

Does Selection Mold Molecular Networks?

Andreas Wagner

(Published 30 September 2003)

Signal transduction pathways, transcriptional regulation circuits, and metabolic pathways are all part of the large molecular interaction networks that sustain life. We know that natural selection has influenced many features of living organisms, both on the level of individual genes and on the level of whole organisms. Has natural selection also molded the structure of molecular networks, either on the largest scale, such as that of the genome-scale protein interaction network shown in Fig. 1A, or on a smaller scale, such as that of the simple transcriptional regulation circuit (Fig. 1B)? Answers to these questions may yield insights into the design principles of molecular networks, as similar questions have for genes and whole organisms.

Aided by small-scale gene-by-gene analyses, functional genomics techniques have produced a great wealth of information about molecular networks. This information is mostly qualitative. It tells us how many genes and which genes a transcriptional regulator regulates, which proteins are part of a protein complex, and which metabolic reactions occur in an organism. However, functional genomics does not provide the fine-grained information, such as association constants and reaction rates, that traditional biochemical methods provide. It gives us a crude qualitative look at the topology of whole networks, exemplified by the protein interaction network shown in Fig. 1.

The topology of a molecular network can be viewed as a feature of an organism like any other. It raises the same basic questions: What is the network's structure? And why does it have

this structure? Most current work is devoted to addressing the first question. On the most coarse level, the description of a molecular network is straightforward: We can characterize it in terms of global statistics, such as the number of interaction partners per protein, the function of the different molecules in the network, or the number of genes regulated by transcription factors. However, on a finer scale, such description already re-

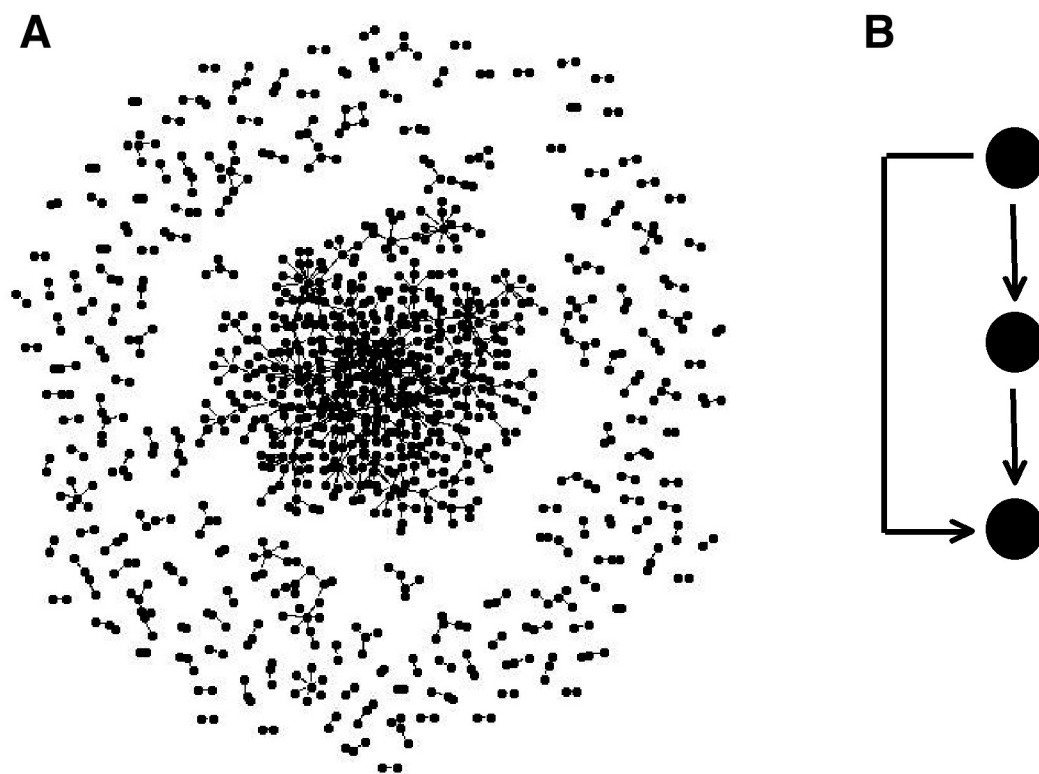


Fig. 1. On what, if any, level of organization does selection influence molecular network structure? Is it on the level of whole genome-scale networks, such as the protein interaction network shown in (A), or on the level of smaller elements, such as that of a transcriptional regulation circuit shown in (B)? This circuit is a feed-forward loop of transcriptional regulation (15), where the expression of the bottom gene is regulated by two transcription factors (middle and top), one of which regulates the other's expression as well.

veals a bewildering degree of complexity. For instance, the number and combination of protein domains that can mediate protein interactions in signal transduction networks are large and growing (1). These domains can be shuffled among proteins in an almost haphazard way. Such network descriptions reveal the tremendous complexity of molecular networks, but they do not go far in revealing underlying principles of their design.

The ensuing question of “why do molecular networks have their structure” could have two principal answers. The structure of molecular networks might reflect their history, much like the jumble of streets in a medieval city reflects the city's growth over centuries. A precedent comes from studies of metabolism.

University of New Mexico, Department of Biology, 167A Castetter Hall, Albuquerque, NM 817131-1091, USA. Telephone, 505-277-2021; fax, 505-277-0304; e-mail, wagnera@unm.edu

The oldest and most central parts of intermediary metabolism in heterotrophic organisms—glycolysis and the tricarboxylic acid cycle—originated earliest in evolution. Subsequently, many chemical reactions were added to this core, so that the most peripheral reactions tend to be those added most recently (2, 3).

The second possibility is that molecular networks have to have a certain structure because this structure is optimally suited to the network's biological function. Only in this case will network topology, both on a large and a small scale, reveal design principles of molecular networks. Two fundamentally different approaches can reveal such design principles. The first involves direct experimentation, often in combination with quantitative modeling, to ask what kind of advantages particular features of a molecular network might convey. To give but one example, cascades of several protein kinase reactions are part of many signal transduction pathways. In such cascades, one kinase serves as the substrate for the next kinase in the pathway. The most prominent example is the mitogen-activated protein kinase (MAPK) cascade, in which a series of three kinases communicates signals from the cytoplasmic membrane to the nucleus. This cascade is involved in a vast number of biological processes as different as neuronal plasticity, maturation of immune cells, and osmoregulation. Is the abundance of this and other kinase cascades a mere accident of evolutionary history, or do such cascades have features beneficial to the reliable transmission of signals? A combination of modeling and experimental work has shown that the MAPK cascade can show a highly cooperative or switch-like response to an input signal, even though its individual parts do not show such cooperativity (4). Such switch-like cooperativity means that the signal transduction pathway is not sensitive to noise (random fluctuations) over a wide range of the input signal's intensity.

This experimental approach has two principal limitations. It requires detailed biochemical knowledge about molecular interactions in part of a network, which is exactly the kind of knowledge that genome-scale functional genomics data do not provide. It can thus only be applied to either small parts of a network like the above cascade, or to networks of moderate size whose workings have been studied intensely over decades with conventional genetic and biochemical methods. Examples of such networks include the lysogeny-lysis switch in bacteriophages, segmentation genes in *Drosophila*, and flower development genes in plants. However, because of the thousands of human years involved in characterizing a molecular network in detail, there are very few such well-studied networks. The second shortcoming is that this approach leaves lingering doubts that a network feature that is seemingly ideal for a purpose, such as switch-like behavior, could not be achieved by different means. The MAPK cascade illustrates this caveat. A kinase cascade is not the only process that can achieve switch-like activation of a target molecule. A single kinase, not an entire cascade, may suffice for switch-like activation of a target molecule if the target molecule has multiple sites that must all be phosphorylated for activation (5). Detailed functional characterization of network features may thus be ideal for providing suggestions, but not conclusions, about optimal network design.

An approach that directly identifies features of a network that are influenced by natural selection could alleviate this problem. The idea has only one major catch: It is much easier to postulate that selection is shaping a network feature than to prove it. A case in point is a recent hypothesis suggesting that natural selection has affected

the degree distribution of metabolic networks and protein interaction networks; that is, the distribution of the number of interaction partners of a molecule in a network. It is based on the observation that in protein interaction networks, metabolic networks, and transcriptional regulation networks, the degree distribution of each network node often has a broad-tailed distribution. In some such networks, this distribution takes the form of a power law, where the probability $P(d)$ that a randomly chosen node has d immediate neighbors is proportional to $d^{-\gamma}$, where γ is a constant characteristic for the network (6-8). In networks with this property, the mean distance between network nodes that can be reached from each other (via a path of edges) is very small and it increases only very little upon random removal of nodes (6). This distance can be thought of as a measure of how "compact" a network is. In graphs with other degree distributions, network compactness can increase substantially upon node removal. These observations have led to the proposition that robustly compact networks confer some (unknown) advantages on cells, and that the frequent power-law degree distribution reflects the action of natural selection on the degree distribution itself. Although intriguing, this hypothesis runs into several problems. First, broad-tailed degree distributions are found in chemical reaction networks that, unlike metabolic networks, have never been under the influence of natural selection to begin with (7). This suggests that such degree distributions may be a general feature of chemical reaction networks and that their emergence does not require natural selection. Second, for protein interaction networks, turnover of individual protein interactions on evolutionary time scales, without natural selection shaping the network's global structure, is sufficient to explain the network's structure (9, 10). A final problem with the hypothesis that power laws reflect selection on robust compactness emerges from an important corollary of it. If the hypothesis were correct, then highly connected nodes in a network should be more important, in the sense that their mutation or outright removal, which change network compactness drastically, should have more severe effects on an organism's fitness than the mutation or removal of nodes with low connectivity (8). However, studies that ask whether highly connected proteins in the yeast protein interaction network can tolerate few mutations and thus evolve slowly provide no support for this prediction (11, 12). In addition, an association between a node's connectivity and importance could exist for a variety of reasons other than a node's effect on network compactness. For instance, highly connected nodes may simply act in a greater variety of biological processes.

Another recent evolutionary hypothesis has yielded more promising results. It regards not the large-scale structure of networks, but their smallest parts: patterns of interactions between only a few network nodes, like that shown in Fig. 1B (13). The example in this figure shows a feed-forward loop of transcriptional regulation, where the expression of the bottom gene is regulated by two transcription factors (middle and top), one of which regulates the other's expression as well. The hypothesis postulates that if any such pattern of interactions has favorable properties, it should be found more often than would be expected by chance alone in a molecular network. Recent analyses of transcriptional regulation networks of both the yeast *Saccharomyces cerevisiae* and the bacterium *Escherichia coli* support this hypothesis. They reveal that several regulatory circuits, including that shown in Fig. 1B, are orders of magnitude more abundant than would be expected by chance alone; that is, in a random network of identical size and degree distribution (14-16). Their large number—48 instances of the feed-forward loop in yeast—suggests

that natural selection favors regulatory circuits of this structure, so that they accumulate in a genome over millions of years. However, it is also possible that the abundance of these circuits is an accident of history. Specifically, these circuits may have originated through the duplication of a smaller number of ancestral circuits. This possibility is not farfetched, given the abundance of duplications of individual genes, chromosome segments, and whole genomes in many organisms. However, a detailed analysis shows that almost all of the identified abundant transcriptional regulation circuits are the result of convergent evolution; that is, they are not derived from some ancestral circuit (17). In addition, analysis of functional properties of some circuits supports the notion that they have desirable properties. For instance, the design of a feed-forward loop serves to activate the regulated (downstream) genes only if the topmost regulator is persistently activated, thus ensuring reliable transcriptional activation in the face of random fluctuations in regulator concentrations (15).

This last example suggests that, in the long run, evolutionary analyses of molecular networks must complement functional analyses to elucidate network design principles. Put differently, Dobzhansky may once again be proven right in his oft-quoted statement that “nothing in biology makes sense except in the light of evolution” (18). Although evolutionary studies of molecular networks are only in their beginning, there is no shortage of design features in molecular networks that are in want of an explanation. They range from the smallest network features (Fig. 1B) to intermediate-scale features [such as the abundance of cyclic structures (7) and clusters of highly connected molecules] to the coarsest, most global features, such as a network’s compactness.

References

1. T. Pawson, P. Nash, Assembly of cell regulatory systems through protein interaction domains. *Science* **300**, 445-452 (2003).
2. H. J. Morowitz, *Beginnings of Cellular Life* (Yale Univ. Press, New Haven, CT, 1992).
3. A. Wagner, D. Fell, The small world inside large metabolic networks. *Proc. R. Soc. London Ser. B* **280**, 1803-1810 (2001).
4. C. Huang, J. Ferrell, Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10078-10083 (1996).
5. P. Nash, X. Tang, S. Orlicky, Q. Chen, F. B. Gertler, M. D. Mendenhall, F. Sicheri, T. Pawson, M. Tyers, Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514-521 (2001).
6. R. Albert, H. Jeong, A. L. Barabasi, Error and attack tolerance of complex networks. *Nature* **406**, 378-382 (2000).
7. P. M. Gleiss, P. F. Stadler, A. Wagner, D. A. Fell, Small cycles in small worlds. *Adv. Complex Syst.* **4**, 207-226 (2001).
8. H. Jeong, S. P. Mason, A.-L. Barabasi, Z. N. Oltvai, Lethality and centrality in protein networks. *Nature* **411**, 41-42 (2001).
9. A. Wagner, How the global structure of protein interaction networks evolves. *Proc. R. Soc. London Ser. B* **270**, 457-466 (2003).
10. R. V. Sole, R. Pastor-Satorras, E. D. Smith, T. Kepler, A model of large-scale proteome evolution. *Adv. Complex Syst.* **5**, 43-54 (2002).
11. I. K. Jordan, Y. I. Wolf, E. V. Koonin, No simple dependence between protein evolution rate and the number of protein-protein interactions: Only the most prolific interactors tend to evolve slowly. *BMC Evol. Biol.* **3**, 1 (2003).
12. H. B. Fraser, D. P. Wall, A. E. Hirsh, A simple dependence between protein evolution rate and the number of protein-protein interactions. *BMC Evol. Biol.* **3**, 11 (2003).
13. U. Alon, Biological networks: The tinkerer as an engineer. *Science* **301**, 1866-1867 (2003).
14. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, *Science* **298**, 824-827 (2002).
15. S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat. Genet.* **31**, 64-68 (2002).
16. T. I. Lee, N. J. Rinaldi, F. Robert, D. T. Odom, Z. Bar-Joseph, G. K. Gerber, N. M. Hannett, C. T. Harbison, C. M. Thompson, I. Simon, J. Zeitlinger, E. G. Jennings, H. L. Murray, D. B. Gordon, B. Ren, J. J. Wyrick, J.-B. Tagne, T. L. Volkert, E. Fraenkel, D. K. Gifford, R. A. Young, Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799-804 (2002).
17. G. C. Conant, A. Wagner, Convergent evolution of gene circuits. *Nat. Genet.* **34**, 264-266 (2003).
18. T. Dobzhansky, Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teacher* **35**, 125-129 (1973).

Citation: A. Wagner, Does selection mold molecular networks? *Sci. STKE* **2003**, pe41 (2003).