SPARSITY IN MODEL GENE REGULATORY NETWORKS*

Marcin Zagórski

The M. Smoluchowski Institute of Physics, Jagiellonian University
Reymonta 4, 30-059 Kraków, Poland

(Received March 31, 2011)

We propose a gene regulatory network model which incorporates the microscopic interactions between genes and transcription factors. In particular the gene's expression level is determined by deterministic synchronous dynamics with contribution from excitatory interactions. We study the structure of networks that have a particular "function" and are subject to the natural selection pressure. The question of network robustness against point mutations is addressed, and we conclude that only a small part of connections defined as "essential" for cell's existence is fragile. Additionally, the obtained networks are sparse with narrow in-degree and broad out-degree, properties well known from experimental study of biological regulatory networks. Furthermore, during sampling procedure we observe that significantly different genotypes can emerge under mutation–selection balance. All the preceding features hold for the model parameters which lay in the experimentally relevant range.

DOI:10.5506/APhysPolBSupp.4.155
PACS numbers: 87.16.Yc, 87.18.Cf, 87.17.Aa

1. Introduction

Modern DNA sequencing methods help to reveal genomes of animals, plants and microbes. What strikes from these studies is that complex organisms do not have many more genes than simple ones. For instance, human genome consists of about 22000 protein-coding genes, fruit fly has approximately 14000 and baker's yeast has around 6000. Hence, human beings do not even have an order of magnitude more genes than simple eukaryotes. Still we are far more complex than them. The reason behind this mystery might be the way genes work together instead of their number. In

* Presented at the 2nd Summer Solstice International Conference on Discrete Models of Complex Systems, Nancy, France, June 16–18, 2010.

particular, there are strong indications that in eukaryotes and prokaryotes with increasing genome size the number of regulatory genes grows faster than linearly in the total number of genes [1, 2]. Thus it seems reasonable to study the network of interactions between regulatory genes.

After billions of years of evolution Earth's life is a very diverse phenomenon, yet all the living organisms are made of simple building blocks called cells. The single cell is a device designed to interpret internal or external signals in order to enhance its survival prospects. Depending on the situation, *i.e.*, either it is a threat of starvation or some internal damage, this simple functional unit reacts by production of appropriate proteins coded by certain genes. In general, gene products can be divided into three groups: structural proteins, enzymes and transcription factors (TFs). The last group is particularly interesting, since TFs serve only to activate (inhibit) other genes causing the increase (decrease) in a production rate of other proteins. As a result, interactions mediated by TFs form a gene regulatory network (GRN), which is a useful concept to analyse different cell states.

From the studies of real GRNs a couple of qualitative properties transpire: *(i)* a given gene is generally influenced by a "small" number of other genes [4, 5, 6], *(ii)* a few genes regulate many other genes (pleiotropic effect) [6, 7], *(iii)* GRNs seem to be robust to random change, yet they are vulnerable to particular mutations in the genotype [8, 9, 10, 11]. In terms of network terminology, the first two features correspond respectively to narrow in-degree distribution and broad out-degree distribution. Furthermore, there is an evidence that the number of in-going links is governed by exponential distribution [6]. However, if one tries to construct models of networks with high robustness, one usually ends up with dense networks that is graphs having a large number of links compared to number of nodes, which is not the case experimentally. Therefore, most of the models so far have had to build in some limitations to possible connectivity [5, 12].

We have recently proposed a model [3] taking care of interactions between genes and transcription factors which overcomes these shortcomings. Basically, we derive the probability of TF binding to its target site by representing DNA receiving site as a string of four bases (A, G, C, T) which can be recognized by a complementary motif in TF molecule. Moreover, the level of gene expression is determined by deterministic nonlinear dynamics, which under mutation–selection balance produces a set of GRNs performing a given function. By analysing statistical properties of the obtained ensemble of networks we qualitatively recover features found in biological networks, and conclude that resulting *GRNs are as sparse as possible being compatible with their function*. What is also very gratifying, is that all the preceding findings are valid in the biologically relevant range of parameters.

The structure of the article is following. First, we describe the model structure by explaining how our framework works as a whole, rather than giving justification for its separate parts; a detailed model description and discussion can be found in [3]. Second, the process of simulation with mutation–selection balance is presented and the emergence of functional GRNs from random genotype is described. Third, we investigate the statistical properties of the obtained ensemble of networks. Particularly, we analyse the Hamming distance distribution between TF molecules and DNA regulatory sites. Afterwards the network resilience against point mutations is evaluated, leading to very heterogeneous distribution of robustness with only a few "fragile" interactions in the genotype. Not surprisingly, if such an interaction is removed from a GRN, the network is unable to perform its function anymore, so these links are "essential" to GRN viability. We end this section by presenting in-degree distribution which is found to be narrow and possibly with exponential decay. The concept of broad out-degree distribution is mentioned only briefly, just to give insight into a way it emerges in our framework when applying population dynamics. Last, we conclude by summing up all the findings and presenting possible generalisations for the future work.

2. The model

2.1. General framework

In order to derive "design principles" of GRNs from the ensemble of regulatory networks a couple of models have been already used, probably the best known being that proposed by Stuart Kauffman (see [13] and references therein) in which a level of gene expression is a *Boolean* variable (1/0 for on/off). Here, however, we adopt a different strategy by allowing gene expression levels to have intermediate values. Particularly, for ith gene its corresponding normalized expression level is stored in $S_i \in [0, 1]$. Furthermore for N genes we can define a vector variable $\boldsymbol{S} = (S_1, S_2, \ldots, S_N)$ which we refer to as a phenotype. Additionally, we assume that each of N genes is able to produce only one type of transcription factor, and each gene can be influenced by any of these TFs. As a result we obtain a $N \times N$ weight matrix $\boldsymbol{W}$ where a given entry W_{ij} corresponds to the strength of interaction between ith gene and jth TF. Hereafter, we refer to $\boldsymbol{W}$ as the genotype and a formula to determine the values of W_{ij} will be given in the following subsection.

To find gene expression pattern $\boldsymbol{S}(t)$ at any given time t, we propose a deterministic dynamics described by a map $\boldsymbol{S}(t+1) = G\left(\boldsymbol{S}(t), \boldsymbol{W}\right)$, where we call initial phenotype $\boldsymbol{S}(0)$. In the context of discrete dynamical systems, G is the global transition function and $\boldsymbol{S}(t)$ denotes the configuration of the

system at time t. This discrete dynamics can be represented by a sequence of steps

$$\underbrace{\boldsymbol{S}(0)}_{\boldsymbol{S}^{(\mathrm{initial})}} \xrightarrow{\boldsymbol{W}} \boldsymbol{S}(1) \xrightarrow{\boldsymbol{W}} \ldots \xrightarrow{\boldsymbol{W}} \underbrace{\boldsymbol{S}(t) \xrightarrow{\boldsymbol{W}} \boldsymbol{S}(t+1)}_{\text{fixed point phenotype}} ,$$

leading to an attractor which is either a cycle or a fixed point. However, for the following realisation[1], after some transient behaviour we observe almost only fixed points. The corresponding phenotypes are here interpreted as cell's "function". This choice [14] is motivated from early embryo development, where a network starting with some initial expression levels $\boldsymbol{S}^{(\mathrm{initial})}$ "performs a desired function" if and only if it converges to a steady state expression pattern $\boldsymbol{S}^{(\mathrm{target})}$ and maintains there.

In this framework we want to sample the space of all genotypes leading from $\boldsymbol{S}^{(\mathrm{initial})}$ to some fixed $\boldsymbol{S}^{(\mathrm{target})}$. Due to the fact that we work with real-valued phenotypes, we need to introduce a method to determine if a given phenotype is close enough to the target one. Hence, we define a fitness function

$$F(\boldsymbol{S}) = \exp\left(-fD\left(\boldsymbol{S}, \boldsymbol{S}^{(\mathrm{target})}\right)\right), \tag{1}$$

where $D(\boldsymbol{S}, \boldsymbol{S}') = \sum_i \mid S_i - S'_i \mid$ is the difference of expression levels for each gene, and $f \in \mathbb{R}$ is a control parameter.

2.2. Microscopic interactions

Since our aim is to incorporate biological interactions between TFs and DNA strand, we represent each TF as well as each binding site by a character string of length L with characters belonging to a 4 letter alphabet. Following the standard practice [15], we assume that the free energy of one TF molecule bound to its target site is, up to an additive constant, equal εd_{ij}, where ε is the single mismatch energy and d_{ij} is a number of mismatches between ith binding site and jth TF (see Fig. 1). Furthermore, one can define the "interaction strengths" W_{ij} via Boltzmann factor

$$W_{ij} = \mathrm{e}^{-\varepsilon d_{ij}} , \tag{2}$$

with normalizing constant set to 1 (*cf.* [16]). This way we are able to go from molecular genotype, which consists of N TFs and N^2 binding sites each of length L to the weight matrix $\boldsymbol{W}$. If there were exactly one TF molecule Eq. (2) would correspond to probability of finding this molecule attached to its target site. In case of n_j TFs of jth type one can derive [16, 17]

[1] The generalisation including repressors and possible cycles is discussed in the last section.

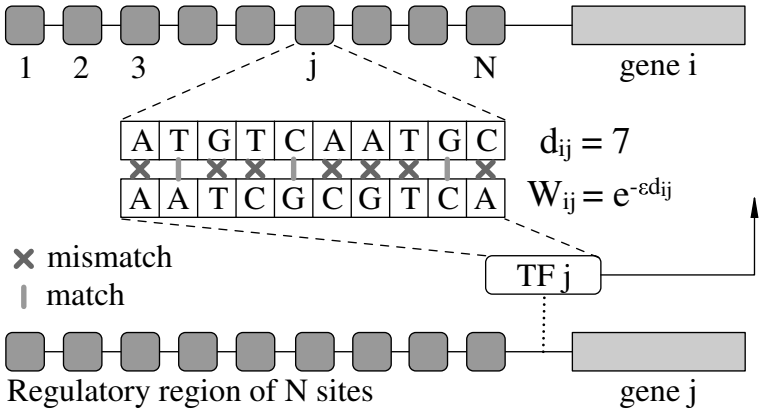

Fig. 1. Schematic representation of the regulatory region of gene i with N binding sites. Represented is the interaction W_{ij} mediated by the binding of TF j to the jth site of that region. Here the string representing the TF is of length $L = 10$, number of mismatches (non-complementary DNA bases) $d_{ij} = 7$, and the resulting W_{ij} is calculated according to Eq. (2).

the probability p_{ij} that precisely one of them is bound to the binding site of ith gene

$$p_{ij} = \frac{1}{1 + 1/(W_{ij} n_j)} = \frac{1}{1 + \exp\left(\varepsilon d_{ij} - \ln(n S_j)\right)}, \tag{3}$$

which dependence is known to physicists as Fermi function (Fig. 2). In the above formula, for the sake of simplicity, we assume that $n_j = nS_j$ where n is a model parameter representing the number of TFs. For living cells the number of transcription factors varies from a few to even an order of 10^4 with no typical value. As for other parameters used in the model, the experimentally found values lead to $10 \leq L \leq 15$ and ε between one and three. Interestingly, from thermodynamics and kinetics of TF-DNA interaction one can derive physical constraints, which lead to roughly the same range of parameters [16]. For instance, the low boundary on L comes from the demand to have at most a number of order one of "perfect" (with no mismatches) binding sites in random genome. This condition is equivalent to $l_g(1/4)^L \leq 1$, where l_g denotes the number of base pairs in a genome; in case of bacteria $l_g \sim 10^6$, so $L \geq 10$. Thus, for the subsequent results we use by default $L = 12$, $\varepsilon = 2.0$, and a range of $10 \leq n \leq 10^4$.

To model the dynamics of GRN, we consider the level of ith gene expression $S_i(t+1)$ to strongly depend on the transcription rate of gene i at time t. This assumption is supported by biological argument, basically, that transcription rate is strongly associated with the degree to which the gene's regulatory region is occupied by TFs. To keep the framework simple, we

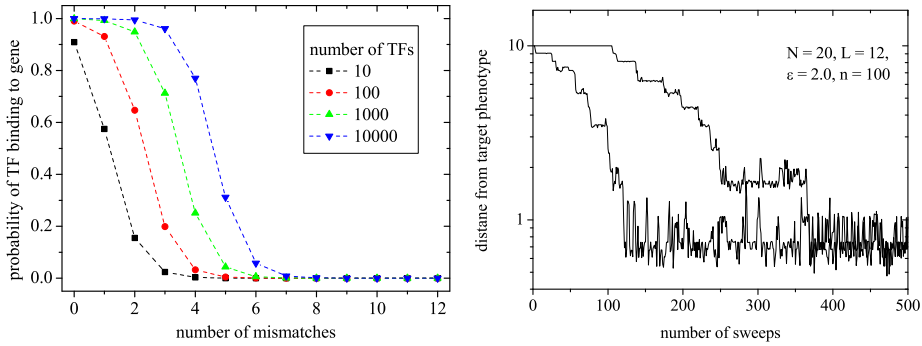

Fig. 2. Left: Probability of finding TF attached to its target site for different values of n; following Eq. (3) with $S_j = 1$ this probability is a Fermi function. Right: Shown is the distance between stationary expression pattern $\boldsymbol{S}$ and $\boldsymbol{S}^{(\mathrm{target})}$ for two trajectories representing the emergence of functional network from a random genotype.

assume that gene i transcription is "on" whenever at least one TF is bound within its regulatory region and otherwise it is "off". This is reminiscent of an "OR" logical gate whereby the output is "on" if and only if at least one of the inputs is "on". The (normalized) mean expression level of a gene is then identified with the probability that transcription is "on". Hence, for gene i we have

$$S_i(t+1) = 1 - \prod_j (1 - p_{ij}(t))\,, \tag{4}$$

where

$$p_{ij}(t) = \frac{1}{1 + \exp\left(\varepsilon d_{ij} - \ln(nS_j(t))\right)}\,. \tag{5}$$

The above holds for TF occupancies in the regulatory regions being statistically independent. This means that any strong correlations in transcription regulation in real systems may be out of reach, yet we claim the qualitative properties of GRN are not affected by the independency assumption, which seems to be the case according to results presented in the following sections. Moreover, we have restricted the transcriptional logic to be of the "OR" type. Our TF thus act only as enhancers, never as repressors, and they do not cooperate using more complicated logic [18]. However, it is possible to include repressors in our framework, which generalization we briefly discuss in the Conclusions and outlooks section. Nevertheless, for this paper findings it is enough to work within minimal hypothesis.

3. Mutation–selection balance

As already mentioned, we want to sample the set of all genotypes leading from $\boldsymbol{S}^{(\mathrm{initial})}$ to $\boldsymbol{S}^{(\mathrm{target})}$, which GRNs we call viable. For simplicity, we shall choose the $\boldsymbol{S}_i^{(\mathrm{initial})}$ and $\boldsymbol{S}_i^{(\mathrm{target})}$ for $i = 1, \ldots, N$ to be 1 ("on") or 0 ("off"). If the initial and target phenotype are drawn at random, the number of components set to 1 will be approximately equal to that set to 0 for large N. Hence, to reduce finite size effects in N, these numbers are set exactly to $N/2$. In order to take into account the permutation symmetry of the model we work with $\boldsymbol{S}_i^{(\mathrm{initial})} = 1$ for $i \leq N/2$ and 0 otherwise; furthermore we also impose without loss of generality $\boldsymbol{S}_i^{(\mathrm{target})} = 1$ for $N/4 < i \leq 3N/4$ and 0 otherwise. Notice that $\sum_i \boldsymbol{S}_i^{(\mathrm{initial})} = \sum_i \boldsymbol{S}_i^{(\mathrm{target})} = N/2$ and $\sum_i \boldsymbol{S}_i^{(\mathrm{initial})} \boldsymbol{S}_i^{(\mathrm{target})} = N/4$; we typically use $N = 20$ genes.

Since the space of viable GRNs is only a tiny fraction of the space of all regulatory networks, we need to introduce some effective sampling method. In particular, we use Metropolis random walk algorithm to explore this ensemble. The procedure is following, we start with random genotype that is we draw all the characters representing gene's regulatory regions and TF molecules randomly, and calculate the corresponding weight matrix $\boldsymbol{W}$. Next, with each step we apply a point mutation to characters representing DNA binding sites, recalculate $\boldsymbol{W}$, and according to Eq. (4) the associated fixed point phenotype $\boldsymbol{S}$. Afterwards, having $\boldsymbol{S}$ we compute fitness of the genotype and accept or reject the attempted move according to Metropolis acceptance probability.

This way by applying mutation–selection balance we obtain, after some initial period, a series of viable genotypes. In Fig. 2 (right) one can trace trajectories of gradual emergence of viable genotype from completely random background. This process can be explained by gradual activation of genes due to appearance of strong interactions in the genotype. Particularly, we refer to an interaction W_{ij} being strong if according to Eq. (3) the related $p_{ij} > 1/2$ for $S_j = 1$. In other words, for a given n and ε we can define a critical value $d_h = [\ln(n)/\varepsilon]$, and any mismatch, let us call it subcritical, that is less or equal to d_h corresponds to a strong interaction. For instance, if the subcritical mismatch appears on the diagonal interaction, it acts as a self-excitatory regulation and the corresponding gene becomes expressed with no need of activation from other genes. This concept of subcritical mismatches being responsible for strong activatory regulation can be extended to non-diagonal interactions, but then it is very difficult to analytically track the leading behaviour of the system. Therefore, we should keep in mind that these subcritical mismatches play a crucial role in the set of viable GRNs, and in the subsequent section we provide a numerical evidences to support this statement.

4. Results

Having obtained the ensemble of viable GRNs, it is interesting to see the distribution $p(d)$ of Hamming distance between DNA binding sites and associated TF molecules. In case of random choice of characters representing the molecular genotype one expects

$$p(d) = \binom{L}{d}(1/4)^{L-d}(3/4)^{d} \,. \tag{6}$$

The above binomial distribution has a very low probability of observing small mismatches. Nevertheless, under selection pressure we expect to see a significant number of functional sites with strong interactions, which are associated with subcritical mismatches. Indeed in Fig. 3 we have apart from binomial part a peak for low values of d. Furthermore, qualitatively the same distribution is observed in studies of transcription factors binding energies [19] where the peak at low energy values corresponds to good matching of TF to DNA target sequence. A similar mechanism of the appearance of a single separated peak in the distribution has also been seen in the balls-in-boxes model [20].

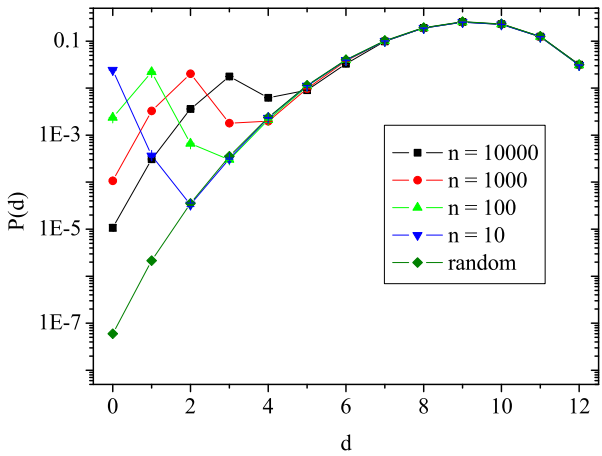

Fig. 3. Distribution of the Hamming distance between a TF and the receiving DNA site for $N = 20$, $L = 12$, $\varepsilon = 2.0$ and for various values of n. Random case corresponds to binomial distribution. The lines are to guide the eye.

Since regulatory networks are very robust to mutations and environmental changes [11, 22, 23], it is interesting to ask what are the consequences of removing one of the interactions from the genotype. In our framework it corresponds to setting $W_{ij} = 0$ for the selected pair (i, j) and finding the phenotype produced by the modified genotype. If the loss of interaction

leads GRN to be a non-viable network, we refer to this interaction as "essential". Moreover, we find that as soon as n is not too large, there is almost always just one essential interaction per row as shown in Fig. 4 (left).

We also consider a stronger measure of essentiality: we ask that viability be lost when the interaction's mismatch is increased by one. Remarkably, the rule "one essential incoming interaction per gene" generally holds here too. The average robustness of fitness with respect to binding site mutations, R_{bs}, is readily estimated: there are $N/2$ sensitive interactions out of N^2, hence one expects $R_{\mathrm{bs}} \approx 1-1/(2N)$. And indeed we find $R_{\mathrm{bs}} = 0.977(2)$ for $N = 20$ (with some weak dependence on n in the third decimal). Thus vast majority of mutations (roughly 97.5%) have no consequence on the fitness, while mutations in the essential interactions are typically deleterious.

One can also interpret the normalized histogram of essential interactions as the in-degree distribution for the associated GRN. In Fig. 4 (right) one can see that the distribution of in-going essential connections qualitatively follows an exponential distribution, which is in agreement with biological findings [6]. On the other hand, the out-degree distribution is of the power law nature, indicating that a small number of TFs regulate a large number of target genes, whereas most TFs regulate few or no target genes [24]. Particularly, such differences between in- and out-degree distribution have been observed for *E. coli* [4] and yeast (*S. cerevisiae*) [25].

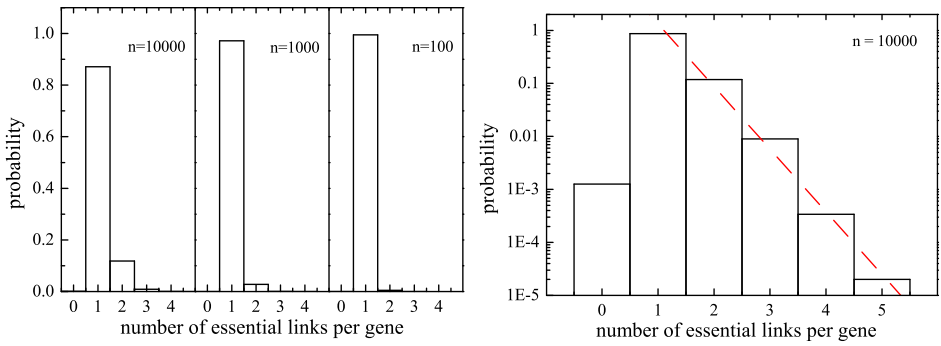

Fig. 4. Left: Probability distribution of the number of essential interactions per row of the matrix specifying a viable network for $N = 20$, $L = 12$, $\varepsilon = 2.0$ and a range of values of n. Right: The same statistics can be also interpreted as in-degree distribution. Data is presented for $n = 10000$ with the dashed line being an exponential fit.

Additionally, from a given viable genotype we can extract the essential interactions and construct the corresponding GRN. Outcome of such a procedure for a small system with $N = 8$ and all genes being expressed in the target phenotype is shown in Fig. 5. At this point one remark should be

made, up to now we have studied GRNs obtained from Metropolis sampling with point mutations made only in strings coding DNA regulatory regions with strings coding TFs being fixed; TF evolve generally slowly compared to DNA binding sites [21]. Notice that when a TF is modified in our framework, a whole column of the matrix $\boldsymbol{W}$ is affected at once. Therefore, we do not observe any selection pressure on the number of essential interactions per column (out-degree of a given gene). Particularly, the interesting problem is to see how the number of out-going links changes in GRN when we simulate a population of genotypes against mutations in TF coding genes. Since mutations in any TF which is associated with at least one essential interaction are highly deleterious, the selection condition favors GRNs with essential links unevenly distributed among columns. From this qualitative argumentation we can see that broad out-degree distributions should be favored, and indeed by doing numerical calculations [3] we arrive at the same conclusion.

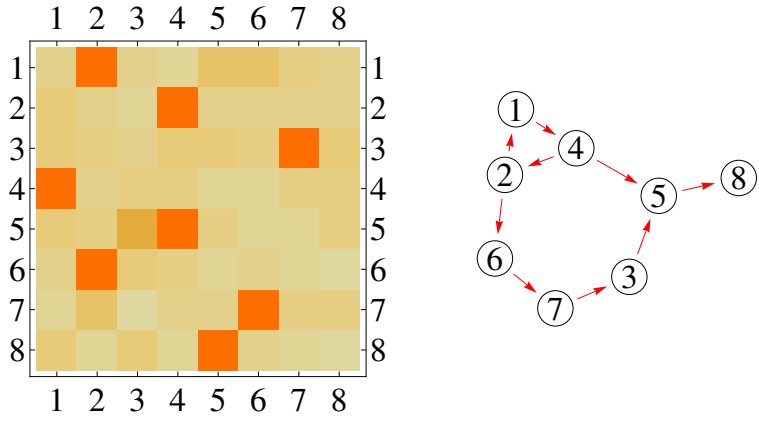

Fig. 5. Left: Genotype with associated weight matrix $\boldsymbol{W}$ for $N = 8$, $L = 12$, $\varepsilon = 2.0$ and $n = 1000$. The dark (brownish) squares represent strong interactions. Right: GRN extracted from this genotype with only essential interactions indicated.

5. Conclusions and outlooks

We have considered a relatively simple model of GRNs in which molecular information is used to obtain a matrix of interactions between DNA binding sites and TF molecules. Furthermore, based on this matrix we proposed a way to find a gene expression pattern reflecting cell's function. Using the dynamics defined in Eq. (4) and Metropolis sampling method we obtained under mutation–selection balance the ensemble of viable genotypes. By investigating the statistical properties of the GRNs we found many similarities

with real regulatory networks: the obtained networks are sparse with narrow in-degree and broad out-degree, network robustness is heterogeneously distributed (mainly only mutations in essential interactions are deleterious), and the results hold for parameters in biologically relevant range.

Although our modeling involves certain idealizations, we have insisted on including interactions through the biophysical mechanism of molecular recognition and affinity. The resulting reasonable GRN topology is the consequence of several causes: the viability constraint, the low probability of a small mismatch between TF and the binding site on the DNA, the size L of this segment, the not-too-small spacing (in units of $k_{\mathrm{B}}T$) between the energy levels that determine the strength of TF-DNA interactions, and finally the value of the parameter n itself which enters the dynamics.

In addition, by introducing the concept of "essential interactions" we are able to quantify the sparsity level of GRNs, and we find that in most cases there is only one "large" interaction per gene although networks are still evolvable and can reach very diverse topologies. Basically, our principal result is that regulatory networks are as sparse as possible being compatible with a given function. We have also generalised the model by introducing repressors and more than one target phenotype, and the preliminary results suggest the above rule still holds. Furthermore, in the extended model we observe various positive and negative circuits [26, 27] which occurrence depends on the network function that can be realised by either multiple target phenotypes or cyclic like behaviour.

Project operated within the Foundation for Polish Science International Ph.D. Projects Programme co-financed by the European Regional Development Fund covering, under the agreement No. MPD/2009/6, the Jagiellonian University International Ph.D. Studies in Physics of Complex Systems. This work was supported by the Polish Ministry of Science and Higher Education Grant No. N N202 229137 (2009-2012).

REFERENCES

[1] E. Van Nimwegen, *Trends Genet.* **19**, 479 (2003).

[2] M.M. Babu *et al.*, *Current Opinion in Structural Biology* **14**, 283 (2004).

[3] Z. Burda, A. Krzywicki, O.C. Martin, M. Zagorski, *Phys. Rev. E* **82**, 011908 (2010).

[4] D. Thieffry, A. Huerta, E. Perez-Rueda, J. Collado-Vides, *BioEssays* **20**, 433 (1998).

[5] S. Kauffman, C. Peterson, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 17102 (2004).

[6] G. Balazsi, A. Heath, L. Shi, M. Gennaro, *Mol. Syst. Biol.* **4**, 225 (2008).

[7] A. Barabási, Z. Oltvai, *Nature Rev. Genet.* **5**, 101 (2004).

[8] S. Bornholdt, K. Sneppen, *Proc. R. Soc. Lond. B Biol. Sci.* **267**, 2281 (2000).

[9] J.A. Edlund, C. Adami, *Artif. Life* **10**, 167 (2004).

[10] A. Wagner, *Robustness and Evolvability in Living Systems*, Princeton University Press, Princeton, NJ. (2005).

[11] M. Chaves, R. Albert, E.D. Sontag, *J. Theor. Biol.* **235**, 431 (2005).

[12] M. Aldana, P. Cluzel, *Proc. Natl. Acad. Sci. USA* **100**, 8710 (2003).

[13] S. Kauffman, *Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, Oxford 1993.

[14] A. Wagner, *Evolution (Lawrence, Kans.)* **50**, 1008 (1996).

[15] P. von Hippel, O. Berg, *Proc. Natl. Acad. Sci. USA* **83**, 1608 (1986).

[16] U. Gerland, J. Moroz, T. Hwa, *Proc. Natl. Acad. Sci. USA* **99**, 12015 (2002).

[17] M. Lässig, *BMC Bioinformatics* **8**, S7 (2007).

[18] N. Buchler, U. Gerland, T. Hwa, *PNAS* **100**, 5136 (2003).

[19] V. Mustonen, J. Kinney, C. Callan, M. Lässig, *Proc. Natl. Acad. Sci. USA* **105**, 12376 (2008).

[20] P. Bialas, Z. Burda, D. Johnston, *Nucl. Phys.* **B493**, 505 (1997).

[21] M. Huynen, P. Bork, *Proc. Natl. Acad. Sci. USA* **95**, 5849 (1998).

[22] U. Alon, M.G. Surette, N. Barkai, S. Leibler, *Nature* **397**, 168 (1999).

[23] C. Espinosa-Soto, P. Padilla-Longoria, E. Alvarez-Buylla, *Plant Cell* **16**, 2923 (2004).

[24] R. Albert, *J. Cell Sci.* **118**, 4947 (2005).

[25] N. Guelzim, S. Bottani, P. Bourgine, F. Képès, *Nat. Genet.* **31**, 60 (2002).

[26] R. Thomas, *Int. J. Dev. Biol.* **42**, 479 (1998).

[27] U. Alon, *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman and Hall/CRC, Boca Raton, FL, 2007.