontrivial r reaction S-matrix can have. This simple model thus shows that even one shared common carrier makes a high variability at the full system despite low variability within all modules.

TABLE I. Coefficients of variation of metabolite node degree distribution in catabolism (C) and amino acid (A), nucleotide (N), lipid (L), and vitamin (V) biosynthetic modules. Each of the carrier, precursor, and other metabolites has a low variability in each module, and their sum results in the high variability in total.

	C	A	N	L	V	All modules
Others	0.38	0.49	0.56	0.67	0.42	0.61
Precursors	0.47	1.05	0	0.35	0.61	0.60
Carriers	0.50	0.81	1.23	0.64	0.92	1.13
All metabolites	0.63	0.88	1.20	0.90	1.04	1.72

These assumptions are so extremely simplified that they would not even allow reactions to chain together to create pathways, but this underscores the point that the mechanism creating power laws in metabolism depends only minimally on the properties of biochemistry *per se*, provided those properties are properly identified. This HOT model is minimal in the biological sense that no simpler reactions are possible and in the mathematical sense that it has minimal assumptions and can trivially be solved analytically. Real S-matrices have broader distributions on both metabolites and reactions and this smears out the distributions and lowers the CV, but the qualitative features are universal and preserved. Indeed, one could argue that high variability itself is thus a relatively uninteresting and certainly unsurprising feature of metabolism. Moreover, the strong invariance properties of power laws means that they can be easily caused by models based on only the most minimal constraints of real metabolism, once they have high CV. Of more interest is that the (both row and column) modular bow tie structure of stoichiometry [7] shows it to facilitate flexibility, adaptability, efficiency, robustness, and evolvability in the face of a large number of constraints

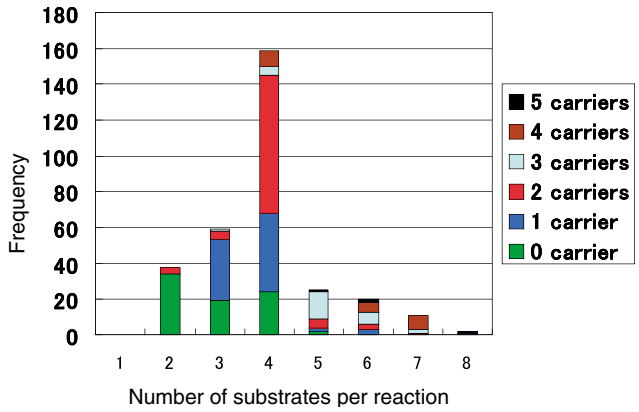

FIG. 5 (color). Reaction node degree (= number of substrates) distribution for metabolic networks of *H. Pylori*. CV = 0.30. The contributions of carrier metabolites to degrees is also indicated.

on conserved quantities involving energy, redox, and many small moieties. This kind of protocol-based modular architecture is ubiquitous throughout biology and advanced technologies as well [7,11].

SF models provide explanations for power-law degree distribution that are not just different, but in every way completely opposite. A difficulty in evaluating SF models is, however, that, beyond having power laws, the definition of scale-free networks and its implications have never been made precise. It is possible to provide a precise definition and rigorous proofs of many of the claimed SF properties [4], and it can be shown that *none* of the properties attributed to SF networks holds for metabolism. Although a full treatment is well beyond the scope of this Letter, a few issues can be briefly sketched. The most familiar feature of SF networks is that their high-degree nodes are responsible for global connectivity and their removal fragments the network. That this does not hold at all for metabolism is readily seen from Fig. 3, where the removal of all carriers leaves the biosynthetic pathways fully intact. While it is obviously true that without carriers metabolism would not function, it is not explained by graph properties alone. Thus the high-degree (carrier) nodes in metabolism are not hubs in the sense of SF graphs. In fact, the metabolic network is highly "self-dissimilar" in the sense that the metabolite degree distributions are very different at the full systems level (power law) and at the module levels (exponentials), which is opposite from SF networks. Metabolism consists of widely *different scales* in modules as is shown in Fig. 4 and thus could more appropriately be called *scale-rich*. Precise statements and proofs of these assertions about self-dissimilarity and scale richness can be made but require additional machinery [4]. That simplistic graph models without any biochemical content can be very misleading was also pointed out in [12].

The contrast here between HOT and SF models is not special to either metabolism or to scale-free itself. Similar results hold for SF models of the Internet [4,5], as well as for models of power laws based on self-organized criticality (SOC) [13]. The essential difference is that HOT tries to minimally capture the trade-offs involved in robust and efficient functionality in the presence of constraints and uncertainties. It allows for and even explains the highly structured and organized networks that result from engineering design or biological evolution, as is illustrated by metabolic stoichiometry. SF and SOC both emphasize explaining a few macroscopic statistics, typically power laws, that emerge from otherwise random models with minimal tuning. When possible, HOT has been applied to such models to illustrate the profound effects of evolution or design. Here it was applied to find a minimal stoichiometry that produced the low/high variability in degrees seen in real S-matrices.

The author thanks John Doyle for helpful discussions.

*Electronic address: reiko@bmc.riken.jp

[1] A.-L. Barabási and Z. Oltvai, Nature Reviews Genetics **5**, 101 (2004).
[2] S. N. Dorogovtsev and J. F. F. Mendes, *Evolution of Networks—From Biological Nets to the Internet and WWW* (Oxford University Press, New York, 2003).
[3] J. M. Carlson and J. Doyle, Phys. Rev. Lett. **84**, 2529 (2000).
[4] L. Li, D. Alderson, R. Tanaka, J. Doyle, and W. Willinger, cond-mat/0501169.
[5] L. Li, D. Alderson, J. Doyle, and W. Willinger, in *Proceedings of ACM SIGCOMM 2004* (AMC Press, New York, 2004).
[6] http://gcrg.ucsd.edu/downloads (all conclusions are essentially the same for the larger S-matrix of *E. Coli*).
[7] M. Csete and J. Doyle, Trends Biotechnol. **22**, 446 (2004).
[8] H. Ma and A.-P. Zeng, Bioinformatics **19**, 1423 (2003).
[9] B. Mandelbrot, *Fractals and Scaling in Finance* (Springer, New York, 1997).
[10] W. Willinger, D. Alderson, J. Doyle, and L. Li, in *Proceedings of the 2004 Winter Simulation Conference* (IEEE, Piscataway, 2004)
[11] M. E. Csete and J. C. Doyle, Science **295**, 1664 (2002).
[12] M. Arita, Proc. Natl. Acad. Sci. U.S.A. **101**, 1543 (2004).
[13] J. M. Carlson and J. Doyle, Proc. Natl. Acad. Sci. U.S.A. **99**, 2538 (2002).