

# Computational architecture of the yeast regulatory network

Sergei Maslov<sup>1</sup> and Kim Sneppen<sup>2</sup>

<sup>1</sup> Department of Physics, Brookhaven National Laboratory, Upton, NY 11973, USA

<sup>2</sup> Niels Bohr Institute, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

E-mail: [sneppen@nbi.dk](mailto:sneppen@nbi.dk)

Received 15 May 2005

Accepted for publication 27 July 2005

Published 9 November 2005

Online at [stacks.iop.org/PhysBio/2/S94](http://stacks.iop.org/PhysBio/2/S94)

## Abstract

The topology of regulatory networks contains clues to their overall design principles and evolutionary history. We find that while in- and out-degrees of a given protein in the regulatory network are not correlated with each other, there exists a strong negative correlation between the out-degree of a regulatory protein and in-degrees of its targets. Such correlation positions large regulatory modules on the periphery of the network and makes them rather well separated from each other. We also address the question of relative importance of different classes of proteins quantified by the lethality of null-mutants lacking one of them as well as by the level of their evolutionary conservation. It was found that in the yeast regulatory network highly connected proteins are in fact less important than their low-connected counterparts.

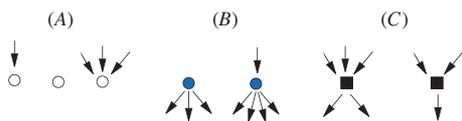
## Introduction

Living organisms need to orchestrate their responses to various environmental conditions, as well as to step through time-dependent programs (cycles) necessary for their function. This is achieved by the virtue of regulation of production, activity and degradation of their proteins by specialized regulatory proteins. All proteins in the genome of an organism thereby form a network with edges pointing away from regulators to targets of their regulations. The architecture of such regulatory networks contains clues about their function and evolution. They were recently shown: (a) to be characterized by a broad distribution of degrees of individual nodes with a few ‘hub’ proteins having a disproportionately large number of neighbors [1]; (b) to have those hubs positioned at the periphery of the network resulting in a global network architecture with soft modular features centered around individual hubs [2]; (c) to contain small yet statistically significant local manifestations of modularity also known as network motifs [3].

An important question is the extent to which one should view the regulatory network as modular as was suggested, e.g. in [4], or as more integrated, implied e.g. by gene disruption studies [5, 6]. The purpose of this work is to analyze the currently known part of the regulatory network of baker’s yeast *Saccharomyces cerevisiae* [7], with respect to global features of its architecture. We will argue that the picture of

hub-regulated modules positioned on the periphery of the network should be augmented with a centrally positioned core region of the network that alone can perform complex combinatorial computations. At the simplest level of description every node in a regulatory network is characterized by just two numbers: its in-degree, given by the number of its regulatory inputs, and its out-degree, given by the number of its outputs. These numbers allow us to divide all proteins into three basic categories illustrated in figure 1.

The first category is formed by the ‘workhorse’ proteins which have a direct functional role such as, e.g. catalyzing necessary metabolic reactions, helping other proteins to properly fold, or just being a part of cell’s structural foundation. These proteins are needed for the normal functioning of the cell or for its response to various external stimuli, but they typically do not regulate the concentration or activity of other proteins. As such they are characterized by zero out-degree  $K_{\text{out}} = 0$  in the regulatory network (see figure 1(A)). Workhorse proteins might have any number of regulatory inputs depending on the number of their functional roles in the cell or, alternatively, on the number of distinct environmental conditions requiring their production. The second and the third category consist of regulatory proteins with  $K_{\text{out}} > 0$ , which are further subdivided into ‘distributors’ (figure 1(B)) with at most one regulatory input ( $K_{\text{in}} = 0$  or  $K_{\text{in}} = 1$ ), and ‘integrators’ (figure 1(C)) regulated by two or more other



**Figure 1.** Three categories of proteins based on their in- and out-degrees in the regulatory network. (A) Workhorse proteins (open circles) with  $K_{\text{out}} = 0$  and any  $K_{\text{in}}$ ; (B) distributors (shaded circles) with  $K_{\text{out}} > 0$  and  $K_{\text{in}} = 0$  or  $1$ ; (C) integrators (filled squares) with  $K_{\text{out}} > 0$  and  $K_{\text{in}} \geq 2$ . The distinction between distributors and integrators emphasizes the difference between regulators that are simply transmitting a given input, and those that perform some nontrivial computation based on at least two different input channels.

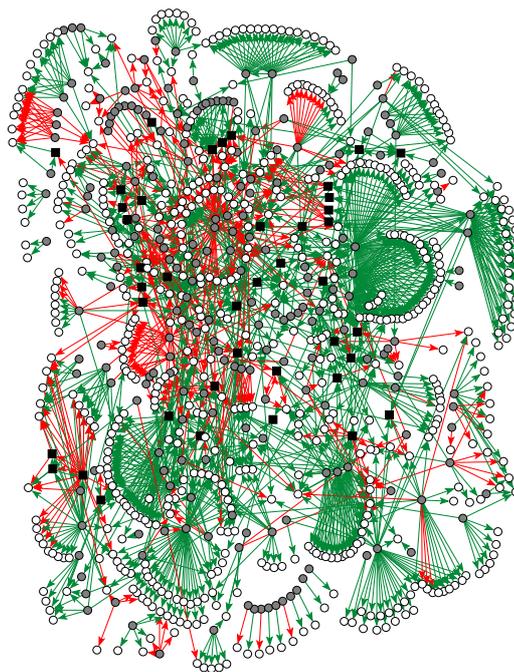
regulators ( $K_{\text{in}} > 1$ ). Distributors' role is to broadcast (and possibly invert) the regulatory signal they receive to a number of downstream proteins involved in a given function. The integrators, on the other hand, have a possibility of a more advanced combinatorial computation, with an output signal approximated by a nontrivial logical function of several inputs [8]. Given that even for the best-studied model organisms our knowledge of their regulatory networks is far from being complete, many distributors may in fact be integrators. Thus the three-tiered classification of proteins attempted in this work is only a practically implementable proxy to their true status.

## Results and discussion

### General properties of the network

The currently known part of the regulatory network in baker's yeast *Saccharomyces cerevisiae* downloaded from the YPD database [7] is visualized in figure 2. It contains 848 protein nodes, out of which 259 regulate the production, activity and degradation of others ( $K_{\text{out}} > 0$ ), linked by 1750 regulations. The list of regulatory mechanisms is dominated by 1276 transcriptional regulations by 125 transcription factors. Positive regulations outnumber negative ones 3 : 1. Histograms of out- and in-degrees of protein nodes in this network are shown in figures 3(A) and (B) respectively. Out-degrees of individual regulators range between 1 and 71 and their histogram approximately follows a power law  $\propto 1/K^\gamma$  with  $\gamma = 1.5$  for about one and a half decades. The in-degree distribution is more narrow than that of out-degrees and is better approximated by an exponential decay  $\exp(-\beta K_{\text{in}})$  with  $\beta = 0.42$ . The parameters of our fits are similar to their values previously reported [9] for a smaller dataset of transcriptional regulations in yeast.

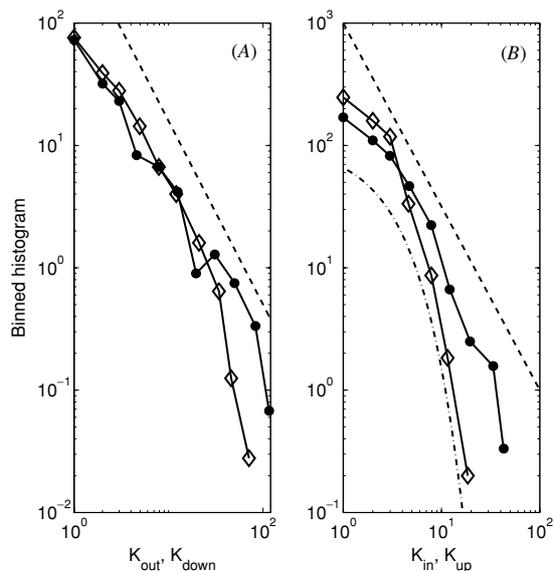
Contrary to [9] we found no correlation between  $K_{\text{in}}$  and  $K_{\text{out}}$  of protein nodes. This can be seen, for example, in the fact that workhorse proteins ( $K_{\text{out}} = 0$ ) have the same average in-degree ( $2.65 \pm 0.1$ ) as regulatory proteins with a non-zero in-degree ( $2.53 \pm 0.25$ ). At a more detailed level, the in-degree distributions of workhorse and regulatory proteins are virtually indistinguishable from each other (see figure 4). The potential area of downstream influence of a given regulatory protein is not limited to its immediate regulatory targets. Indeed,



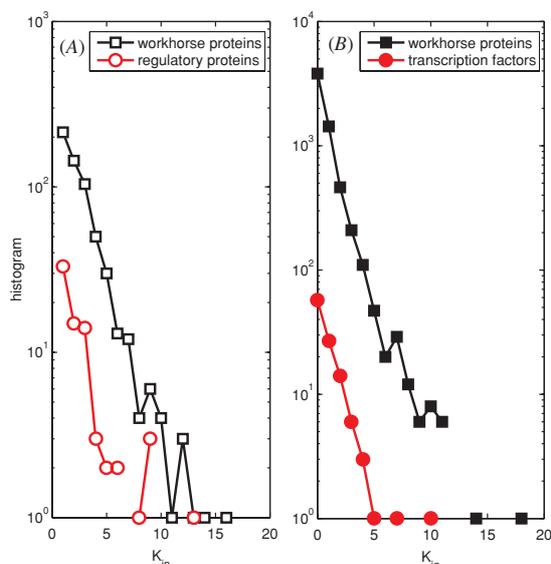
**Figure 2.** The presently known part of the regulatory network in baker's yeast *S. cerevisiae*. This network obtained from Yeast Proteome Database (YPD) [7] consists of 1750 regulations among 848 proteins by 259 regulatory proteins. Green and red arrows denote the positive and negative regulations, respectively. Filled squares correspond to integrators, gray circles to distributors and open circles to workhorse proteins.

some of these targets themselves may have regulatory outputs broadcasting the signal further downstream. To quantify this effect for every regulator we found the number  $K_{\text{down}} \geq K_{\text{out}}$  of all of its downstream targets (both direct and indirect). Similarly, for every node we found the number  $K_{\text{up}} \geq K_{\text{in}}$  of all regulators that are positioned directly or indirectly upstream from it in the regulatory network and which thus could in principle affect its abundance and activity. In figures 3(A) and (B) we show histograms of these two integrated properties of nodes. The  $K_{\text{down}}$  distribution is somewhat broader than that of out-degrees  $K_{\text{out}}$  and can be approximated by a power law with the same exponent  $\gamma = 1.5$  (dashed line in figure 3(A)) but over a wider range.

The  $K_{\text{up}}$  distribution is considerably broader than that of in-degrees, as it in fact can be approximately fitted by a power law with an exponent 1.5 (dashed line in figure 3(B)). While the exponent of the distribution can in principle have any value, the exponents of both  $K_{\text{down}}$  and  $K_{\text{up}}$  are known to be equal to  $3/2$  for any random network at a special value of its average connectivity [10, 11]. In fact, we find that distributions of  $K_{\text{down}}$  and  $K_{\text{up}}$  in the yeast regulatory network are indeed close to their random counterparts defined by randomly reshuffling the network using the algorithm proposed in [2]. This algorithm explicitly conserves in- and out-degrees of all nodes in the network while randomizing all of its higher-level topological properties. We found no statistically significant

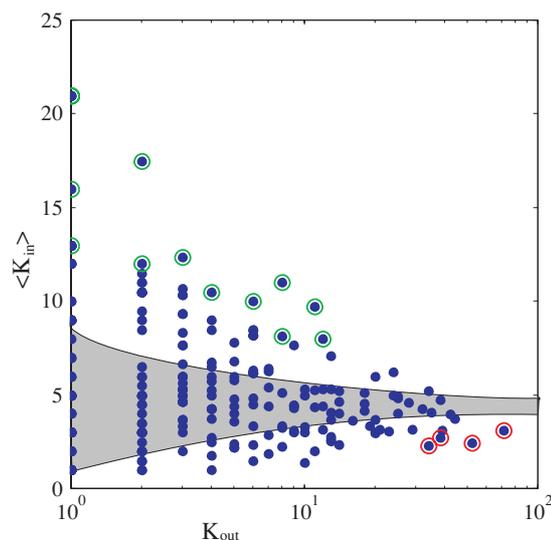


**Figure 3.** Basic topological properties of the network. (A) Logarithmically binned histograms of out-degrees  $K_{out}$  (open diamonds), and total downstream regions of influence  $K_{down} \geq K_{out}$  (filled circles) of all regulatory proteins. (B) Logarithmically binned histograms of in-degrees  $K_{in}$  (open diamonds), and upstream regions (filled circles) of all proteins. Dashed lines in both panels correspond to a power law with the exponent  $-1.5$ . The dot-dashed curve in the right panel is the exponential distribution  $N(K_{in}) \propto \exp(-0.42 K_{in})$ .



**Figure 4.** In-degrees of regulatory and workhorse proteins are identically distributed. Histograms of in-degrees of regulatory (red circles) and workhorse (black squares) proteins in (A) the YPD data set; (B) chip-on-chip dataset [1] with the  $P$ -value cutoff of 0.001.

differences between  $K_{down}$  (or alternatively  $K_{up}$ ) distributions in real and randomized networks. The close agreement between the upstream (as well as downstream) regions in real and randomized networks indicates the relative lack of modularity in the real network.



**Figure 5.** Anti-correlation of in- and out-degrees of neighboring nodes. The average in-degree  $\langle K_{in} \rangle$  in a module consisting of all direct targets of a given regulatory protein plotted versus its out-degree  $K_{out}$ . Filled blue symbols represent 259 such modules for each of the regulatory proteins listed in the YPD database. Modules with  $\langle K_{in} \rangle$  which is at least three standard deviations below its value in a randomized network are marked with red circles. They are controlled by regulators with very large out-degrees  $K_{out}$ . In contrast, modules with higher than average  $\langle K_{in} \rangle$  (again at three standard deviations or more) tend to be controlled by regulators with low  $K_{out}$ . They are marked with green circles. The standard deviation was calculated using an ensemble of 100 random networks generated by the algorithm [2] which strictly conserves  $K_{in}$  and  $K_{out}$  of every node. The shaded area denotes the typical (one standard deviation) range of  $\langle K_{in} \rangle$  observed in such randomized networks. Its width decreases with  $K_{out}$  due to the improved statistics.

#### Peripherally positioned hub-regulated modules

One of the important topological properties of the yeast regulatory network is that regulatory proteins with high out-degrees (regulatory hubs) tend to regulate targets with lower than average in-degrees [2]. This is explicitly demonstrated in figure 5, which plots the average in-degree  $\langle K_{in} \rangle$  of all targets controlled by a given regulator with the out-degree  $K_{out}$ . One sees that highly connected regulators tend to regulate proteins with lower than average  $\langle K_{in} \rangle$  and vice versa. This effect can also be quantified by the correlation coefficient (see materials and methods) of these variables equal to  $-0.22$  is statistically significant at 9 standard deviations ( $P$ -value around  $10^{-19}$ ), while their Spearman rank correlation (see materials and methods) of  $-0.21$  is similarly statistically significant. The deficit of connections between hub-regulators and proteins with multiple regulatory inputs is not a product of a possible anthropogenic bias of the data in the YPD database. To confirm this we repeated our analysis using a completely unbiased full genome assay of binding between 106 yeast transcription factors and cis-regulatory regions of all 6200 of yeast genes [1]. Again we found a highly statistically significant anti-correlation between out- and in-degrees of

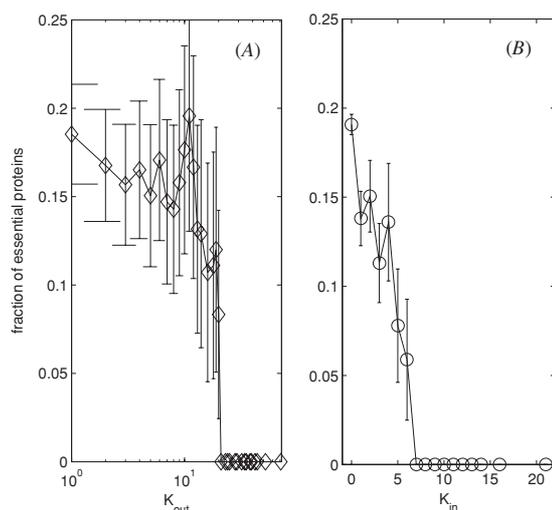
protein nodes connected by edges of this network: the Pearson correlation coefficient equal to  $-0.17$  is statistically significant at about 12 standard deviations ( $P$ -value around  $10^{-33}$ ).

#### The computational core of the regulatory network

The central region of the yeast regulatory network is formed by 206 known regulations of one of the 259 regulatory proteins by another. In the absence of such regulations multi-step programs (cycles) of gene expression as well as non-trivial combinatorial responses to external or internal stimuli would be impossible. In this work we advocate further subdividing regulatory proteins into integrators and distributors. While the distributors' role is to broadcast the regulatory signal they receive to their downstream targets, it is integrators (marked as black squares in figure 2) that predominantly perform computations. We have found that not only each integrator by definition receives several regulatory inputs but also it tends to send several output signals to other regulators. One can show that these two properties are in fact strongly correlated with each other: the Spearman rank correlation coefficient between the  $K_{in}$  and  $K_{out}^R \leq K_{out}$ , the number of regulatory outputs directed towards other regulators, is equal to 0.21, which is statistically significant at  $P$ -value of  $4 \times 10^{-4}$  (a very respectable number for only 259 data points). It should be contrasted to a complete lack of correlation between a regulatory protein and its total number of regulatory outputs (the Pearson rank correlation of 0.05 with  $P$ -value of 0.22). This confirms and extends our previous observation that distributions for regulatory and non-regulatory proteins are in fact identical to each other.

To quantify how interconnected is the computational core of the regulatory network we counted the number of edges connecting pairs of 41 integrators to each other. There are 30 such regulations in the actual network. This number should be compared to its value  $20 \pm 4$  in a randomized version of the network in which in- and out-degrees of individual nodes were strictly conserved [2]. This excess is offset by fewer regulations of integrators by distributors (124 in the real network versus  $134 \pm 4$  in a random one). The distributors do not appear to be more connected to each other than expected by pure chance alone (28 connections versus  $29 \pm 2$ ). Similarly edges pointing away from an integrator to a distributor are about equally scarce in the real network (5) and in its null-model ( $4 \pm 2$ ).

We also found a positive correlation between in-degrees of integrators that regulate each other. The Pearson correlation coefficient between such in-degrees is 0.36, which is statistically significant at around 2 standard deviations ( $P$ -value of 0.03). The Spearman rank correlation equal to 0.38 has a comparable  $P$ -value. A larger network of regulations among all regulators also contains a trace of the positive correlation between in-degrees of connected nodes but not at a statistically significant level. This indicates that it is indeed integrators which are responsible for this correlation. All these observations point to integrators forming the centrally positioned heavily interconnected computational core of the yeast regulatory network.



**Figure 6.** Anti-correlation between the importance of a protein and its in- and out-degrees in the regulatory network. The fraction of yeast proteins found to be essential for the survival of the cell [13] as a function of out-degree  $K_{out}$  (A) and in-degree  $K_{in}$  (B). Both panels demonstrate that the ‘importance’ of a protein decreases with its degree in the regulatory network. To improve the statistics the degree is used in a cumulative fashion: we plot the fraction of essential proteins among all tested proteins with the degree at or above  $K_{out}$  (in panel (A)) or  $K_{in}$  (in panel (B)).

#### Relative importance of proteins as a function of their in- and out-degrees in the regulatory network

A relative importance (essentiality) of a protein for the basic functioning of the cell can be crudely assessed by whether a mutant cell lacking this particular protein survives (viable null-mutant) or dies (lethal null-mutant). Hub-proteins are commonly believed to be more important and therefore null-mutants lacking one of them are expected to have higher than average chances of being lethal. Such positive correlation between protein’s connectivity and the likelihood of its null-mutant to be lethal was indeed observed in the protein–protein interaction network in yeast [12]. To test this hypothesis for the yeast regulatory network in figure 6(A) we plot the fraction of yeast regulators found to be essential for the survival of the cell [13] as a function of their out-degree. Contrary to this hypothesis we found that very highly connected regulators  $K_{out} > 20$  are in fact never essential: null mutants of all 22 of them are viable, which is statistically significant at the  $0.815^{22} = 0.011$  level. For low to intermediately connected regulators we found no statistically significant correlation between their out-degree and their chances to be essential.

The likelihood of a protein to be essential is also negatively correlated with its in-degree. Indeed, the average in-degree of essential proteins present in our dataset is equal to  $2.3 \pm 0.1$ , while that of non-essential ones is significantly higher than  $2.7 \pm 0.1$ . Figure 6(B) quantifies the same effect by showing that the fraction of essential proteins goes down with the in-degree in the regulatory network. This apparent insignificance of ‘in-degree hubs’ can be tentatively attributed to the following simple observation: the expression level of a protein controlled by many regulators tends to go up and down

by a significant factor in response to internal and external stimuli. Therefore, the organism is likely to be adapted to survive even in the complete absence of such proteins (a 100% decrease in the expression level). This is in agreement with a recent observation that essential proteins are characterized by a reduced level of uncertainty and noise in their expression levels [14].

Another measure commonly used to quantify the ‘importance’ of a given protein is the extent of its evolutionary conservation quantified by the ratio  $K_A/K_S$  between frequencies of amino acid changing ( $K_A$ ) and silent ( $K_S$ ) nucleotide substitutions. More important proteins tend to be more conserved in the course of evolution and hence are characterized by smaller values of  $K_A/K_S$ . When we used  $K_A/K_S$  [15] as an alternative measure of the importance of a protein again we found no significant correlations with its in- and out-degrees in the regulatory network: the  $P$ -value of the Spearman rank correlation between  $K_A/K_S$  and  $K_{in}$  is in a statistically insignificant 0.1–0.2 range. This was found to be true for both the database-derived (YPD) and the high-throughput (ChIP-on-chip) datasets. If anything, regulatory hubs proved to be less ‘important’ than their low-connected counterparts: a positive Spearman correlation coefficient of 0.13 between  $K_A/K_S$  and  $K_{out}$  of regulatory proteins in the YPD database (which at  $P$ -value 0.04 is just marginally statistically significant) hints that highly connected regulators tend to be less evolutionary conserved.

Once again our findings about the yeast regulatory network are opposite to those previously reported for the yeast protein–protein interaction network, where a small yet statistically significant negative correlation between the  $K_A/K_S$  and the number of interaction partners of a protein was detected [14]. This original observation was subsequently questioned [16, 17] and tentatively attributed to indirect effects due to a very strong negative correlation between the evolutionary rate and the abundance of a protein [18].

## Conclusions

Above we presented a number of empirical observations describing the topology of the yeast regulatory network. As clearly visible in figure 3 this network does not break up into multiple isolated modules corresponding to cell’s responses to various perturbations. In contrast, its different parts appear to be rather well interconnected with each other: the majority of all proteins are simultaneously controlled by several regulatory proteins, which in their turn control multiple targets.

We found no correlation between the number of regulatory inputs and regulatory outputs of a protein. In other words, the average number and the distribution of regulatory inputs of non-regulatory (workhorse) proteins are the same as that of their regulatory counterparts. This observation is somewhat surprising since proteins from these two categories process their regulatory inputs in different ways: while regulatory proteins propagate the regulatory signal further downstream, the workhorse proteins simply act based on it.

The complete lack of correlation between in- and out-degrees of the same protein should be compared to a

strong negative correlation between out-degrees of regulatory proteins and in-degrees of their targets. In particular, highly connected (hub) regulators tend to regulate proteins with lower than average in-degree. This causes modules controlled by such hub-regulators to be relatively well separated from each other and the rest of the network [2]. Indeed, targets of a given hub-regulator tend to belong exclusively to its module and are therefore coupled to the rest of the network only through this regulator. Visually, such modules tend to be positioned on the periphery of the network away from its densely interconnected core region (see figure 2). This trend was reported [19] to be even more pronounced in simpler (prokaryotic) organisms such as *E. coli*. It is evident from a large number of the so-called single input modules [19] consisting of a hub-regulator exclusively controlling a group of proteins with  $K_{in} = 1$ . More complicated wiring of the regulatory network in eukaryotes [8] manifests itself in a softer nature of this correlation in yeast.

The core of the regulatory network is formed by the set of regulations exerted by one regulatory protein on another. Integrators (regulators with two or more inputs) constitute the most complex and interconnected part of this core region. It was shown above that on average integrators also tend to regulate more other regulatory proteins than their distributor counterparts. Moreover, the positive correlation between numbers of inputs of integrators which are directly connected to each other in the regulatory network hints at several hierarchical levels present in this network. These levels are characterized by progressively increasing numbers of inputs of the regulatory proteins involved. This means that a somewhat arbitrary dividing point (2 or more regulatory inputs) we chose to separate integrators from distributors can be shifted up and down without changing our qualitative findings. To summarize we demonstrated that the higher is the number of regulatory inputs of a regulator, the higher is the number of its outputs directed toward other regulators and coincidentally the higher is the average number of inputs of those regulators.

We next addressed the question of relative importance of different proteins quantified by the lethality of null-mutants lacking one of them as well as the level of their evolutionary conservation. We have found negative correlation between the importance of a regulatory protein and the number of its downstream targets. This observation contradicts a naive point of view that the function ‘sits’ on edges of the network and hence the importance of the protein is directly proportional to the number of its immediate neighbors (its degree in the network in question). That is to say, if each edge in the network has a certain probability of being indispensable for the survival of the cell and this probability is independent for different edges starting at the same protein, then hub-proteins should be much more essential than their low-connected counterparts. This argument is obviously incorrect for the regulatory network where downstream targets of a given regulator typically correspond to just one function and thus are not independent of each other. Hence our results simply indicate that there is no clear correlation between the importance of a task and the connectivity of regulatory proteins involved in it.

Since out-degree of a regulator measures only the number of its direct targets and not to the total number of proteins

**Table 1.** The likelihood of a protein to be essential versus the centrality of its position in the regulatory network. Note that the fraction of essential proteins among the centrally positioned integrators is higher than that among more peripherally positioned distributors, which in its turn is somewhat higher than for workhorse proteins.

| Protein type          | Essential genes | Nonessential genes | Tested genes | Fraction of essential genes |
|-----------------------|-----------------|--------------------|--------------|-----------------------------|
| Workhorses            | 67              | 459                | 526          | 0.13(2)                     |
| Distributors          | 33              | 162                | 195          | 0.17(3)                     |
| Integrators           | 10              | 27                 | 37           | 0.27(9)                     |
| Overall in data       | 110             | 648                | 758          | 0.15(1)                     |
| Overall in yeast [13] | 1103            | 4678               | 5781         | 0.1915(5)                   |

participating in a given task, our results do not contradict [20] where it was reported that the importance of a given functional module tends to increase with the total number of proteins involved in it. Our observation that highly connected regulators tend to be somewhat less essential than their low-connected counterparts can be tentatively attributed to the peripheral position of the former in the regulatory network. Indeed, the fraction of essential proteins among the centrally positioned integrators is higher than that among more peripherally positioned distributors, which in its turn is somewhat higher than for workhorse proteins (see table 1)

## Materials and methods

The information about the yeast regulatory network was downloaded from the YPD database [7] at [www.proteome.com/YPD](http://www.proteome.com/YPD) in 2001, when the access to this database was still free of charge for academic institutions. With 848 protein nodes linked by 1772 regulations (including self-regulations) this dataset is considerably larger than the previously studied [9] transcription regulatory network consisting of 837 transcriptional regulations of 491 genes by 124 transcription factors. To eliminate potential artifacts in our classification scheme we excluded 22 self-regulations present in this dataset. That left us with 1750 regulations and 259 regulatory proteins. In some parts of this work we also utilized the results of the ‘ChIP-on-chip’ high-throughput experiment [1] which tested the *in vivo* binding of 106 yeast transcription factors to the upstream regulatory regions of genes encoding all 6270 of yeast proteins. The regulatory network with 4418 regulations was obtained from the raw experimental data by imposing a conservative *P*-value cutoff [1].

The system-wide data on viability of *S. cerevisiae*’s null-mutants used in our study were taken from [13] in which 1103 essential (non-viable null-mutants) and 4678 non-essential (viable null-mutants) yeast proteins were reported. The lists of viable and non-viable null-mutants as reported in [13] were downloaded from the *Saccharomyces* Genome Database (<http://genome-www.stanford.edu/Saccharomyces>). The ratios were reported in [15] where a comparative analysis of genomes of four yeast species was performed. The data were downloaded from the website maintained by the authors of

this paper: [http://www.broad.mit.edu/annotation/fungi/comp\\_yeasts/downloads.html](http://www.broad.mit.edu/annotation/fungi/comp_yeasts/downloads.html).

When comparison with a random network was used in this paper such random network was generated using the edge switching algorithm introduced in [2]. This algorithm strictly conserves both in- and out-degrees of every node, while randomly reassigning the neighbors. The set of Matlab programs performing such randomization is available at <http://www.cmth.bnl.gov/~maslov>. All networks were visualized using the Kamada–Kawai algorithm [21] built into the Pajek software tool for Windows 32 [22]. Pajek is free to the academic users and can be downloaded at <http://vlado.fmf.uni-lj.si/pub/networks/pajek>.

In sections of the paper we used Pearson’s rank coefficient. For example, between  $K_i(\text{out})$  and  $K_i(\text{in})$  across all regulations  $i$  the Pearson coefficient is the product moment coefficient

$$r = \frac{\sum_i (K_i(\text{out}) - \langle K(\text{out}) \rangle)(K_i(\text{in}) - \langle K(\text{in}) \rangle)}{\sqrt{\sum_j (K_j(\text{out}) - \langle K(\text{out}) \rangle)^2 \sum_i (K_i(\text{in}) - \langle K(\text{in}) \rangle)^2}}. \quad (1)$$

For our data set with  $n = 1750$  regulatory links the obtained  $r = -0.22$  has an estimated standard deviation of about  $1/\sqrt{n-3} = 0.03$ . Therefore we report a highly significant negative correlation between  $K(\text{out})$  of a regulator and  $K(\text{in})$  of its downstream target. The used Spearman rank coefficient equals the Pearson rank coefficient between the ranks of the datasets, in our case between ranking of regulatory links according to the  $K(\text{out})$  of their regulator, respectively their rank according to the  $K(\text{in})$  of its target. The Spearman coefficient only depends on the relative ordering of the variables, and is thus not sensitive to extreme connectivity values.

## Acknowledgments

Work at the Brookhaven National Laboratory was carried out under Contract no DE-AC02-98CH10886, Division of Material Science, US Department of Energy. Work at the Niels Bohr Institute was supported by Danmarks grundforskningsfond center ‘Models of Life’. Kim Sneppen also thanks the Institute for Strongly Correlated and Complex Systems at Brookhaven National Laboratory for hospitality and financial support.

## References

- [1] Lee T I *et al* 2002 *Science* **298** 799–804
- [2] Maslov S and Sneppen K 2002 *Science* **296** 910–3
- [3] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 *Science* **298** 824–7
- [4] Hartwell L H, Hopfield J J, Leibler S and Murray A W 1999 *Nature* **402** C47–52
- [5] Hughes T R *et al* 2000 *Cell* **102** 109–26
- [6] Elena S F and Lenski R E 1997 *Nature* **390** 395–8
- [7] Costanzo M C *et al* 2001 *Nucl. Acids Res.* **29** 75–9
- [8] Yuh C H, Bolouri H and Davidson E H 1998 *Science* **279** 1896–902

- [9] Guelzim N, Bottani S, Bourguin P and Kepes F 2002 *Nat. Genet.* **31** 60–3
- [10] Harris T E 1963 *The Theory of Branching Processes* (Berlin: Springer)
- [11] Newman M E, Strogatz S H and Watts D J 2001 *Phys. Rev. E* **64** 026118
- [12] Fraser H B, Hirsh A E, Giaever G, Kumm J and Eisen M B 2004 *PLoS Biol.* **2** e137
- [13] Giaever G *et al* 2002 *Nature* **418** 387–91
- [14] Fraser H B, Hirsh A E, Steinmetz L M, Scharfe C and Feldman M W 2002 *Science* **296** 750–2
- [15] Kellis M, Patterson N, Endrizzi M, Birren B and Lander E S 2003 *Nature* **423** 241–54
- [16] Jordan I K, Wolf Y I and Koonin E V 2003 *BMC Evol. Biol.* **3** 1
- [17] Hahn M W, Conant G C and Wagner A 2004 *J. Mol. Evol.* **58** 203–11
- [18] Bloom J D and Adami C 2003 *BMC Evol. Biol.* **3** 21
- [19] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 *Nat. Genet.* **31** 64–8
- [20] Bergmann S, Ihmels J and Barkai N 2004 *PLoS Biol.* **2** e9
- [21] Kamada T and Kawai S 1989 *Inf. Process. Lett.* **31** 7–15
- [22] Batagelj V and Mrvar A 1998 *Connections* **21** 47–57