Available online at www.sciencedirect.com
ScienceDirect
Journal of Theoretical Biology 248 (2007) 618–626
Journal of Theoretical Biology
www.elsevier.com/locate/yjtbi

Comparative study of the transcriptional regulatory networks of *E. coli* and yeast: Structural characteristics leading to marginal dynamic stability

Deok-Sun Lee*[,1], Heiko Rieger

Theoretische Physik, Universität des Saarlandes, 66041 Saarbrücken, Germany

Received 19 October 2006; received in revised form 18 May 2007; accepted 3 July 2007
Available online 14 July 2007

Abstract

Dynamical properties of the transcriptional regulatory network of *Escherichia coli* and *Saccharomyces cerevisiae* are studied within the framework of random Boolean functions. The dynamical response of these networks to a single point mutation is characterized by the number of mutated elements as a function of time and the distribution of the relaxation time to a new stationary state, which turn out to be different in both networks. Comparison with the behavior of randomized networks reveals relevant structural characteristics other than the mean connectivity, namely the organization of circuits and the functional form of the in-degree distribution. The abundance of single-element circuits in *E. coli* and the broad in-degree distribution of *S. cerevisiae* shift their dynamics towards marginal stability overcoming the restrictions imposed by their mean connectivities, which is argued to be related to the simultaneous presence of robustness and adaptivity in living organisms.

Keywords: Gene regulatory network; Boolean model; Dynamic stability

1. Introduction

Living organisms depend simultaneously on a stable internal environment and a capability to adapt to a fluctuating external environment (Causton et al., 2001). Since the biological characteristics of an organism are determined by the interplay between its gene repertoire and the regulatory apparatus (Babu et al., 2004), robustness and adaptiveness should be generic features of the molecular interactions composing the gene regulation machinery. The organization of the gene transcriptional regulatory network has been analyzed for numerous organisms, in particular for the prokaryote *Escherichia coli* (*E. coli*) (Thieffry et al., 1998; Dobrin et al., 2004; Shen-Orr et al., 2002) and the eukaryote *Saccharomyces cerevisiae* (*S. cerevisiae*) (Guelzim et al., 2002; Lee et al., 2002; Luscombe et al., 2004).

Adaptivity of an organism implies the production of different cell types with different functions from the same genome. This begins with a regulated transcription by certain proteins, transcriptional factor (TF) (Orphanides and Reinberg, 2002). The identification of the target genes for each TF allows the construction of a gene transcriptional regulatory network, where the nodes are the genes or operons that produce TF's or are regulated by TF's, and the directed edges indicate a regulatory dependence: A directed edge from node A to node B implies that a TF encoded by gene A is involved in the regulation of the expression of gene B. The expression level of each gene defines the dynamical state of the network. To achieve robustness and adaptiveness at the same time one expects the regulatory network dynamics to be neither chaotic nor fully insensitive to perturbations, but marginally stable. Structural characteristics of the network must support these dynamical features.

*Corresponding author.
E-mail address: deoksun.lee@gmail.com (D.-S. Lee).
[1]Present address: Department of Physics, University of Notre Dame, Notre Dame, Indiana 46556, USA.

0022-5193/$ - see front matter © 2007 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2007.07.001

Our study reveals specific topological features in the transcriptional regulatory network architecture of *E. coli* and *S. cerevisiae* that shift the dynamics towards marginal stability. *E. coli*'s network has a very low mean connectivity, the number of edges per node, which would lead to a high stability thus deteriorating adaptiveness in random networks, where all regulating rules are equally probable. But we find that single-element circuits which are anomalously rich in *E. coli*'s network help mutations triggered by random perturbations to persist, favoring an unstable dynamical behavior. *S. cerevisiae* on the other hand has a sufficiently high mean connectivity which favors chaotic dynamics in the random networks deteriorating stability. Here we find that *S. cerevisiae*'s network has a broad (algebraic) node degree distribution and we demonstrate the stabilizing effect of this feature upon the dynamics.

Practically, the information about the transcriptional regulatory network structure—which TF binds to which gene—is available, for example, via the chromatin-immunoprecipitation microarray experiments (Lee et al., 2002). The question, whether a specific TF enforces or inhibits the expression of a specific target gene has to be studied separately. However, those individual interactions do not necessarily occur independently and these regulatory interactions are often combinatorial (Buchler et al., 2003) and time-, cell cycle-, or environment-dependent, limiting the available information on the complete regulation profile. Generic dynamical features then have to be extracted using model interactions as suggested by Kauffman (1969, 1993): One digitizes the continuous expression level to a Boolean variable, 0 (inactive) and 1 (active), and assumes a random static regulation rule for each gene in the form of a random Boolean function for each gene determining its state at the next time step by the current states of its regulators. Here *random* means that the output value of these Boolean functions is 0 or 1 with equal probabilities.

Based on considerations of random Boolean networks with a fixed number of regulators k for every element, Kauffman (1969, 1993) hypothesized that distinct stationary states—limit cycles—correspond to different types of cells. This idea got some support from the agreement of the scaling behavior of the number of limit-cycles for $k = 2$-random Boolean networks and the number of cell types with respect to the genome size, but was also debated (Samuelsson and Troein, 2003; Klemm and Bornholdt, 2005). Among networks with fixed in-degree, $k = 2$ is a critical point distinguishing two different dynamical phases: stable and unstable against perturbations, suggesting that the regulatory network dynamics of living organisms is "on the edge" between order and chaos (Kauffman, 1969, 1993).

However, real regulatory networks do not have a fixed in-degree but a heterogeneous connectivity, even their average in-degree $\langle k \rangle$ is usually different from 2. Nevertheless the Boolean model itself is useful, and recently the effects of the nature of the regulating rules on the dynamical stability were studied within its framework (Harris et al., 2002; Kauffman et al., 2003, 2004), which will be discussed later in connection with our results. We propose that the network structure itself is also relevant for the stability/instability aspect mentioned before. Therefore we construct a network from the data for the transcriptional regulatory interactions for *E. coli* and *S. cerevisiae*, and study how a point mutation, i.e., an altered dynamical state of a single element, spreads over the whole network by inducing another mutation through regulatory interactions in the random Boolean functional form.

2. Method

2.1. Datasets

The transcriptional regulatory network in *E. coli* has long been studied and the obtained results are well integrated e.g., in RegulonDB database (http://regulondb.ccg.unam.mx). We used the dataset of Ref. Shen-Orr et al. (2002), which are based on the Regulon DB and enhanced by literature search. The resultant network consists of 418 operons and 573 interactions with 111 nodes having at least one outward edge. On the contrary, the transcriptional regulation of *S. cerevisiae* on the genome scale became available only very recently via the combination of chromatin-immunoprecipitation and DNA microarray analysis (Lee et al., 2002). We used the data of Ref. [Lee et al. (2002)] and chose the P value threshold 0.01 to yield a network of 4555 nodes and 12 455 directed edges with 112 nodes having at least one outward edge. Isolated nodes and those possessing only self-regulation have been excluded in both networks since they have no interaction with other elements.

2.2. Random Boolean functions

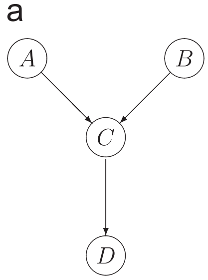
These experimental data establish a directed network G of N nodes, and we assign a dynamic Boolean variable σ_i (that can take on the values 0 or 1 only, corresponding to an inactive or active state, respectively) to each node i. These dynamical variables evolve synchronously via $\sigma_i(t+1) = f_i(\sigma_{i_1}(t), \sigma_{i_2}(t), \ldots, \sigma_{i_{k_i}}(t))$, with the nodes $i_1, i_2, \ldots, i_{k_i}$ having the outward edges incident on the node i and k_i the in-degree of the node i. The output value of f_i for each input configuration $\{\sigma_{i_1}(t), \sigma_{i_2}(t), \ldots, \sigma_{i_{k_i}}(t)\}$ is 1 with probability p or 0 with probability $1 - p$. Once determined at the beginning, the regulating rule f_i does not change with time. If $k_i = 0$, σ_i is fixed at f_i; $\sigma_i(t+1) = f_i$ regardless of the value of $\sigma_i(t)$. Here we introduced a parameter p controlling the distribution of the regulating rules. If $p = 0$ (1), the output value should be 0 (1) for any input configuration, yielding $\sigma_i = 0(1)$ for all i. On the other hand, if $p = \frac{1}{2}$, an input configuration can lead to the output value 0 and 1 with the same probability $\frac{1}{2}$, and as a result, all $2^{2^{k_i}}$ regulating rules for a node with in-degree k_i

are equally probable. While presenting the results, we will use another parameter λ defined as the probability that a regulating rule yields different output values from different input configurations. This is useful in understanding the dynamic stability introduced below, and one can find easily that it reduces to $2p(1-p)$ in the random Boolean networks, which is just twice the probability that one output value is 0 and the other output value is 1. An example network with this Boolean dynamics is given in Fig. 1.

2.3. Stability measure

The stability of a time-trajectory $\Sigma(t)$ is assessed by the effects of a point mutation $\sigma_i \to 1-\sigma_i$ on the dynamical evolution of the subsequent states. For this, we choose a configuration $\Sigma = \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$, and prepare its mutant, $\hat{\Sigma} = \{\hat{\sigma}_1, \hat{\sigma}_2, \ldots, \hat{\sigma}_N\}$, where $\hat{\sigma}_i = \sigma_i$ for all i except j with j chosen arbitrarily. Evolving Σ and $\hat{\Sigma}$ on the same network with the same regulating rules, we count $n_m(t)$, the number of element i's with $\sigma_i(t) \neq \hat{\sigma}_i(t)$, at each time step t. A node with $\Delta\sigma_i(t) \equiv |\sigma_i(t) - \hat{\sigma}_i(t)| > 0$ is considered as mutated. We average $n_{\mathrm{m}}(t)$ over different realizations of the regulating rules and different initial pairs of configurations to get the average, $N_m(t) = \langle n_m(t) \rangle$, which converges to its stationary value N_m. For each individual normal–mutant pair $(\Sigma, \hat{\Sigma})$, one can measure the relaxation time t_{r} after which $n_m(t)$ reaches its stationary value. Its distribution $P(t_{\mathrm{r}})$ is investigated as well.

a



c

t	σ_A	σ_B	σ_C	σ_D
0	0	0	0	1
1	1	0	1	1
2	1	0	0	1
3	1	0	0	0

b

$\sigma_A(t)$	$\sigma_A(t+1)$	$\sigma_B(t)$	$\sigma_B(t+1)$
0	1	0	0
1	1	1	0

$\sigma_A(t)$	$\sigma_B(t)$	$\sigma_C(t+1)$	$\sigma_C(t)$	$\sigma_D(t+1)$
0	0	1	0	0
0	1	0	1	1
1	0	0		
1	1	1		

Fig. 1. An example of Boolean dynamics. (a) A Boolean network of four nodes and three directed edges. Each node has a Boolean variable σ_i ($i = A, B, C, D$). (b) Regulating rules f_i's determining the node i's state at time $t+1$ with its regulators' states at time t as input. The nodes A and B have no regulator and their Boolean variables take constant values, respectively, at time $t+1$ regardless of their values at time t. (c) An example of the time evolution of those Boolean variables under the regulating rules in (b).

3. Results

3.1. Time evolution of the number of mutated elements

Fig. 2(a) and (b) present the results for the number of mutated elements $N_m(t)$ and N_m. At $p = \frac{1}{2}$ or equivalently at $\lambda = 2p(1-p) = \frac{1}{2}$, $N_m(t)$ decreases very rapidly from $N_m(0) = 1$ to a much smaller value in *E. coli*. On the other hand, N_m for *S. cerevisiae* increases with time up to about 3, indicating the occurrence of a mutation cascade. As the parameter λ decreases, N_m decreases both in *E. coli* and *S. cerevisiae*. In *E. coli*, N_m is smaller than 0.3 for all $\lambda \leqslant \frac{1}{2}$ indicating that system-wide mutations are suppressed in the random Boolean dynamics. Fig. 2 also shows that in *S. cerevisiae*, N_m is smaller than in *E. coli* for $\lambda \lesssim 0.2$ but increases with λ very rapidly to become larger for $\lambda \gtrsim 0.2$.

The functional form of $P(t_{\mathrm{r}})$ for $p = \frac{1}{2}$ in Fig. 2(c) is strikingly different between *E. coli* and *S. cerevisiae*: it is exponential for *E. coli* and a power-law, $P(t_{\mathrm{r}}) \sim t_{\mathrm{r}}^{-1.5(2)}$, for *S. cerevisiae*. This long tail of $P(t_{\mathrm{r}})$ implies that in the case of *S. cerevisiae* an element can be mutated and recovered even at very late times in the dynamics.

3.2. Mean connectivity

These differences in the mutation spread dynamics may be primarily attributed to a difference in the mean connectivity and can be understood by a mean-field approach (Derrida and Pomeau, 1986; Aldana and Cluzel, 2003): The probability $H(t) = \lim_{N\to\infty} N_m(t)/N$ that a randomly chosen node i is mutated at time t, also called the Hamming distance, is given in terms of the probability that a regulator of the node i is mutated, which we denote by $\bar{H}(t)$, and the probability that the regulating rule f_i yields different output values from different input configurations, λ, as

$$H(t+1) = \sum_{k_{in}} \lambda(1-(1-\bar{H}(t))^k) P_d(k),$$
$$\bar{H}(t+1) = \sum_{k,q} \lambda(1-(1-\bar{H}(t))^k) \frac{qP_d(k,q)}{\langle q \rangle}. \quad (1)$$

Here $P_d(k,q)$ is the joint probability that a node has in-degree k and out-degree q and is related to the in-degree distribution $P_d(k) = \sum_q P_d(k,q)$. $H(t)$ and $\bar{H}(t)$ evolve towards their stationary values H and $\bar{H}$. Setting $\bar{H}(t+1) = \bar{H}(t) = \bar{H}$ and expanding the second line of Eq. (1) for small $\bar{H}$, one finds $\bar{H} \simeq \bar{H}\lambda\langle kq\rangle/\langle q\rangle - \bar{H}^2\lambda\langle k^2 q\rangle/(2\langle q\rangle) + \mathcal{O}(\bar{H}^3)$ provided $\langle q\rangle$, $\langle kq\rangle$, and $\langle k^2 q\rangle$ are all finite. Therefore $\bar{H}$ and H are zero for λ smaller than a critical value λ_c with $\lambda_c = 1/K$ and $K \equiv \langle kq\rangle/\langle q\rangle$ and non-zero otherwise. The expression $\lambda_c = K^{-1}$ for the critical point holds true as long as K is finite. Since the Hamming distance H can be positive only if $K > 2$, $N_m \simeq HN$ for finite N should be small in *E. coli* that has the value $K \simeq 1.08$ and can be large, of order N, for $\lambda \gtrsim 0.42$ in *S. cerevisiae* that has $K \simeq 2.35$. Although the Hamming

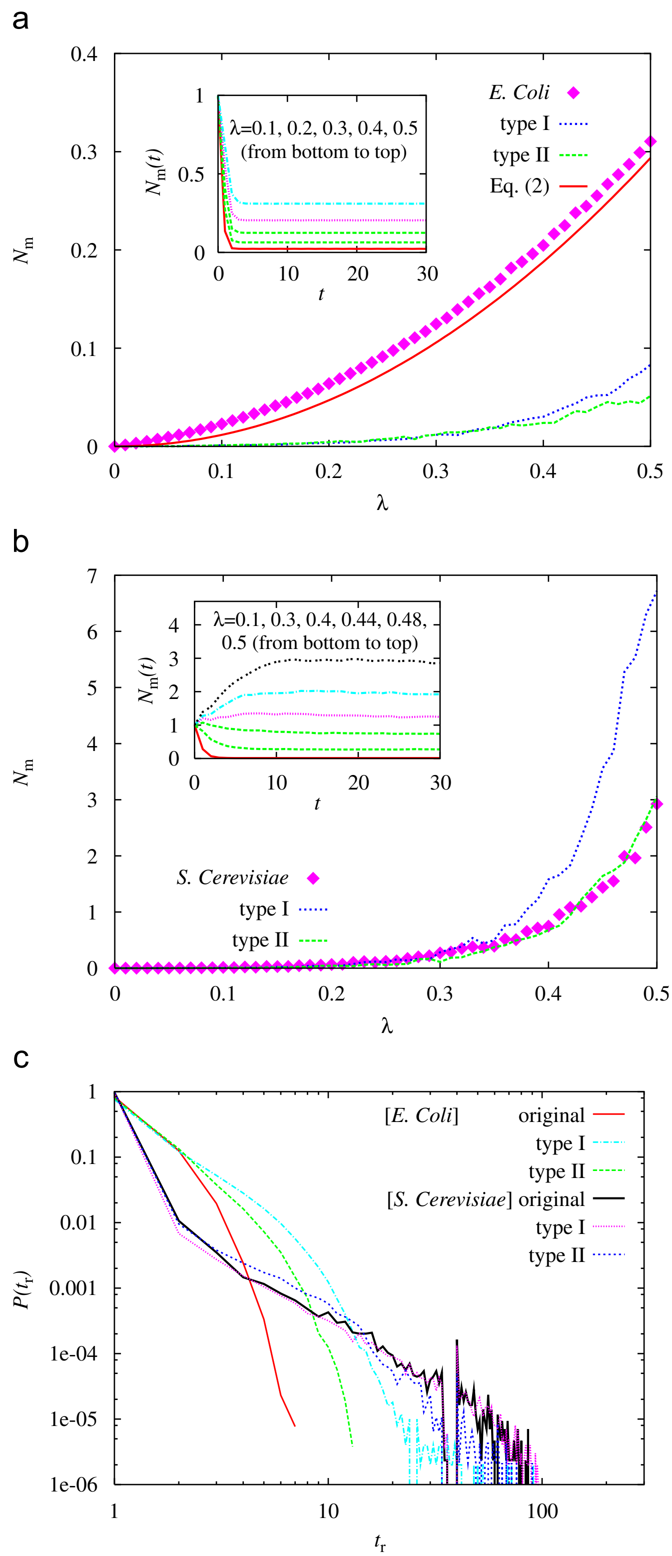


Fig. 2. Number of mutated elements $N_m(t)$ and $N_m = \lim_{t\to\infty} N_m(t)$ and distribution of the relaxation time $P(t_r)$. (a) Plot of the stationary value N_m versus $\lambda = 2p(1-p)$ in the original network and two types of randomized graphs (see the text for the definition) for *E. coli*. The data are averages over 10^2 initial pairs of configurations for each of more than 10^3 realizations of regulating rules. The approximation given in Eq. (2) is drawn together. The inset shows the time developments $N_m(t)$ for selected values of λ in the original *E. coli* network. (b) The same data as (a) for *S. cerevisiae*. (c) Plots of $P(t_r)$ with $p = \frac{1}{2}$ ($\lambda = \frac{1}{2}$) on the original networks and the randomized graphs for *E. coli* and *S. cerevisiae*.

distance is not necessarily of order N^{-1} at λ_c, one finds the value of λ for which $N_m = 1$ very close to the value $K^{-1} \simeq 0.42$ in the latter. The in-degree k and the out-degree q show no significant correlation in the two networks according to our analysis not presented here, that is, $P_d(k,q) \simeq P_d(k)P_d(q)$, which yields $\langle kq\rangle \simeq \langle k\rangle\langle q\rangle$ and $K \simeq \langle k\rangle$.

3.3. Comparison with randomized networks

Next we studied the same dynamics in two kinds of randomized networks derived from the regulatory networks of *E. coli* and *S. cerevisiae*. The first type of randomized graphs (type I) are constructed by the repetition of removing an edge connecting nodes v_1 and w_1 and creating a new one between v_2 and w_2, where both v_1 and v_2 had at least one outward edge and the node pair v_2 and w_2 were not connected before this change. Thus these type-I randomized networks have the same number of nodes, edges, and TF's as the original networks, but the edges connect randomly chosen pairs of TF and target gene. Our results for N_m and $P(t_r)$ are shown in Fig. 2. For type-I randomized graphs derived from *E. coli*, N_m is substantially suppressed as compared with the original network. In type-I random graphs derived from *S. cerevisiae*, N_m increases much more rapidly passing $\lambda \simeq 0.3$. The relaxation time distribution for the random graphs from *E. coli* is broader than for the original network but still decays faster than that for *S. cerevisiae*. The type-I randomization does not change significantly the relaxation time distribution for *S. cerevisiae*.

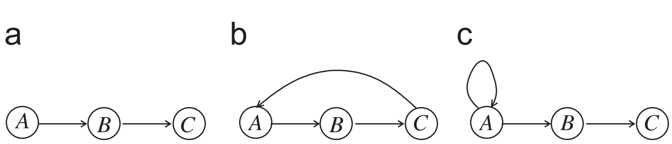
The type-II randomized graphs we considered are constructed by exchanging the end points of two edges: Two randomly chosen edges $e_1 = (v_1, w_1)$ and $e_2 = (v_2, w_2)$ are replaced by $e_1' = (v_1, w_2)$ and $e_2' = (v_2, w_1)$, respectively. These graphs preserve the joint degree distribution $P_d(k,q)$, but their local connectivity patterns may be different from that in the original network. We present the plots of N_m and $P(t_r)$ in Fig. 2. This type-II randomization does not change the relaxation time distribution for *S. cerevisiae* neither. Thus much faster decay of the relaxation time in the original and randomized networks for *E. coli* than in those for *S. cerevisiae* can be ascribed to the much lower mean connectivity, $\langle k\rangle \simeq 1.24$, of the former than that of the latter, $\langle k\rangle \simeq 2.73$. Interestingly the quantities N_m and $P(t_r)$ for these randomized graphs agree well with those for the original network of *S. cerevisiae*, but not for *E. coli*: This implies that it is the degree distribution that is mainly responsible for the spread of mutation in *S. cerevisiae* while other (local) structural factors must be important in *E. coli*.

3.4. Abundance of single-element circuits in E. coli

One might expect that circuits (directed closed paths) in the regulatory network play an important role for the spread of mutations, because in networks with a tree-structure, i.e., without circuits, point mutations spread without circulation and a node that is mutated will recover at the next time step and never become mutated again as indicated in Fig. 3(a). The nodes on a circuit, on the other hand, can return to a mutated state even after recovery (Fig. 3(b)). The nodes lying on circuits or those on bridges connecting distinct circuits can in principle switch their status permanently and thus they can be considered as comprising a core in the dynamics of mutation spread. As a subnetwork including all such circuits and the bridges connecting them, we define the core of a network as the maximal subgraph in which each node has at least one inward edge coming from and at least one outward edge incident to an element of the core.

By deleting the edges having at either end a node that does not meet the requirement for the core elements, we found the core subnetwork in the regulatory networks of *E. coli* and *S. cerevisiae*. Note that if an edge has the same node at both ends, the node, which regulates itself, becomes the element of the core. The relevance of the core to the mutation spread dynamics can be understood e.g., by investigating the relaxation time distribution $P(t_r)$ in *S. cerevisiae* depending on the location of the initial point mutation. Our analysis shows that initial mutations in the core lead to a qualitatively equal (power-law with the same exponent) distribution of the relaxation time. On the other hand, initial mutations in the output module, consisting of all nodes that have inward edges coming from the nodes in the core and their edges, decay very fast since the output module has a tree structure and cannot cause mutations in the core.

The organization of the core turns out to be very different in *E. coli* and *S. cerevisiae* as shown in Fig. 4(a) and (b), respectively. Most of all, the nodes are much more densely connected in *S. cerevisiae* than in *E. coli*. This difference can be first ascribed to different mean connectivities of the nodes in the core: it is about 1.47 in *E. coli* and 2.65 in *S. cerevisiae*. However, a more striking difference exists in their core organization. In *E. coli*, all 54 circuits are identified, all of which are single-element



t	$\Delta\sigma_A$	$\Delta\sigma_B$	$\Delta\sigma_C$
0	1	0	0
1	0	1	0
2	0	0	1
3	0	0	0

t	$\Delta\sigma_A$	$\Delta\sigma_B$	$\Delta\sigma_C$
0	1	0	0
1	0	1	0
2	0	0	1
3	1	0	0

t	$\Delta\sigma_A$	$\Delta\sigma_B$	$\Delta\sigma_C$
0	1	0	0
1	1	1	0
2	1	1	1
3	1	1	1

Fig. 3. Network structure dependence of mutation spread. The regulating rules are given by $f_i(\sigma) = \sigma$ or $1 - \sigma$ for nodes i's with one input and $f_i = 1$ or 0 for nodes i's with no input. Thus a mutated regulator necessarily makes its target node mutated at the next time step. Time evolution of $\Delta\sigma_i = |\sigma_i - \hat{\sigma}_i|$ for each node is shown in tables. (a) No circuit (tree structure). All nodes recover at $t = 3$ and thus the Hamming distance H is zero. (b) A circuit of length 3. The point mutation circulates with period 3, resulting in $H = \frac{1}{3}$. (c) A single-element circuit together with tree structure. All nodes are mutated at $t = 2$ and thus $H = 1$.

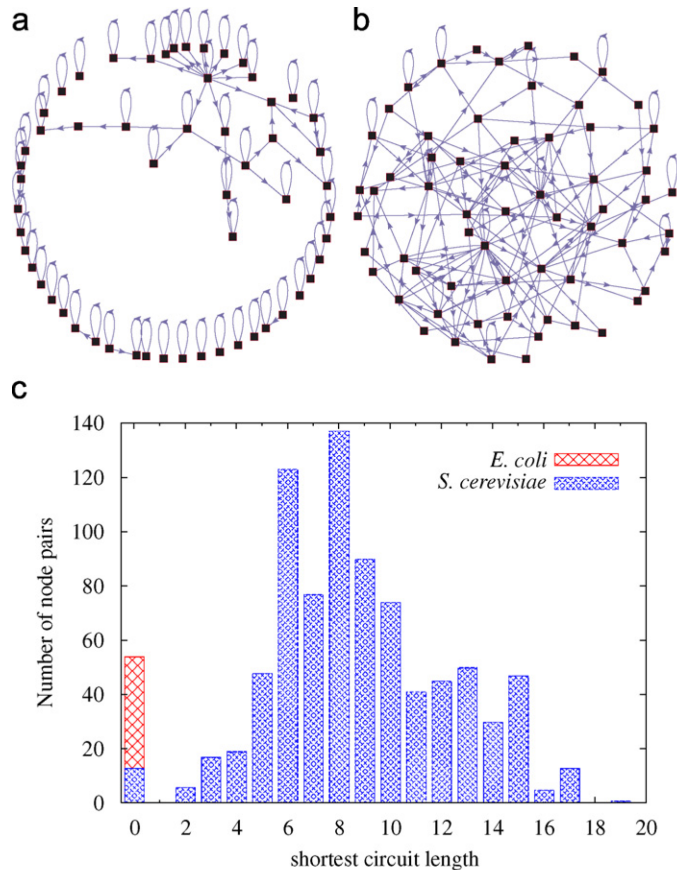


Fig. 4. Organization of the core in *E. coli* and *S. cerevisiae*. (a) Core of *E. coli*. It consists of 57 nodes and 84 edges. (b) Core of *S. cerevisiae*. It has 63 nodes and 167 edges. (c) Histogram of the shortest circuit lengths. In *E. coli*, a circuit longer than 1 is not observed but all 54 circuits are single-element ones. In *S. cerevisiae*, 836 pairs of nodes among all possible 1953 pairs in the core are connected by circuits and the shortest circuit length ranges from 0 to 19.

circuits representing self-regulation. There are no circuits whose length (i.e. the number of edges on the cycle) is larger than 1 (Thieffry et al., 1998). On the contrary, only one or two single-element circuits are formed in its randomized graphs. This organization of circuits in *E. coli* is also contrasted with the one in *S. cerevisiae*. We computed the shortest circuit for each pair of nodes in the core and counted the numbers of node pairs for each given shortest-circuit length. The distribution of shortest-circuit length obtained for *S. cerevisiae* is broad as shown in Fig. 4(c). We propose that the presence of single-element circuits in *E. coli* is the main reason for the enhancement of N_m of *E. coli* compared with both of its randomized graphs. Once a node i regulating itself is mutated, the input configurations to the regulating rule f_i are necessarily different between the normal–mutant pair $(\Sigma, \hat{\Sigma})$ since it is guaranteed that at least one of its regulators, the node i itself, is mutated. Recalling that a node can be mutated at the next time step only if the input configurations from the normal–mutant pair are different, one can see that single-element circuits have a higher probability to be mutated than nodes which do not regulate themselves (see Fig. 3(c)).

Therefore networks with more single-element circuits can be more adaptive. The absence of multi-element circuits in *E. coli* may be coming from incompleteness of the dataset we used and a few multi-element circuits may exist in reality. Even so, there is no difference in the contribution of abundant single-element circuits to the adaptivity of the network.

In the core of *E. coli* network, 54 edges are used for single-element circuits and the remaining 30 edges connect pairs of distinct nodes. As a result, the network has many isolated nodes and few small connected components, resulting in the rapid decay of the relaxation time. In Fig. 2(c), we find that the relaxation times observed in *E. coli* are mostly 1 or 2. From this, we can analytically predict the value of N_m as a function of λ. Suppose $N_m(t)$ saturates no later than time 2. From Eq. (1), $\bar{H}(1) = \lambda K N^{-1} + \mathcal{O}(N^{-2})$ since $\bar{H}(0) = N^{-1}$ and

$$N_m \simeq NH(2) \simeq N\lambda K\bar{H}(1) \simeq \lambda^2 K^2. \tag{2}$$

This is in good agreement with the true value as shown in Fig. 2(a).

3.5. Broad in-degree distribution in *S. cerevisiae*

In *S. cerevisiae*, the most significant dynamical feature that we found and that we need to explain is the slower increase of N_m with λ as compared with the type-I randomized graph, shown in Fig. 2(b). Contrary to the type-II randomized graphs, those of type-I do not preserve the degree distribution of the original network. From this, we can conjecture that the degree distribution of *S. cerevisiae* causes the slow increase of N_m. To check this, we analyze in detail the dependence of the Hamming distance on the degree distributions.

With uncorrelated in- and out-degree as is the case in the regulatory networks considered here, Eq. (1) is reduced to $H(t) = \bar{H}(t)$ and

$$H(t+1) = \lambda \sum_k [1 - (1 - H(t))^k] P_d(k). \tag{3}$$

Thus the in-degree distribution $P_d(k)$ determines the behavior of the Hamming distance $H(t)$. The in-degree distributions of *E. coli* and *S. cerevisiae* shown in Fig. 5(a) are quite different from each other. The maximum degree is 31 in *S. cerevisiae* while it is only 6 in *E. coli*. The in-degree distribution of *S. cerevisiae* is broader than that of its type-I randomized networks, too. The log–log plot of $P_d(k)$ in *S. cerevisiae* suggests that it fits into a power-law $P_d(k) \sim k^{-\gamma}$ with $\gamma \simeq 2.7 \pm 0.2$ as shown in Fig. 5. In Ref. (Guelzim et al., 2002) where about 900 regulatory interactions among the yeast genes are analyzed, the in-degree distribution was shown to be of an exponential-form with the maximum degree 13. This discrepancy may be attributed to analyzing different datasets of different sizes. While it is true that the data in Fig. 5(a) do not fit perfectly into the power-law, we use the power-law as a good approximation for the asymptotic behavior of the in-degree distribution of the

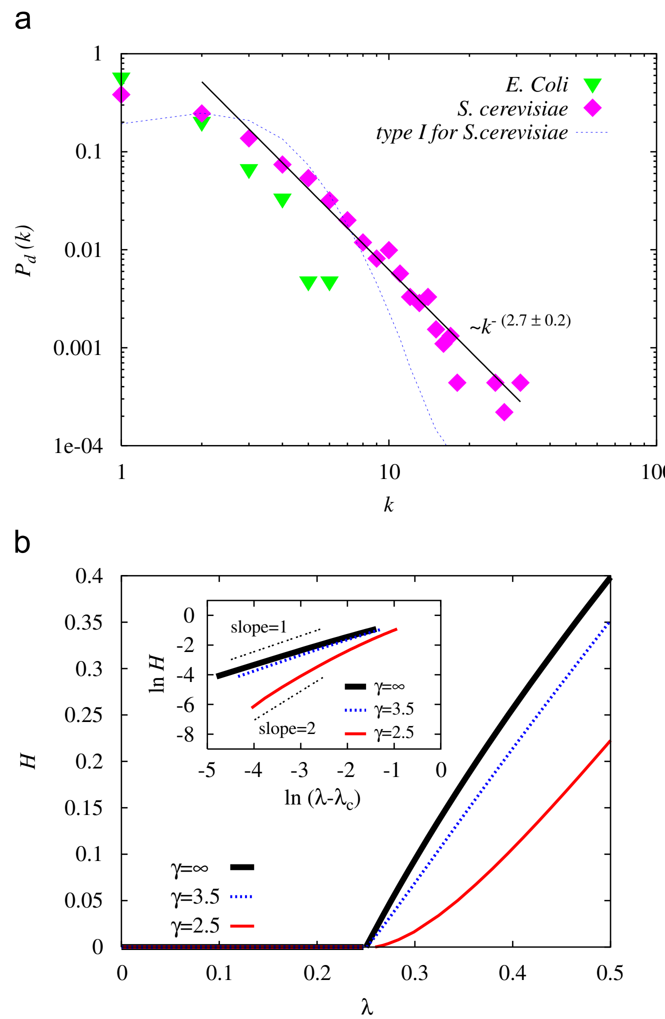


Fig. 5. Connectivity pattern and its effect on the critical behavior of the Hamming distance. (a) In-degree distributions $P_d(k)$ for *E. coli*, *S. cerevisiae*, and its type-I randomized networks. $P_d(k)$ of *S. cerevisiae* is broader than that of its type-I randomized networks or *E. coli*. Fitting $P_d(k)$ of *S. cerevisiae* with a power-law gives an approximate expression, $P_d(k) \sim k^{-\gamma}$ with $\gamma \simeq 2.7 \pm 0.2$. (b) Hamming distance H as a function of λ numerically obtained from Eq. (3) with $P_d(k)$ of the static model (Lee et al., 2004), which has a power-law tail as $P_d(k) \sim k^{-\gamma}$ with the exponent γ tunable. The inset shows that $H \sim \Delta$ commonly for $\gamma \to \infty$ and $\gamma = 3.5$, and that $H \sim \Delta^2$ for $\gamma = 2.5$, in agreement with Eq. (4).

yeast network, distinguishing it from that of the type-I randomized graphs following a Poisson distribution, $P_d(k) = \langle k \rangle^k e^{-\langle k \rangle}/k!$. Given $P_d(k) \sim k^{-\gamma}$, we find from Eq. (3) that the Hamming distance in the stationary state behaves as $H \sim \Delta^{\beta}$ for λ larger than the critical value λ_c with $\Delta \equiv \lambda/\lambda_c - 1$ and the critical exponent β given by

$$\beta = \begin{cases} 1 & (\gamma > 3), \\ 1/(\gamma - 2) & (2 < \gamma < 3). \end{cases} \tag{4}$$

The derivation of Eq. (4) is given in the Appendix. We restricted the range of γ to $\gamma > 2$ because the mean connectivity diverges with $\gamma < 2$. When the in-degree is subject to a Poisson distribution or an exponentially decaying distribution, it corresponds to $\gamma \to \infty$ and the critical behavior is the same as that for $\gamma > 3$. We present the numerical solution to Eq. (3) in Fig. 5(b) for $\gamma \to \infty$ (Poisson distribution), $\gamma = 3.5$, and $\gamma = 2.5$.

The increase of β with decreasing γ below $\gamma = 3$ indicates a difference in the behavior of the Hamming distance near the critical point between networks with $\gamma > 3$ and those with $2 < \gamma < 3$. Suppose we have two networks with a power-law in-degree distribution $P_d(k) \sim k^{-\gamma}$: One has $\gamma = 3.5$ and the other has $\gamma = 2.5$, and both have $\langle k \rangle = 4$. Then, in the region $0 < \Delta = \lambda/\lambda_c - 1 \ll 1$, the Hamming distance behaves as $H \sim \Delta$ for $\gamma = 3.5$ and $H \sim \Delta^2$ for $\gamma = 2.5$: the former increases more rapidly than the latter in the region $\Delta \ll 1$. Also the region where the Hamming distance remains non-zero but small, e.g., $H \leqslant 0.05$ is larger with $\gamma = 2.5$ than with $\gamma = 3.5$: it is given by $\lambda \in (0.25 : 0.29]$ with $\gamma = 3.5$ and $\lambda \in (0.25 : 0.35]$ with $\gamma = 2.5$. Such dependence of the Hamming distance on the in-degree exponent γ can thus explain different network responses between *S. cerevisiae* and its type-I randomized graphs. It is the broad in-degree distribution with $\gamma = 2.7(2)$ that makes the number of mutated elements increase with λ more slowly than in the corresponding type-I randomized graphs that have $\gamma \to \infty$. Due to such a slow increase of the Hamming distance, *S. cerevisiae* can keep the size of mutation small for a wider range of the parameter p or λ, which would be much larger with random structures.

3.6. Canalyzing Boolean functions

We study in this section the dynamic stability of the regulatory networks under the canalyzing rules, instead of random rules, as suggested in recent studies (Harris et al., 2002; Kauffman et al., 2003, 2004). Harris et al. have compiled about 150 regulating rules of various eukaryotic genes from literatures and reported on the possibility of a bias in their distribution (Harris et al., 2002). They were nested canalyzing functions described as follows (Kauffman et al., 2003). A rule f_i for a node with in-degree k_i has its input nodes ranked from 1 to k_i and the canalyzing Boolean input values $I_1, \ldots, I_{k_i}$ together with their respective canalyzing output values $O_1, \ldots, O_{k_i}$. For a given input configuration, $\{\sigma_1, \sigma_2, \ldots, \sigma_{k_i}\}$, the output value is equal to O_ℓ if $\sigma_\ell = I_\ell$ and $\sigma_j \neq I_j$ for all $j < \ell$. If $\sigma_j \neq I_j$ for all $1 \leqslant j \leqslant k_i$, the output value is $1 - O_{k_i}$. While *E. coli* is a prokaryotic organism and the canalyzing rules have not been shown to dominate all the regulatory interactions in *S. cerevisiae* on the genome scale, the canalyzing rule may be a biologically relevant principle in the regulatory interaction and thus it is worth studying the stability of the regulatory networks which allow only the canalyzing functions for their regulatory interactions.

The canalyzing rules are known to make the regulatory network of the yeast (Kauffman et al., 2003) and model networks with general in-degree distributions (Kauffman et al., 2004) stable against a point mutation. We also obtained the same results for *E. coli*'s and *S. cerevisiae*'s regulatory networks. Introducing the distributions for I_j and O_j commonly given by $P(I_j = 1) = P(O_j = 1) = \exp(-2^{-j}\alpha)/[1 + \exp(-2^{-j}\alpha)]$ with $\alpha = 7$ as in Refs. (Kauffman et al., 2003, 2004), we computed the number of mutated elements N_m in the stationary state as shown in Fig. 6. In both networks, N_m does not reach 1, implying the absence of mutation on a global scale. While the networks with larger mean connectivities become more unstable and *S. cerevisiae*'s network is unstable under random Boolean functions, large mean connectivities lead to stability (Kauffman et al., 2004) and the regulatory network of the yeast is stable, under canalyzing Boolean functions. This shows the sensitivity of the dynamic stability to the nature of the regulating rules.

However, even under the canalyzing Boolean functions, the architecture of the regulatory networks of *E. coli* and *S. cerevisiae* serve to shift the dynamics towards marginal stability. We present the number of mutated elements in the original network and those in the type I and type II randomized networks in Fig. 6. While all those networks are stable under the canalyzing regulation rules, the values of N_m of the real regulatory networks are far larger than those in their respective randomized networks, implying that the organization of *E. coli*'s or *S. cerevisiae*'s network is far from random but serving for enhancing the spread of mutation. This corroborates our finding of the bias towards marginal dynamic stability in the organization of the regulatory networks. It would be desirable to identify the network properties responsible for enhancing the mutation spread under the canalyzing Boolean functions, which is in progress.

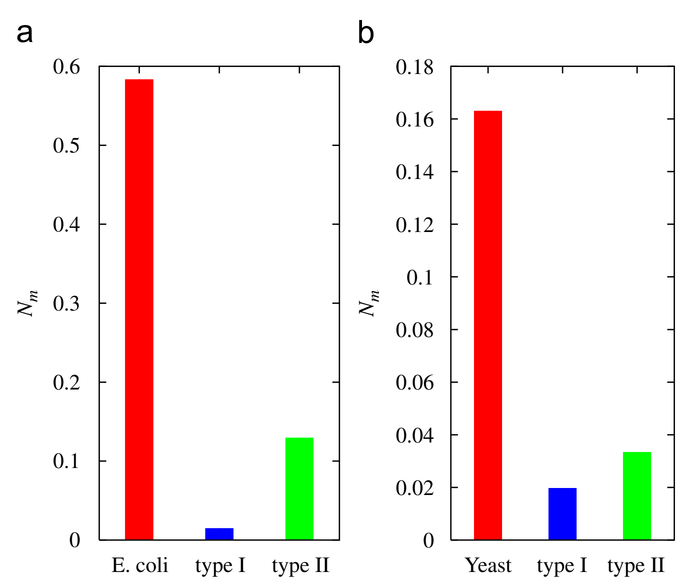


Fig. 6. Number of mutated elements $N_m(t)$ with the canalyzing Boolean functions used for the regulating rules. (a) Values of N_m in the original *E. coli* network and in the two types of randomized networks. The data are averages over 10^3 initial pairs of configurations for each of more than 10^3 realizations of regulating rules. (b) The same data as (a) for *S. cerevisiae*.

4. Conclusion

Dynamical robustness of biological networks such as the yeast cell-cycle network (Li et al., 2004) or the yeast transcriptional regulatory network (Kauffman et al., 2003, 2004) has been intensively studied and is well understood. Our results in this paper illuminate another aspect of the biological networks, adaptivity, as well as robustness, and suggest the possibility of an evolutionary pressure in the biological network organization towards the coexistence of robustness and adaptivity.

We performed numerical experiments—spread of mutation—to probe the dynamic stability of the recently unveiled networks of gene transcriptional regulation of *E. coli* and *S. cerevisiae* and provided analytical confirmation for the results by analyzing their structural features. While the small number of edges per node in *E. coli* fundamentally prohibits a global spread of mutation, a relatively large number of edges in *S. cerevisiae* enables a global mutation conditionally depending on the regulating rules. We further identified the relevant structural features which are distinguished from those of random graphs: All circuits of the regulatory network of *E. coli* are single-element circuits and the in-degree distribution of *S. cerevisiae* takes a power-law form. Single-element circuits in *E. coli* have higher probability to be mutated than nodes without self-regulation. The broad in-degree distribution in *S. cerevisiae* smoothens the increase of the number of mutated elements. This increase would be sharper for an exponential distribution, as is the case in the random graphs.

These biological networks appear to follow design principles that tend to balance the size of mutation. The small mean connectivity of the regulatory network of *E. coli* would restrict the size of mutations drastically, which is compensated by the abundance of single-element circuits that lead to the required enhancement of the mutation size. In the case of *S. cerevisiae*, its global characteristics of the regulatory network, a mean connectivity larger than 2, would lead to a very large mutation size, but a very heterogeneous interconnectivity pattern suppresses it. These local structural features demonstrate that both genetic networks have evolved, in spite of the restrictions imposed by the global characteristics, in such a direction that they can stay dynamically between stable (i.e., rarely mutated on a global scale) and unstable (easily mutated). Being neither stable nor unstable appears to be necessary for living organisms to maintain their stable internal state and adapt itself to fluctuating external environment simultaneously. Therefore our finding suggests that such a marginal dynamic stability of the whole system is supported by a selected structural organization of the internal systems on smaller scales, as the transcriptional regulatory network studied in this work. While we have concentrated only on the average in-degree, the organization of circuits, and the in-degree distribution of the network, further structural analysis will be helpful to illuminate how structure supports function.

Acknowledgments

We thank Uri Alon and Richard A. Young for allowing us to use their data. This work was supported by Deutsche Forschungsgemeinschaft (DFG).

Appendix A. Derivation of Eq. (4) from Eq. (3)

To find the behavior of $H = \lim_{t\to\infty} H(t)$ as a function of λ near the critical point $\lambda_c = \langle k\rangle^{-1}$, we set $H(t+1) = H(t) = H$ and expand Eq. (3) for small H since H is small around the critical point. Using the approximation $1 - H \simeq e^{-H}$ for $|H| \ll 1$ and the expansion $e^{-x} = \sum_{n=0}^{\infty}(-1)^n x^n/n!$, we obtain

$$H = \lambda \sum_{n=1}^{\infty} \frac{(-1)^{n+1}\langle k^n\rangle}{n!} H^n. \tag{A.1}$$

Here $\langle k^n\rangle$ is the nth moment of the in-degree distribution $P_d(k)$, i.e., $\langle k^n\rangle \equiv \sum_k k^n P_d(k)$. It is finite for all n only if $P_d(k)$ decays exponentially. In this case, all the terms in the right-hand side of Eq. (A.1) are analytic and keeping the first two leading terms, one finds that Eq. (A.1) is expressed as $H \simeq \lambda\langle k\rangle H - \lambda\langle k^2\rangle H^2/2$. This allows us to see that $H = 0$ for $\lambda < \lambda_c = \langle k\rangle^{-1}$ and $H \simeq 2(\lambda\langle k\rangle - 1)/(\lambda\langle k^2\rangle)$ or $H \sim \Delta$ with $\Delta \equiv \lambda/\lambda_c - 1$ for $\lambda > \lambda_c$.

When the in-degree distribution is a power-law asymptotically, $P_d(k) \sim k^{-\gamma}$, all the moments $\langle k^n\rangle$ are not finite: $\langle k^n\rangle$ for $n > n_*$ with $n_* = \lceil \gamma - 2\rceil$ diverges as $k_{\max}^{n-\gamma+1}/(n-\gamma+1)$, where $\lceil x\rceil$ is the smallest integer not smaller than x and $k_{\max}$ is the (average) largest in-degree. The largest in-degree diverges as $N^{1/(\gamma-1)}$, which is derived from the relation $\sum_{k>k_{\max}} P_d(k) \sim N^{-1}$. Thus $\langle k^n\rangle \sim N^{(n-\gamma+1)/(\gamma-1)}$. Such diverging terms are arranged as $H^{\gamma-1}\sum_{n>n_*}(-1)^{n+1}[k_{\max}H]^{n-\gamma+1}/[n!(n-\gamma+1)]$ in the right-hand side of Eq. (A.1). Here the summation converges to a constant in the limit $k_{\max}\bar{H} \to \infty$ due to alternating signs and fast decay of the coefficients (Lee, 2005). Thus the small-H expansion of Eq. (A.1) reads as $H = \lambda\sum_{n=1}^{n_*}(-1)^{n+1}\langle k^n\rangle H^n/n! + \lambda(\text{constant})H^{\gamma-1} + \cdots$. The $H^{\gamma-1}$ term is relevant to the critical behavior of H for $\gamma < 3$ since it holds for $\gamma < 3$ that $H \simeq \lambda\langle k\rangle H + \lambda(\text{const.})H^{\gamma-1}$, yielding $H \sim \Delta^{1/(\gamma-2)}$. On the other hand, the linear and quadratic terms are relevant for $\gamma > 3$ as for exponentially decaying in-degree distributions. In summary, the Hamming distance H with a power-law in-degree distribution $P_d(k) \sim k^{-\gamma}$ behaves near the critical point as

$$H \sim \begin{cases} \Delta, & (\gamma > 3), \\ \Delta^{1/(\gamma-2)}, & (2 < \gamma < 3). \end{cases} \tag{A.2}$$

References

Aldana, M., Cluzel, P., 2003. A natural class of robust networks. Proc. Natl Acad. Sci. 100, 8710–8714.

Babu, M.M., et al., 2004. Structure and evolution of transcriptional regulatory networks. Curr. Opin. Struct. Biol. 14, 283–291.

Buchler, N., Gerland, U., Hwa, T., 2003. On schemes of combinatorial transcription logic. Proc. Natl Acad. Sci. USA 100, 5136–5141.

Causton, H.C., et al., 2001. Remodeling of yeast genome expression in response to environmental changes. Mol. Biol. Cell 12, 323–337.

Derrida, B., Pomeau, Y., 1986. Random networks of automata: a simple annealed approximation. Europhys. Lett. 1, 45–49.

Dobrin, R., Beg, Q.K., Barabási, A.-L., Oltvai, Z.N., 2004. Aggregation of topological motifs in the *Escherichia coli* transcriptional regulatory network. BMC Bioinformatics 5, 10.

Guelzim, N., Bottani, S., Képès, F., 2002. Topological and causal structure of the yeast transcriptional regulatory network. Nature Genetics 31, 60–63.

Harris, S.E., Sawhill, B.K., Wuensche, A., Kauffman, S., 2002. A model of transcriptional regulatory networks based on biases in the observed regulation rules. Complexity 7, 23–40.

Kauffman, S., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. J. Theor. Biol. 22, 437–467.

Kauffman, S., 1993. The Origins of Order: Self-organization and Selection in Evolution. Oxford University Press, Oxford.

Kauffman, S., Peterson, C., Samuelsson, B., Troein, C., 2003. Random Boolean network models and the yeast transcriptional network. Proc. Natl Acad. Sci. USA 100, 14796–14799.

Kauffman, S., Peterson, C., Samuelsson, B., Troein, C., 2004. Genetic networks with canalyzing Boolean rules are always stable. Proc. Natl Acad. Sci. USA 101, 17102–17107.

Klemm, K., Bornholdt, S., 2005. Stable and unstable attractors in Boolean networks. Phys. Rev. E 72, 055101.

Lee, D.-S., 2005. Synchronization transition in scale-free networks: clusters of synchrony. Phys. Rev. E 72, 026208.

Lee, T.-I., et al., 2002. Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298, 799–804.

Lee, D.-S., Goh, K.-I., Kahng, B., Kim, D., 2004. Evolution of scale-free random graphs: Potts model formulation. Nucl. Phys. B 696, 351–380.

Li, F., et al., 2004. The yeast cell-cycle network is robustly designed. Proc. Natl Acad. Sci. USA 101, 4781–4786.

Luscombe, N.M., et al., 2004. Genomic analysis of regulatory network dynamics reveals large topological changes. Nature 431, 308–312.

Orphanides, G., Reinberg, D., 2002. A unified theory of gene expression. Cell 108, 439–451.

Samuelsson, B., Troein, C., 2003. Superpolynomial growth in the number of attractors in Kauffman networks. Phys. Rev. Lett. 90, 098701.

Shen-Orr, S., Milo, R., Mangan, S., Alon, U., 2002. Network motifs in the transcriptional regulation network of *Escherichia coli*. Nature Genetics 31, 64–68.

Thieffry, D., Huerta, A.M., Pérez-Rueda, E., Collado-Vides, J., 1998. From specific gene regulation to genomic networks: a global analysis of transcriptional regulation in *Escherichia coli*. Bioessays 20, 433–440.