

# **Longest paths and cycles in signal transduction and transcriptional regulation networks: a signature of natural selection?**

Andreas Wagner<sup>1</sup> and Jeremiah Wright,  
Department of Biology, The University of New Mexico

<sup>1</sup>Corresponding author:

Department of Biology,  
MSC03 2020  
1 University of New Mexico,  
Albuquerque, NM, 87131-0001  
email: [wagnera@unm.edu](mailto:wagnera@unm.edu)  
FAX: (505) 277-0304

## **Abstract**

We ask whether natural selection has shaped three biologically important features of 15 signal transduction networks and 2 genome-scale transcriptional regulation networks. These features are regulatory cycles, the lengths of the longest pathways through a network – a measure of network compactness –, and the abundance of node pairs connected by many alternative regulatory pathways. We determined whether these features are significantly more or less abundant in biological networks than in randomized networks with the same distribution of incoming and outgoing connections per network node. We find that autoregulatory cycles are of exceptionally high abundance in transcriptional regulation networks. All other cycles, however, are significantly less abundant in several signal transduction networks. This suggests that the multistability caused by complex feedback loops in a network may interfere with the functioning of such networks. We also find that several of the networks we examine are more compact than expected by chance alone. This raises the possibility that the transmission of information through such networks, which is fastest in compact networks, is a biologically important characteristic of such networks.

## Introduction

Metabolic, regulatory, and molecular interaction networks have been characterized in multiple organisms (7, 10, 11, 13, 16, 19, 21, 25, 30, 34). However, we still know very little about the design principles of such networks. On what level of organization does natural selection shape the structure of such networks, on that of the entire network which may comprise between dozens and thousands of genes, on that of local neighborhoods around individual genes, or on the smallest level, that of individual network genes and their interactions with other genes? With few exceptions, we have no empirically founded answers to this question. One such exception regards the protein interaction network of the yeast *Saccharomyces cerevisiae*. Here, interactions between highly connected proteins (proteins with many interaction partners) occur much less frequently than expected by chance alone. In contrast, interactions between highly connected proteins and lowly connected proteins occur much more frequently than expected by chance alone (23). To arrive at this statement, one compares the connectivity of proteins between the experimentally observed protein interaction network, and random networks in which each protein has the same number of interaction partners as in the yeast protein interaction network (23). The statistically significant pattern suggests that natural selection suppresses interactions between highly connected nodes. Unfortunately, it does not tell us why.

Other limited evidence for the influence of natural selection on network structure comes from transcriptional regulation networks (5, 15, 21, 32, 34). Here, short regulatory motifs involving few transcriptional regulators and their target genes are much more abundant than expected by chance alone (21, 32). Examples include a so-called feedforward loop where a transcriptional regulator  $R_1$  regulates the expression of a regulator  $R_2$ , which regulates the expression of a target gene  $T$ . In addition,  $R_1$  also regulates the expression of  $T$  directly. Experimental and theoretical evidence suggests that such motifs can serve specific biological functions, such as the suppression of gene expression noise (22). In addition, they appear to have arisen not from a common ancestral motif but independently through convergent evolution (5). Both of these findings support the notion that natural selection has contributed to the abundance of such network motifs. Other current claims that the structure of genetic networks is optimized by natural selection are speculative and may not hold up to closer scrutiny. A case in point is the notion that the power-law distribution of connectivity often observed in genetic networks (17, 18) results from an evolutionary optimization of network “robustness” to mutations. The very ubiquity of power-law connectivity distributions in networks that have not been under the influence of natural selection sheds doubt on this notion (2).

Aside from connectivity correlations and small regulatory motifs, biological networks have several structural features that natural selection might have influenced. We here examine 15 different signal transduction networks and two large transcriptional regulation networks for the presence of several such features. We represent these networks as directed graphs, in which directed edges point from a regulatory molecule to its regulatory targets, molecules whose concentration or activity the regulators influence. We take the above null-hypothesis approach, where the structure of a network is compared to that of randomized networks with the same number of incoming and

outgoing connections for each node (23). Despite its limitations (3, 24), this is currently the most promising approach to identify candidates for important network features, even though it needs to be supplemented by other kinds of evidence to be conclusive.

For our analysis, we chose two maximally different kinds of networks about which systems biology is generating information. The first kind, transcriptional regulation networks (21, 34), are large (genome-scale) networks about which single functional genomics experiments have revealed much information. In these networks, one kind of regulatory molecule, a transcriptional regulator, influences the expression of its target genes. Signal transduction networks, in contrast, are much smaller, containing of the order of dozens of genes. Their structure is painstakingly elucidated by thousands of man-years of experimentation in different laboratories. In addition, signal transduction networks include not only one but multiple levels of regulation: Their regulatory moieties may be proteins, but also small molecules or even ions. We here use information curated by experts on 15 different signal transduction networks that are important in processes as different as slime mold aggregation and mammalian development (1). We are acutely aware of the potential shortcomings of such data – especially nonuniform representation due to different expert “styles” – but absent standardized representations of signal transduction networks, such data provides currently the only viable avenue to analyze the structure of signal transduction networks. Below, we analyze two, transcriptional regulation networks and 15 signal transduction networks for the abundance of three features, cycles indicating feedback regulation, the lengths of the longest paths, and the incidence of nodes with paths between them.

## Materials and Methods

**Signal transduction networks.** The Science Signal Transduction Knowledge Environment (<http://stke.sciencemag.org/cgi/cm>) contains a collection of signal transduction pathways manually assembled by experts on these networks. We analyzed the structure of all 15 signal transduction networks with more than 30 nodes that were available in this repository in May 2004. These networks are the adrenergic pathway (<http://www.stke.org/cgi/cm/>; CMP\_8762), a network that mediates the responses of cells to epinephrine and norepinephrine; the *Dictyostelium discoideum* cAMP chemotaxis network (CMP\_7918), which is involved in the aggregation of cells in response to starvation; the differentiation pathway in PC12 cells (CM\_8038), a network that mediates the differentiation of a rat adrenal tumor cell line under the influence of nerve growth factor; the extracellular signal regulated kinase 1 and 2 (ERK1/2, or mitogen associated kinase [MAPK] p42 and p44) network (CMP\_10705), the c-Jun N-terminal MAPK network (CMP\_10827), and the p38 MAPK network (CMP\_10958), which are activated by a variety of mitogenic stimuli, differentiation signals, and cellular stresses; the B and T lymphocyte receptor signaling network (CMP\_6909 and CMP\_7019), which mediate the response of B and T cells to antigens and antigen-presenting cells; the networks that mediate the action of  $G_{\alpha 13}$  (CMP\_8809) and  $G_{\alpha i}$  (CMP\_7430), two variants of the  $\alpha$ -subunit of heterotrimeric guanine nucleotide binding proteins (G-proteins), which have innumerable functions in cell biological processes; the insulin signaling network (CMP\_12069), which modulates the storage and release of energy after nutrient deprivation and nutrient uptake; the mammalian Toll-like receptor networks

(CMC\_8644), which are involved in the inflammatory response of tissues to microbial infections; the Wnt/ $\beta$ -catenin network (CMP\_5533), which influences cell proliferation and other aspects of cell behavior in vertebrates and invertebrates through Wnt proteins, which are secreted glycoproteins; the FAS signal transduction network (CMP\_7966), one of whose functions is to trigger apoptosis; and the Integrin signaling network (CMP\_6880), which senses the environment in the extracellular matrix and are necessary for cell migration, growth, and survival. Note that, as opposed to transcriptional regulation networks, nodes in all of these networks are heterogeneous: They can represent proteins, small molecules, or ions. A directed edge links node A to node B if A influences the concentration or activity of B.

**Transcriptional regulation networks.** For our analysis of the transcriptional regulation network of the yeast *Saccharomyces cerevisiae*, we used data on likely transcriptional regulatory interactions obtained from a genome-scale chromatin immunoprecipitation analysis (20, 21). In this experiment, 106 epitope-tagged transcriptional regulators were used in three replicate chromatin immunoprecipitation experiment to identify genomic DNA to which these regulators were bound (26). The immunoprecipitated DNA was hybridized to DNA microarrays containing the regulatory regions upstream of known yeast genes. The fluorescence intensity of a spot (regulatory region) on the array indicates the binding strength of a transcriptional regulator to the regulatory region. This indication of binding is quantitative, but for many analyses, a qualitative (all-none) indication of binding and transcriptional regulation is more useful. The authors thus developed an error model of binding that allowed them to assign a probability or P-value of binding for each transcriptional regulator to a gene's regulatory region (21). This P-value indicates the confidence one has in a factor's binding to a specific DNA region. We here generally follow the authors' suggestion of equating *bona fide* binding of a transcriptional regulator to a target gene if this P-value is smaller than  $10^{-3}$ . This value minimizes the number of false-positive binding interactions, while maximizing the number of true positive regulator-target binding interactions (21).

For the transcriptional regulation network of *Escherichia coli*, we used a database of direct transcriptional interactions published by Shen-Orr and collaborators (32). This database was compiled from an existing database (RegulonDB) and an extensive literature search; it contains 578 transcriptional regulation interactions among 423 genes or operons, of which 116 encode regulators.

## Results

**Cyclic structures.** Regulatory cycles or feedback loops have long been recognized as important for biological networks (8, 29). They were perhaps first appreciated in metabolic networks, where the end product of a metabolic pathway can inhibit an upstream reaction and thus repress its own production (33). Cycles also occur in gene regulation networks, where they may endow a gene or network with robustness to environmental change or intracellular noise. Alternatively, regulatory cycles can cause a biological network to adopt one of multiple alternative states (9, 12). The simplest possible cycles are autoregulatory cycles, in which a regulator affects its own concentration or activity.

The importance of cyclic structures raises the question whether they are more abundant in biological networks than one might expect chance alone. We asked this question for the transcriptional regulation networks of *E. coli* and yeast, as well as for 15 signal transduction networks, each of them with at least 30 regulatory molecules (see Materials and Methods). To this end, we created from each of these networks 1000 randomized networks, using the following procedure(1). To create a randomized network, we select two directed edges at random from the network. These edges can connect up to four different nodes A, B, C, and D ( $A \rightarrow B$ ,  $C \rightarrow D$ ). If node A is different from node C and node B is different from node D, then we swap the two edges between the nodes ( $A \rightarrow D$ ,  $C \rightarrow B$ ). We repeat this procedure until  $4n$  edge pairs have been swapped, where  $n$  is the number of edges in the network. (Note that the above condition on node differences does not require that A is unequal to B, nor that C is unequal to D. This means that autoregulatory (self-)loops are legitimate edges and can be created or destroyed in the randomization process.) We then compared the number of cycles that occur in the biologically realized networks with the distribution of the number of cycles in the randomized network. In general, exhaustive counting of cycles in large graphs may be prohibitive computationally, requiring alternative means to estimate the abundance of cycles (14). However, because of the moderate size of our networks, we were able to determine the number of cycles in them exhaustively.

In our analysis, we distinguish between cycles of length one, i.e., autoregulatory loops, and cycles of length greater than one. None of the signal transduction networks show a significant difference of autoregulatory cycles to randomized networks. However, 7 out of 15 signal transduction networks show a significantly smaller (!) number of cycles between length two and ten than the actual network. These are the B-cell antigen receptor network (2 cycles,  $P=0.002$ ), the  $G_{\alpha 13}$  network (2 cycles,  $P=0.033$ ), the Toll-like receptor network (6 cycles,  $P=0.044$ ), the ERK1/2 MAP kinase network (no cycles,  $P < 10^{-3}$ ), the c-Jun N-terminal MAPK network (0 cycles,  $P < 10^{-3}$ ), the p38 MAPK network (0 cycles,  $P=10^{-3}$ ), and the Integrin signaling network (52 cycles,  $P=0.009$ ). The only exception is the FAS signaling network, which shows more cycles than expected by chance alone (1298 cycles,  $P=0.003$ ). Cycle data on the six networks with the most significant deviation from the random expectation are shown in Figure 1.

In contrast to the signal transduction networks, autoregulatory cycles are overabundant in the transcriptional regulation network of both *E. coli* (59 cycles,  $P < 10^{-3}$ ) and yeast (12 cycles,  $P < 10^{-3}$ ; Figure 2). However, no cycles of any other length are overabundant or underabundant in either the yeast (e.g.,  $P=0.46$  for cycles with length between 2 and 10) or the *E. coli* ( $P=0.22$ ) transcriptional regulation network.

**Network Compactness.** In many biological and other networks, the number of edges connecting any two nodes is small (37). One could loosely call such networks highly “compact”, because they can be traversed very rapidly along any path of edges. Similar compactness is also observed in simple random networks, such as the classical Erdosz-Renyi network with a Poisson distribution of the number of interactions per node. High compactness means that information can rapidly spread from any one node to all other nodes connected to it. In biological networks, compactness might be advantageous, because it allows a network to react rapidly to perturbations, for example in the concentrations of regulatory molecules. This need has perhaps first been acknowledged for metabolic networks, where transition times between metabolic states in different

environments need to be minimized (4, 6, 31). These observations raise the question whether biological networks are compact beyond what would be expected by chance alone. We here addressed this question by asking whether the longest paths through a network are significantly shorter in biological than in randomized networks where every node has the same number of incoming and outgoing connections. More specifically, we determined the average length of the  $n$  longest paths through a network ( $n=1, 2, 5, 10,$  and  $15$ ) and compared it with the average length of the  $n$  longest paths in 1000 randomized networks.

Only a minority of biological networks is significantly differently compact than randomized networks. The networks where this difference is statistically most pronounced all are more compact than randomized networks. Figure 3a shows the P-values for the four networks that are significantly more compact for all  $n$  than randomized networks. These are the c-Jun N-terminal MAP kinase network ( $P < 10^{-3}$ ), the p38 MAP kinase network ( $P=0.006$ ), the Integrin network, and the *E. coli* transcriptional regulation network ( $P < 10^{-3}$ ). There are also two networks that show marginally significantly different compactness at some  $n$ . These are the ERK1/2 MAP kinase network, which is marginally significantly more compact, and the T-cell signal transduction network, which is marginally significantly less compact at some  $n$ . Figure 3b and 3c show the distribution of the average length of the  $n=15$  longest path for the two networks with the most significantly elevated compactness. These are the c-Jun N-terminal MAP kinase network (Figure 3b) and the *E. coli* transcriptional regulation network (Figure 3c).

**Alternative paths between node pairs.** Regulatory interactions between a regulatory molecule and its regulatory target need not be direct. That is, they may be mediated by one or more intermediate regulators. Moreover, there may be more than one regulatory path through a network connecting any one regulator to its target. It is possible that such alternative paths provide robustness to mutations in the intermediate regulators. That is, if many alternative paths connect a regulator to its target gene, a loss of function in one of the intermediate regulators may be compensated by one of the alternative pathways through the network. We have indeed recently found evidence for this phenomenon (36). Specifically, the yeast transcriptional regulation network is most robust to mutations in intermediate regulators that are part of many alternative pathways between a regulator-target pair: Such regulators evolve at faster rates, and have weaker effects on cell growth when eliminated. This observation raises the question whether natural selection has influenced the number of pathways between at least some regulator-target pairs.

We addressed this question by comparing the number of regulator-target pairs connected by many (10, 20, or 30) alternative pathways between an actual network and its randomized counterpart. The results are equivocal, because in some networks regulator-target pairs are over-represented, whereas in others they are under-represented. Specifically, we find that the Insulin signaling pathway has significantly more (101) regulator target pairs with more than 20 pathways between them than 1000 randomized networks ( $P=0.024$ ). Alternative pathways between some nodes are similarly abundant in the Wnt/ $\beta$ -catenin network (101 and 72 regulator-target pairs with more than 20 and 30 alternative pathways;  $P=0.007$  and  $P=0.001$ , respectively), and in the FAS signaling network (1409, 1270, and 1203 regulator-target pairs with more than 10, 20, and 30

alternative pathways;  $P=0.007$ ,  $0.005$ , and  $<10^{-3}$ , respectively.) In contrast, the Integrin signaling network has fewer regulator-target pairs with more than 10, 20, or 30 alternative pathways than expected (197, 132, and 94 such pairs;  $P=0.002$ ,  $0.006$ ,  $0.005$ , respectively). The c-Jun N-terminal MAPK network shows a slight underabundance of regulator-target pairs (36 & 12 pairs, respectively) with more than 10 and 20 alternative pathways ( $P=0.019$  and  $P=0.04$ , respectively). The p38 MAPK network has no regulator-target pairs with more than 10 pathways between them, which is encountered only in a fraction  $P=0.021$  of randomized networks. For the two transcriptional regulation networks we studied, no significant differences between the actual networks and their randomized counterparts existed.

## Discussion

There may be many as yet undiscovered features of biological networks that reflect design principles of such networks, features that have been shaped by natural selection. We have focused on three such features, two of which show a consistent pattern for a number of networks. These are the abundance of cycles and network compactness, as indicated by the average length of the  $n$  longest paths in the network ( $n \leq 15$ ). Both of these features have received some earlier attention. For example, it has been noticed that the *E. coli* and the yeast transcriptional regulation network contains many autoregulatory feedback loops, and that both the *E. coli* and the yeast transcriptional regulation network contain few long linear pathways compared to similar networks in other species (15, 27, 34). However, these studies did not answer in statistical terms whether one should be surprised by such patterns, which might occur in all or most networks of similar size and number of regulatory interactions. Our analysis shows that these features are of significantly different abundance only for some biological networks.

Perhaps the most striking case is the overabundance of autoregulatory loops of both the yeast and the *E. coli* transcriptional regulation networks. This pattern underscores the importance and usefulness of autoregulation, which plays an important role in switching genes stably on, suppressing intracellular noise, and changing response times to gene expression (12, 28, 29). Perhaps more surprising is the underabundance of all other cycles in signal transduction networks. Because cycles are important for feedback in many biological systems, one might think that they should also be abundant. However, cycles may not only provide for feedback regulation. They may also cause complex dynamical behavior, in particular multistability (9, 35). Complex dynamics and multistability may be either beneficial or detrimental for signal transduction. On one hand, multistability may be essential for a cell's adoption of a stable response to an extracellular signal. On the other hand, too many different states or too complex dynamical behavior may interfere with the robust functioning of a network. We speculate that the underabundance of cycles in some signal transduction networks may reflect a purging of cycles for this reason.

With one exception, network compactness, where significantly different from that in a randomized network, is significantly greater in biological networks. Long paths are suppressed in biological networks. We emphasize that this compactness goes beyond that found in small-world networks, which need not be more compact than random networks with a Poisson distribution of the number of interactions per node (37). What is

the reason for this pattern? Both signal transduction and transcriptional regulation networks need to respond to stimuli outside the cell. The rate at which information propagates through a network may be important for a rapid response, which would best be achieved by keeping pathways short. While this line of reasoning provides a rationale for the observed compactness, we realize that it is no proof, which could only come from appropriate experiments.

Our results also show that a network feature with a striking deviation from a random organization in one organism or network need not show the same deviation elsewhere. One case in point are the differences between the compactness of the transcriptional regulation networks of yeast and *E. coli*. In contrast to the *E. coli* network, the yeast network is no more compact than expected by chance alone. We do not know the reason for this discrepancy, although the profound differences in transcriptional regulation between prokaryotes and eukaryotes may be partly responsible. A second example is the overabundance of autoregulatory cycles in the *E. coli* and yeast transcriptional regulation networks, but the lack of such an overabundance in signal transduction networks. This may well be due to artifacts stemming from the representation and manual curation of such networks. (Only 2 of the 15 signal transduction networks we analyzed here show any autoregulatory cycles at all.) Third, only a minority of signal transduction networks showed statistically significant deviations from random organization. This raises the possibility that natural selection may influence some networks in ways too subtle to detect with this statistical approach. Alternatively, the global organization of these networks may be only marginally affected by natural selection, which may constrain the network only on a smaller level of organization.

The last observations point to some of the caveats of the approach we pursued here. First, like any other approach, its results depend on the quality of the available data. Genome-scale experiments on regulatory networks are subject to substantial experimental error. Manually curated data on smaller networks have the advantage that every interaction has been subject to extensive literature review, but the representation of the data leaves space for inappropriate choices. Second, qualitative data (“who interacts with whom”), may be too coarse to reveal much information of biological interest. It may be necessary to incorporate quantitative information on molecular interactions, which is very difficult to come by. The same holds for data on only one regulatory modality, such as in the transcriptional regulation networks we analyzed. For example, many feedback cycles involve modes of regulation different from and in addition to transcriptional regulation, such as allosteric inhibition of enzymes and posttranslational regulation. The final word on the abundance of any network feature is not spoken until a network representation incorporates such different modalities of regulation, as is the case for the signal transduction networks we studied. All of these caveats can only be addressed with improved data. We suggest that until such data becomes available, the approach we pursued here is best suited to identify candidate features for further investigation by different means. Cyclic structures and network compactness are such candidate features.

**Acknowledgments.** Support through the NIH grant GM63882 is gratefully acknowledged, as is continued support through the Santa Fe Institute.

## Figure captions

**Figure 1: Cycles (exclusive of autoregulation) in signal transduction networks.** Six signal transduction networks with significantly different numbers of cycles than 1000 randomized networks with the same distribution of incoming and outgoing edges. **a)** B-cell antigen receptor network; **b)** c-Jun N-terminal MAP kinase network; **c)** ERK1/2 MAP kinase network; **d)** FAS signaling network; **e)** Integrin network; **f)** p38 MAP kinase network. Shown is the distribution of the numbers of cycles with lengths between 2 and 10 for 1000 randomized networks. The arrow indicates the number of cycles in the actual biological network.

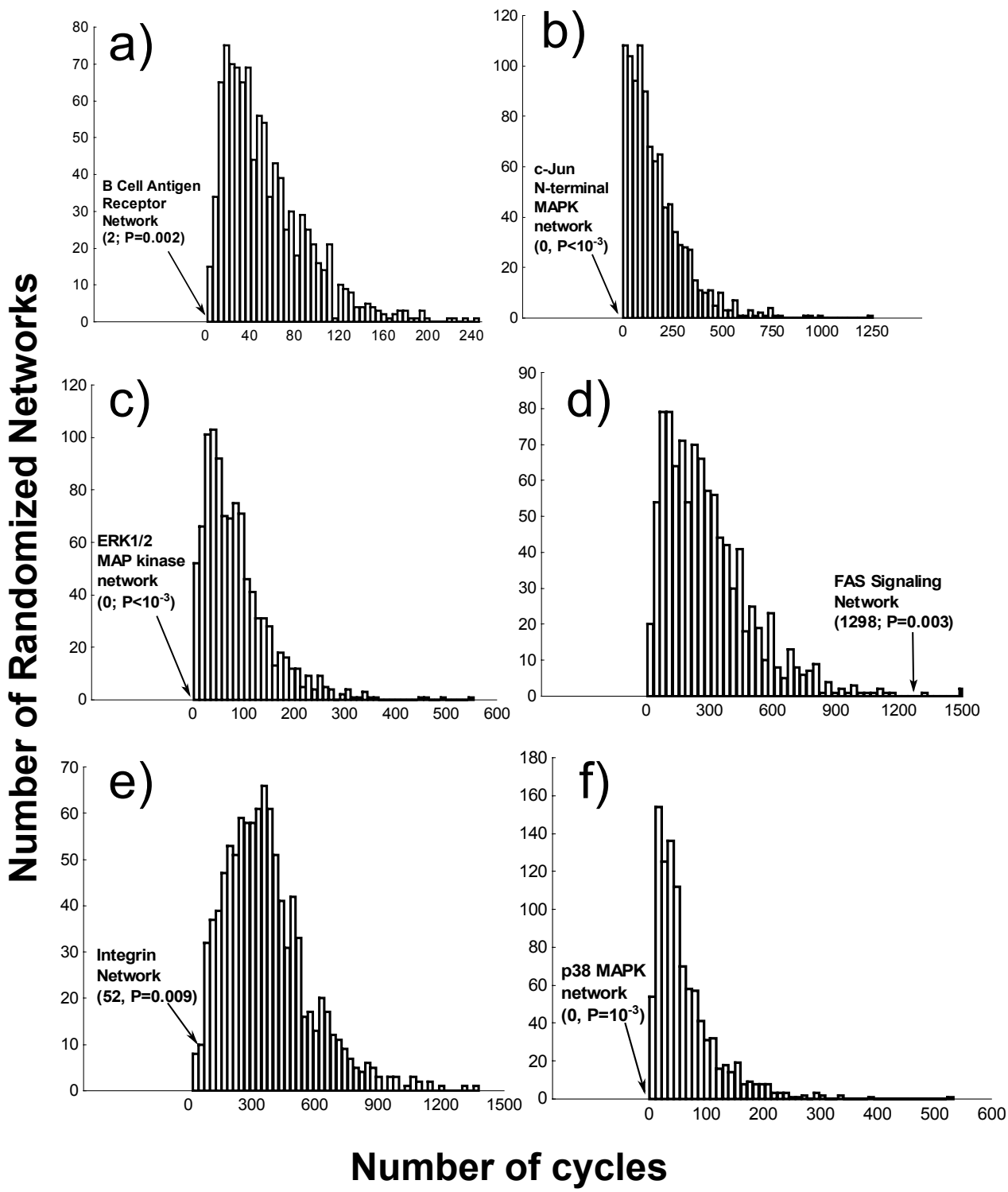
**Figure 2: Autoregulation cycles in transcriptional regulation networks.** Shown is the distribution of the numbers of autoregulatory cycles for 1000 randomized transcriptional regulation networks of *E. coli* (**a**) and yeast (**b**). The arrow indicates the number of cycles in the actual biological network.

**Figure 3: Network compactness.** **a)** Biological networks where the average length of the  $n$  longest pathways through the network differs significantly from that in random networks for all  $n$  shown. The P-values shown correspond to the fraction of randomized networks (among 1000 such networks) where this average length is greater than that or equal to that in the actual network (T-cell signal transduction network) or less than that or equal to that of the actual network (all 5 other networks). NS:  $P \geq 0.05$ ; **b)** Distribution of the average length of the 15 longest paths in the c-Jun N-terminal MAP kinase network; **c)** distribution of the average length of the 15 longest paths in the *E. coli* transcriptional regulation network.

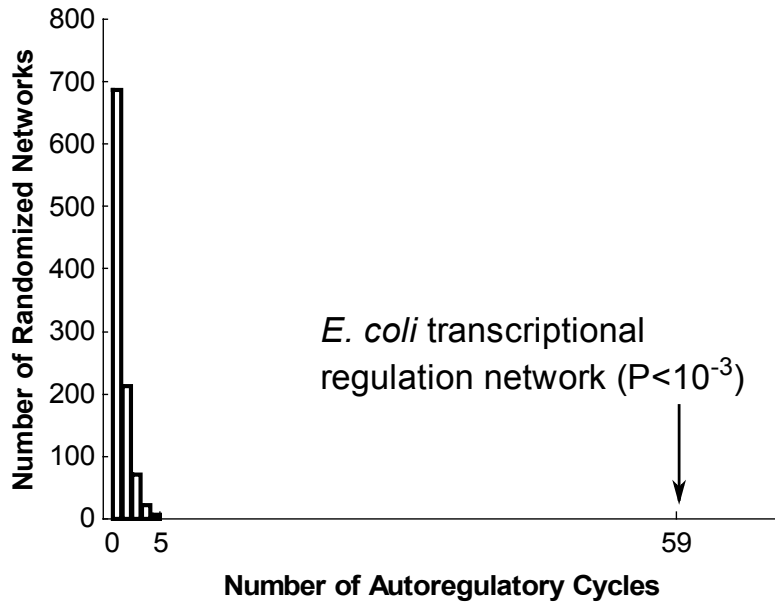
## Literature Cited

1. Science Signal Transduction Knowledge Environment (Sci. STKE); <http://stke.sciencemag.org/cgi/cm>.
2. Albert R, Barabasi AL. 2002. Statistical mechanics of complex networks. *Reviews of Modern Physics* 74: 47-97
3. Artzy-Randrup Y, Fleishman SJ, Ben-Tal N, Stone L. 2004. Comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks". *Science* 305
4. Cascante M, Melendez-Hevia E, Kholodenko BN, Sicilia J, Kacser H. 1995. Control analysis of transit-time for free and enzyme-bound metabolites - physiological and evolutionary significance of metabolic response-times. *Biochemical Journal* 308: 895-9
5. Conant GC, Wagner A. 2003. Convergent evolution in gene circuits. *Nature Genetics* 34: 264-6
6. Easterby JS. 1986. The effect of feedback on pathway transient response. *Biochemical Journal* 233: 871-5
7. Edwards JS, Palsson BO. 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: 5528-33
8. Fell D. 1997. *Understanding the control of metabolism*. Miami: Portland Press
9. Ferrell JE. 2002. Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability. *Current Opinion in Cell Biology* 14: 140-8
10. Flajolet M, Rotondo G, Daviet L, Bergametti F, Inchauspe G, et al. 2000. A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* 242: 369-79
11. Forster J, Famili I, Fu P, Palsson B, Nielsen J. 2003. Genome-scale reconstruction of the *Saccharomyces cerevisiae* metabolic network. *Genome Research* 13: 244-53
12. Freeman M. 2000. Feedback control of intercellular signalling in development. *Nature* 408: 313-9
13. Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, et al. 2002. Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415: 141-7
14. Gleiss PM, Stadler PF, Wagner A, Fell DA. 2001. Small cycles in small worlds. *Advances in complex systems*. 4: 207-26
15. Guelzim N, Bottani S, Bourguin P, Kepes F. 2002. Topological and causal structure of the yeast transcriptional regulatory network. *Nature Genetics* 31: 60-3
16. Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et al. 2002. Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415: 180-3
17. Jeong H, Mason SP, Barabasi A-L, Oltvai ZN. 2001. Lethality and centrality in protein networks. *Nature* 411: 41-2

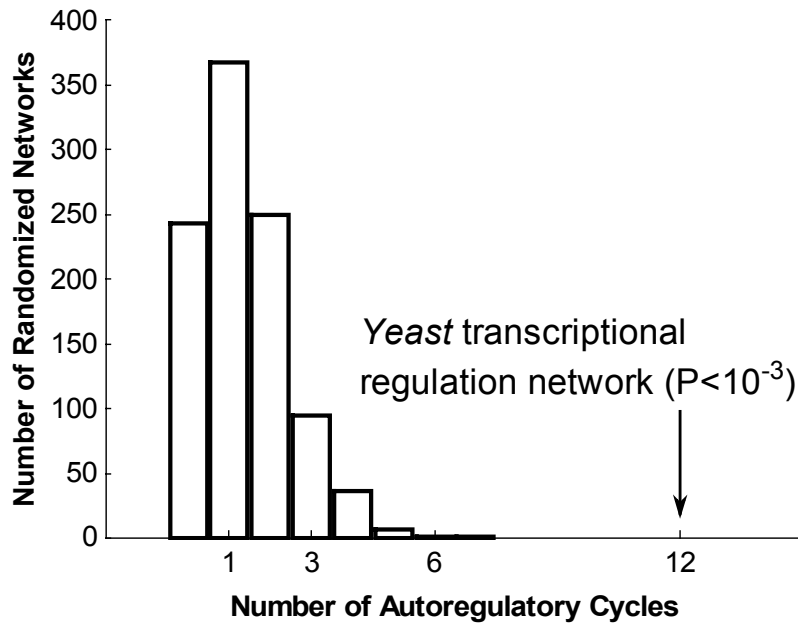
18. Jeong H, Tombor, B. Albert, R. Oltvai, Z.N., Barabasi, A.L. 2000. The large-scale organization of metabolic networks. *Nature*. 407: 651-4
19. Karp P, Riley M, Paley S, Pellegrini-Toole A, Krummenacker M. 1999. EcoCyc: electronic encyclopedia of E.coli genes and metabolism. *Nucleic acids research* 27: 55-
20. Knop M, Siegers K, Pereira G, Zachariae W, Winsor B, et al. 1999. Epitope tagging of yeast genes using a PCR-based strategy: More tags and improved practical routines. *Yeast* 15: 963-72
21. Lee T, Rinaldi N, Robert F, Odom D, Bar-Joseph Z, et al. 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298: 799-804
22. Mangan S, Alon U. 2003. Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the U.S.A.* 100: 11980-5
23. Maslov S, Sneppen K. 2002. Specificity and stability in topology of protein networks. *Science* 296: 910-3
24. Milo R, Itzkovitz S, Kashtan N, Levitt R, Alon U. 2004. Response to comment on "network motifs: Simple building blocks of complex networks" and "superfamilies of evolved and designed networks". *Science* 305
25. Rain JC, Selig L, De Reuse H, Battaglia V, Reverdy C, et al. 2001. The protein-protein interaction map of *Helicobacter pylori*. *Nature* ; 409: 211-5
26. Ren B, Robert F, Wyrick J, Aparicio O, Jennings E, et al. 2000. Genome-wide location and function of DNA binding proteins. *Science* 290
27. Rosenfeld N, Alon U. 2003. Response delays and the structure of transcription networks. *Journal of Molecular Biology* 329: 645-54
28. Rosenfeld N, Elowitz M, Alon U. 2002. Negative autoregulation speeds the response times of transcription networks. *Journal of Molecular Biology* 323: 785-93
29. Savageau MA. 1976. *Biochemical systems analysis: a study of function and design in molecular biology*. Reading, MA: Addison-Wesley
30. Schilling CH, Palsson BO. 2000. Assessment of the metabolic capabilities of *Haemophilus influenzae* Rd through a genome-scale pathway analysis. *Journal of Theoretical Biology* 203: 249-83
31. Schuster S, Heinrich R. 1987. Time Hierarchy in Enzymatic-Reaction Chains Resulting From Optimality Principles. *Journal of Theoretical Biology* ; 129: 189-209
32. Shen-Orr S, Milo R, Mangan S, Alon U. 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* 31: 64-8
33. Stryer L. 1995. *Biochemistry*. New York: Freeman
34. Thieffry D, Huerta A, Perez-Rueda E, Collado-Vides J. 1998. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *E. coli*. *BioEssays* 20: 433-40
35. Thomas R, D'Ari R. 1990. *Biological feedback*. Boca Raton, FL: CRC Press
36. Wagner A, Wright J. 2004. Alternative pathways and mutational robustness in molecular networks (submitted).
37. Watts DJ. 1999. *Small worlds*. Princeton, New Jersey: Princeton University Press



**a)**



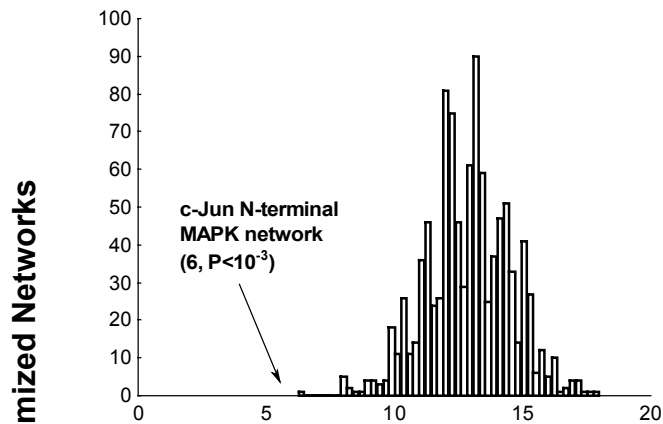
**b)**



a)

Network	P-value for average length of $n$ longest paths				
	n=1	n=2	n=5	n=10	n=15
c-Jun N-terminal MAP kinase	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>
p38 MAP kinase	0.006	0.006	0.006	0.006	0.006
Integrin signaling	0.011	0.011	0.009	0.009	0.007
<i>E. coli</i> transcriptional regulation	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>	<10 <sup>-3</sup>

b)



c)

