PHYSICAL REVIEW E **75**, 015101(R) (2007)

Design and statistical properties of robust functional networks: A model study of biological signal transduction

Pablo Kaluza,[1] Mads Ipsen,[2] Martin Vingron,[3] and Alexander S. Mikhailov[1]
[1]*Department of Physical Chemistry, Fritz Haber Institute of the Max Planck Society, Faradayweg 4-6, D-14195 Berlin, Germany*
[2]*Department of Chemistry, H. C. Ørsted Institute, University of Copenhagen, Universitetsparken 5, Copenhagen Ø2100, Denmark*
[3]*Department of Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Ihnestrasse 63-73, D-14195 Berlin, Germany*

(Received 10 March 2006; published 19 January 2007)

A simple flow network model of biological signal transduction is investigated. Networks with prescribed signal processing functions, robust against random node or link removals, are designed through an evolutionary optimization process. Statistical properties of large ensembles of such networks, including their characteristic motif distributions, are determined. Our analysis suggests that robustness against link removals plays the principal role in the architecture of real signal transduction networks and developmental genetic transcription networks.

DOI: 10.1103/PhysRevE.75.015101 PACS number(s): 89.75.Hc, 89.20.−a, 89.75.Fb

Cells in a biological organism function in a stable, precise way despite noise and destructive mutations [1]. If the principles of biological robustness were understood, they could be further applied for the design of complex industrial production and transportation systems, or for understanding social processes [2,3]. The cells operate as dynamical networks and their robustness is, to a large extent, determined by a special network architecture. This has been demonstrated for various biological functions, including chemotaxis [4], metabolism [5], signal transduction [6], and the cell cycle [7]. It was shown that genetic networks with robust expression patterns may spontaneously develop through biological evolution [8]. The problems of robustness have also been discussed in an abstract context for large random networks, aimed at developing optimal defense strategies for the Internet and the WWW [9–11]. The networks of a living cell are not random. They are selected by biological evolution to execute certain functions. In particular, a cell should activate a fixed group of genes in response to each arriving stimulus. These networks should maintain their *prescribed,* specific functions, although possibly exposed to random damage or parameter variations. The structure of such networks reflects their functions. Is it possible, by adjustment of the network structure, to develop systems with prescribed functions that are, furthermore, robust against damage? How strongly would the requirements of robustness against a particular kind of damage affect their architecture?

In this paper, we study a toy flow model of biological signal transduction. The network transports signals, applied to input nodes, through a number of middle redistribution nodes to a set of output nodes. In a cell, the analogy would be to a particular set of genes that are turned on upon arrival of a certain stimulus at the cell surface. This mapping between input (stimulus) and output (gene activity) is mediated by a network of interactions among proteins in the cell. These proteins are modeled as nodes in our networks, while interactions between them are reflected in the existence of links. Physically, a signal from the cell surface is passed on through processes like protein phosphorylation or dephosphorylation, translocation, structural change, etc. In a gross oversimplification of the real processes, we model this signal transduction process by an abstract network flow. Proteins undergo mutations which in turn can affect the links among them or completely delete some nodes and introduce new ones. Thus the network topology is subject to random local changes.

By running an optimization process with structural mutations and subsequent selection, we show that networks with predefined output patterns can be constructed. Then, we extend the optimization criterion and design networks which, while approximately retaining a fixed output pattern, become robust against removal of randomly chosen links or nodes. Statistical properties of robust functional networks, for an ensemble of different optimization trajectories starting with various initial conditions, are considered and distributions of structural motifs in two kinds of networks, robust against link or node removals, are then determined.

A considered network of size $N=N_{in}+M+N_{out}$ consists of N_{in} input nodes, M middle nodes, and N_{out} output nodes. Its architecture is specified by a directed graph of connections between the nodes with the adjacency matrix A_{ij} (we have $A_{ij}=1$, if there is a link from node j to node i, and $A_{ij}=0$ otherwise). An input node can be connected only with the middle nodes, a middle node can be connected with other middle nodes and with the output nodes (see Fig. 1). Each link $j\rightarrow i$ carries some *signal flux* u_{ij}. The sum of all incoming fluxes for any node is equal to the sum of all outgoing fluxes. For any node, all outgoing fluxes are equal in intensity and are obtained by splitting the total incoming signal flux in equal parts between the outgoing connections. Thus we have $u_{ik}=(\Sigma_l A_{lk})^{-1}\Sigma_j A_{kj}u_{kj}$ for any node k. Introducing the total fluxes $x_i=\Sigma_j A_{ji}u_{ji}$ passing through nodes i, this redistribution law can also be written as $x_i=\Sigma_j A_{ij}x_j(\Sigma_k A_{kj})^{-1}$ for $i=1,2,\ldots N$. External fluxes can be applied to the input nodes and sinks are attached to the output nodes. An external unit flux $x_\alpha=1$, applied to an input node $\alpha=1,2,\ldots,N_{in}$, becomes distributed after passing through the network and fractions $x_\beta=Q_{\alpha\beta}$ of the applied flux reach different output nodes $\beta=1,2,\ldots,N_{out}$. The matrix $\mathbf{Q}$ with the elements $Q_{\alpha\beta}$ represents the output pattern of a given network. Note that $\Sigma_\beta Q_{\alpha\beta}=1$. The performance $\mathbf{F}(G)$ of a given network G is its output pattern $\mathbf{Q}$, i.e., $\mathbf{F}(G)=\mathbf{Q}$. The ideal performance of a

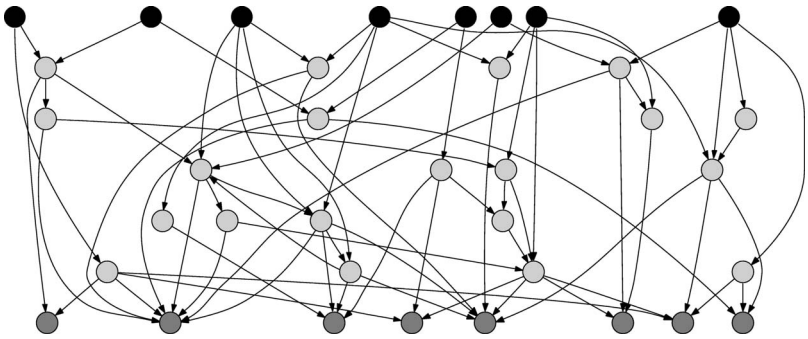

FIG. 1. Example of a flow distribution network.

network corresponds to some fixed output pattern $\mathbf{Q}_0$, specifying to which final destinations and in what amounts a particular kind of signal must be supplied. In a biological setting, the output pattern would correspond to a particular set of target genes responding to the stimulus, together with their strength of expression. A network G_0 is optimal if $\mathbf{F}(G_0)=\mathbf{Q}_0$. The distance ϵ of any network G from the ideal performance is $\epsilon=|\mathbf{F}(G)-\mathbf{Q}_0|$.

To characterize functional robustness of a network G, a set S of all networks, obtained by applying local damage (such as deleting a link or deleting a node, e.g., as a consequence of a mutation in a protein) to the network G, should be considered. This set can be viewed as the *damage shell* of the considered network. We introduce some tolerance threshold h, such that all networks with distances $\epsilon>h$ with respect to the ideal performance $\mathbf{Q}_0$ are *abortive*. Then, the robustness ρ of the network G is defined as the fraction of all networks in its damage shell which are *not* abortive,

$$\rho=\frac{1}{N_S}\sum_{G'\in S}\theta(h-|\mathbf{F}(G')-\mathbf{Q}_0|), \tag{1}$$

where N_S is the number of networks in the damage shell S and $\theta(z)$ is the step function, $\theta(z)=1$ for $z>0$ and $\theta(z)=0$ otherwise.

To construct a network with a prescribed output pattern $\mathbf{Q}_0$ which is robust against local random damage, the following optimization algorithm is implemented:

(i) At each iteration step, the flow error ϵ and the robustness of the network G are determined. The flow error is computed as $\epsilon=(2N_{in})^{-1}\sum_{\alpha=1}^{N_{in}}\sum_{\beta=1}^{N_{out}}(Q_{\alpha\beta}-Q^0_{\alpha\beta})^2$. To determine ρ, we consider a set of M networks G' obtained from G by deleting a single middle node. For each of them, the flow error ϵ' is calculated. The robustness ρ is the fraction of the networks G' with $\epsilon'<h$. Next, an evolutionary structural mutation is applied to the network G, yielding a new network G_1 with a different flow error ϵ_1 and a different robustness ρ_1. To decide, whether to accept this evolutionary mutation, we examine the value of ϵ_1 and the differences $\Delta\epsilon=\epsilon_1-\epsilon$ and $\Delta\rho=\rho-\rho_1$.

(ii) If the error of the new network is beyond the tolerance threshold ($\epsilon_1>h$), the decision is based on the flow errors. The mutation is always accepted, if $\Delta\epsilon\leqslant 0$ and accepted with probability $p=\exp(-\Delta\epsilon/\epsilon\sigma)$ if $\Delta\epsilon>0$. If the error of the new network is below the tolerance threshold ($\epsilon_1<h$), the decision is based on the robustness. The mutation is always accepted, if $\Delta\rho\leqslant 0$, and accepted with probability $p=\exp[-\Delta\rho/(1-\rho)\sigma]$, if $\Delta\rho>0$. Hence the parameter σ plays the role of temperature in our Metropolis algorithm.

(iii) If the mutation has been accepted, the network G is replaced by G_1 and the iteration step is repeated with the new network. If the mutation is rejected, the next iteration step is performed for the old network G.

The evolution is started with a random initial network G_0. Thus the optimization process is based on the flow error, while it remains larger than the tolerance threshold, and is switched to the selection based on the robustness, when the flow errors become sufficiently small. A similar optimization approach has been previously used to reconstruct symmetric networks from their Laplacian spectra [13].

We have tried several different mutation schemes and found that, in the considered optimization process, the best result is reached if an evolutionary mutation represents adding a randomly generated path from an input node to an output node to a given network or subtracting such a path. To specify a mutation, one input node, one output node, and k middle nodes (between 1 and M) are randomly chosen and linearly connected, thus forming a graph g. We further decide whether this graph g should be added to the existing graph G or subtracted from it (in the sense of operations with the graphs, see [14]). Under this mutation scheme, all middle nodes are preserved. During the evolution, we retain a node, even if it has no incoming connections (such nodes, however, are not taken into account when estimating robustness). In the final network, nodes without incoming connections are deleted. In our numerical investigations, networks with eight input and eight output nodes and with 20 (initial) middle nodes were used. The target output matrices $Q^0_{\alpha\beta}$ were generated at random. Each row α in such a matrix has $K<N_{out}$ nonzero elements, which are randomly located within the row. Their values are chosen at random, in such a way that their sum is equal to unity. Thus in response to the activation of an input node, the network always activates K output nodes, but the positions of these nodes and their activation degrees are randomly chosen. Figure 2 shows how the flow error and the robustness (dash curve) evolve with time during a typical optimization process. After the optimization based on the robustness was switched on at about $t_1=6486$, the robustness increased by a factor of more than 17, while keeping the error within the tolerance window.

To undertake a statistical analysis, many independent evolutions starting from randomly chosen initial networks were performed. In Fig. 3(a), the black diagram shows the histogram of flow errors ϵ in the ensemble of 100 networks obtained by running the evolution based only on the flow error.

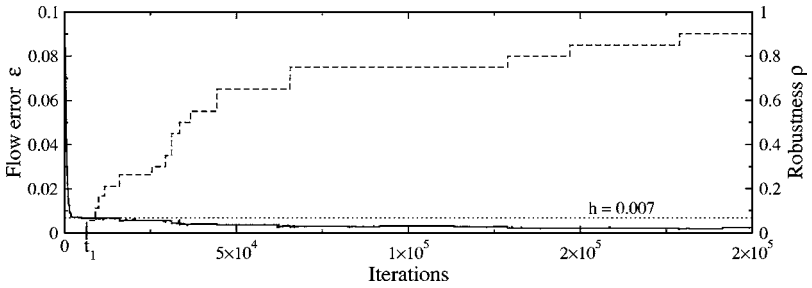

FIG. 2. Evolution of the flow error (solid line) and robustness (dash line) in the optimization process ($h=0.007$ and $\theta=0.0004$).

In the same diagram, the gray filled diagram shows the distribution of flow errors ϵ' in the networks of this ensemble after applying all possible local damages (deleting a middle node). While the flow errors ϵ of the selected networks remain all small, we see that the errors in the networks forming their damage shells are much larger and there is a long tail in their statistical distribution. In contrast to this, Fig. 3(b) displays analogous histograms for the evolution based on the robustness (with the tolerance threshold $h=0.007$). The distribution of flow errors (black) is almost the same as in Fig. 3(a), but the distribution of errors ϵ' after the application of local damages is strongly different. Now, most of the networks within a damage shell have errors below the tolerance threshold.

The local damage can also represent removal of a link, not a node. In this case, the robustness ρ of a given network is defined as the fraction of the networks, obtained by deleting one of its links, which would show an error ϵ' below the threshold. With this modification, the same optimization algorithm, as described above, can be employed. By running evolutions, based on this modified selection criterion, networks, which are robust against removal of links, could be constructed.

Statistical properties of the two classes of networks, robust against the removal of either nodes or links, have been compared for different tolerance thresholds. We have found that the mean degree of the networks, which are robust against the removal of nodes, is not significantly different from that of the networks, selected only on the basis of their flow error, whereas the networks robust against links removals have a much larger number of connections. The clustering coefficient is increased with respect to the purely functional networks for the networks robust against link removal and decreased for the networks robust against removal of nodes. The average path length, defined as the mean shortest path connecting input and output nodes, shows a similar behavior.

Distributions of structural motifs [12] have been determined for the designed networks robust against link and node removals (Fig. 4). To do this, numbers of different three-node subgraphs in a given network are first found. These are further compared with the frequencies at which various three-node motifs are present in the respective randomized graphs and the "scores" of different motifs for a given network are computed. A positive (negative) score means that this particular motif is found more (less) frequently in the considered graph than in its randomized version. In this way, any graph can be characterized by a certain structural motif diagram.

Milo *et al.* [12] have analyzed various real-world networks and discovered four superfamilies with distinct motif diagrams. The second superfamily included signal transduction networks and genetic developmental networks of multicellular organisms and the neural network of a nematode *C. elegans*. These networks share a common function of information processing, transmitting signals from input nodes via middle nodes to a set of output nodes. It was suggested that this common function should explain the observed convergence of motif distributions of these networks. Our study, based on a simple model of signal transduction, shows that this explanation is not sufficient. Analyzing large ensembles of functional networks, designed to implement various output patterns, we have found that the common function does not yet fix motif distributions, i.e., networks with the same output pattern may still have very different motif diagrams. Introducing additional evolutionary optimization of such functional networks, aimed at making them robust against link or node removals, we could see how motif diagrams of these two groups of the networks diverged during the evolution starting from the same initial ensemble and approached two different asymptotic forms when the optimization saturation has been reached. The final motif distribution of the networks robust against link removals [Fig. 4(a)] shows a striking agreement with the respective diagrams for biologi-

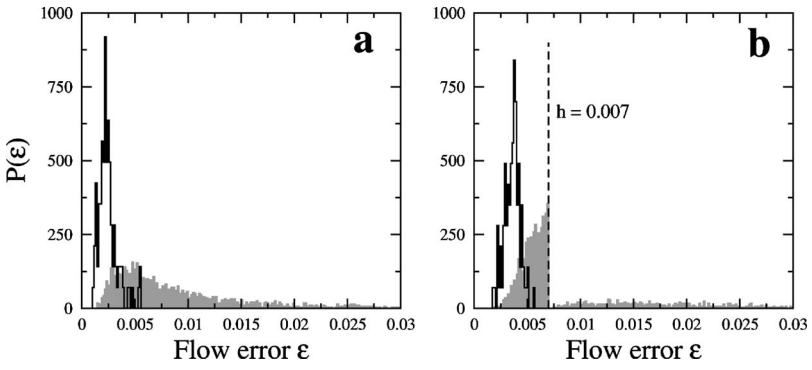

FIG. 3. Histograms of different flow errors in the ensembles of 100 networks (a) optimized only by flow error and (b) optimized for robustness against deletion of a node. Black diagrams show distributions of flow errors in the ensembles; gray filled diagrams show distributions of flow errors after application of various local damages. Optimization was run for 3×10^5 iteration steps for each network.

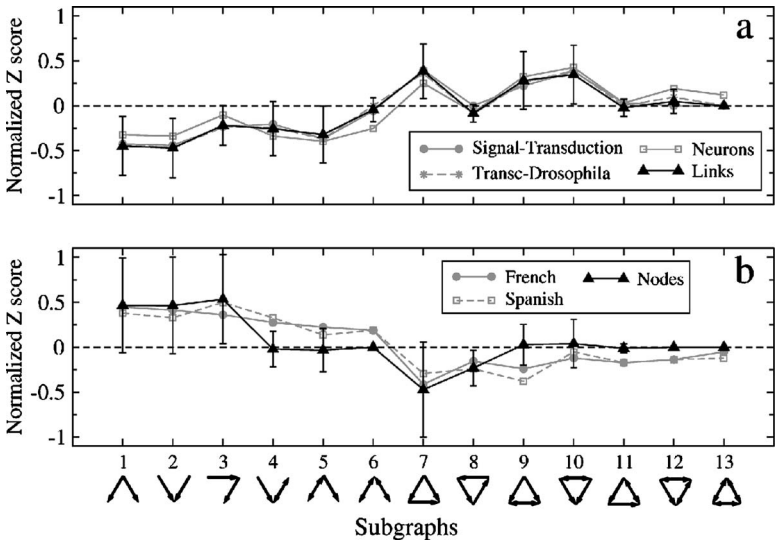

FIG. 4. Motif distributions of networks robust against removal of (a) (triangles) links and (b) (triangles) nodes. For each network type, an ensemble average over a subset of functional networks with the error less than $h=0.007$ in the ensemble of 100 independently evolved networks with different, randomly generated, output patterns is performed. Bars show statistical dispersion of data. For comparison, the data [12] for (a) signal transduction networks, developmental genetic transcription networks in drosophila, neural system of *C. elegans*, and (b) two linguistic networks are included (gray curves).

cal signal transduction networks and the genetic developmental network of *Drosophila*, while the motif distribution of the model networks robust against link removals resembles that of the linguistic networks in the fourth superfamily [Fig. 4(b)].

Proceeding from this analysis, we suggest that the principal role in determining the characteristic motif distribution in the second superfamily is played by the condition of robustness against link removals, imposed by biological evolution. Signal transduction networks and genetic developmental networks of multicellular organisms may have evolved to become robust against random disruption of interactions between proteins, rather than the removal of entire proteins (nodes) from these networks. In this respect, we note that a genetic point mutation modifies only one amino acid in the protein chain and this would typically lead only to a minor change in the folded protein conformation. Therefore it can be expected that, after such a mutation, a protein would lose only some of its interactions with other proteins in the network, corresponding to breaking of individual links in the model. Moreover, the neural system of *C. elegans* may have evolved to become robust against breaking of synaptic connections, rather than against the death of whole neural cells.

The spectacular agreement between motif distributions of real biological information processing networks and model networks, optimized to be robust against link removals, cannot be explained by purely qualitative arguments. It may indicate the existence of a general statistical *universality class* of such networks. The considered minimalistic model neglects many details of actual signal transduction and genetic regulation processes. Nonetheless, it is apparently able to correctly reproduce major structural properties of networks in the second superfamily.

Although the focus in this study has been on biological signal transduction networks, the results are more general. The same model can be, for example, applied to describe industrial logistic networks where a set of different goods should be transported, in prescribed fractions, to a variety of destinations.We have effectively shown how functional networks can be designed and made robust against different kinds of local damage by running an artifical evolution process.

We are grateful to U. Alon for discussing these results and for permission to use his motif identification software.

[1] M. A. Savageau, Nature (London) **229**, 542 (1971).
[2] *Networks of Interacting Machines: Production Organization in Industrial Systems and Biological Cells*, edited by D. Armbruster, K. Kaneko, and A. S. Mikhailov (World Scientific, Singapore, 2005).
[3] A. S. Mikhailov and V. Calenbuhr, *From Cells to Societies: Models of Complex Coherent Action* (Springer, Berlin, 2002; 2nd updated printing 2006).
[4] U. Alon, M. G. Surette, N. Barkai, and S. Leibler, Nature (London) **397**, 169 (1999).
[5] Th. Handorf, O. Ebenhöh, and R. Heinrich, J. Mol. Evol. **61**, 498 (2005).
[6] B. Binder and R. Heinrich, *Genome Informatics 2004*, edited by H. Mamitsuka (Universal Academy Press, Tokyo, 2004), p.13–23.
[7] F. Li, T. Long, Y. Lu, Q. Ouyang, and Ch. Tang, Proc. Natl. Acad. Sci. U.S.A. **101**, 4781 (2004).
[8] S. Bornholdt and K. Sneppen, Proc. R. Soc. London, Ser. B **267**, 2281 (2000).
[9] R. Albert, H. Jeong, and A. L. Barabasi, Nature (London) **406**, 378 (2000).
[10] D. S. Callaway, M. E. J. Newman, S. H. Strogatz, and D. J. Watts, Phys. Rev. Lett. **85**, 5468 (2000).
[11] T. Tanizawa, G. Paul, R. Cohen, S. Havlin, and H. E. Stanley, Phys. Rev. E **71**, 047101 (2005).
[12] R. Milo, S. Itzkovitz, N. Kashtan, R. Levitt, S. Shen-Orr, I. Ayzenshtat, M. Sheffer, and U. Alon, Science **303**, 1538 (2004).
[13] M. Ipsen and A. S. Mikhailov, Phys. Rev. E **66**, 046109 (2002).
[14] R. Diestel, *Graph Theory* (Springer, Berlin, 1997).