Emergence of Motifs in Model Gene Regulatory Networks

Marcin Zagórski

Institute of Physics, Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland
marcin.zagorski@uj.edu.pl

Abstract. Gene regulatory networks arise in all living cells, allowing the control of gene expression patterns. The study of their circuitry has revealed that certain subgraphs of interactions or motifs appear at anomalously high frequencies. We investigate here whether the overrepresentation of these motifs can be explained by the functional capabilities of these networks. Given a framework for describing regulatory interactions and dynamics, we consider in the space of all regulatory networks those that have a prescribed function. Markov Chain Monte Carlo sampling is then used to determine how these functional networks lead to specific motif statistics in the interaction structure. We conclude that different classes of network motifs are found depending on the functional constraint (multi-stability or oscillatory behaviour) imposed on the system evolution. The discussed computational framework can also be used for predicting regulatory interactions, if only the experimental gene expression pattern is provided.

Keywords: gene regulatory networks, network motifs, transcription factors, cell cycle.

General Description

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. One of the key mechanisms responsible for processing available information are regulatory interactions between genes. For instance, when a yeast cell finds itself in the environment rich in sugar it starts to produce enzymes to process this nutrient into energy. If we go down to a molecular level, the sugar presence or absence can be treated as an input signal for a cell's processing unit, *i.e.* gene regulatory network (GRN). The set of interactions between genes along with the gene expression machinery allows all living cells to control their gene expression patterns. In the last decade, our knowledge how any given gene can affect another's expression has been significantly extended through various experiments. For example, small gene networks have been constructed to implement simple functions *in vivo* [3,4], and much larger sets of interactions have been derived from a number of organisms [6,13,7].

L. Vanneschi, W.S. Bush, and M. Giacobini (Eds.): EvoBIO 2013, LNCS 7833, pp. 212–215, 2013.

Therefore it has been possible to show that several subgraphs of interactions ("motifs") arise more frequently than might be expected [14,9,8,16]. In a very recent study [5], the motif statistics were reported for the human regulatory network, indicating overrepresentation of certain structures. Hence, the question of design principles or conditions under which certain motifs appear in biological networks is of great interest.

In [2] by Z. Burda, A. Krzywicki, O.C. Martin and myself published in PNAS 108, 17263-17268 (2011) we propose a computational framework within which gene regulatory networks with a predefined functional capabilities can be sampled *in silico*. Thus, it is possible to study various statistical properties of networks generated with certain constraints imposed. Specifically, the proposed model incorporates microscopic interactions between genes and transcription factors through a weight matrix (genotype). Next, the gene's expression level is determined by deterministic synchronous dynamics with contribution from both excitatory and inhibitory interactions. Having defined transcriptional dynamics and providing initial gene expressions, we can easily obtain gene expression pattern (phenotype) which is interpreted as a function of GRN. By imposing some arbitrary target gene expression pattern (target phenotype) we would like to know which genotypes do lead to this predefined target. In practice, we quantify how well a given genotype is adapted to the imposed pattern by a fitness function depending on a distance between target phenotype and phenotype produced by that genotype.

The main computational difficulty lays in a problem: how from the huge space of all genotypes obtain a sample of genotypes with a high fitness? This goal can be achieved by Markov Chain Monte Carlo (MCMC) method, which generates a biased random walk in the space of genotypes, enforcing at each step the accept/reject Metropolis rule [11]. Note that the MCMC introduces no bias: at large times the *a priori* specified distribution is obtained exactly. Hence, we can understand how phenotypic properties constrain the genotypes, in particular at the level of the architecture of the genetic interactions. Additionally, we can use Metropolis rule to determine which of genetic interactions described by genotype are essential for its functionality. If the element of the genotype matrix corresponding to interaction strength between two genes is set to zero, and the viability of network is lost (the fitness drops and the Metropolis rule rejects modified genotype), the interaction between these genes is considered essential. As a result, a set of all essential interactions that constitutes the gene regulatory network for the underlying genotype is obtained.

A very gratifying point is that obtained GRNs are evolvable and a given target expression pattern can be realized through different topologies. Particularly, we consider two classes of constraints which resemble two types of biological processes: (i) different stable gene expression patterns can be interpreted as different types of cells during cell development, (ii) cyclic gene expression is characteristic for cell cycle, where different genes are excited/inhibited during different stages of cell division process. In order to reveal significant network motifs we compare the number of subgraphs of a given type between generated GRNs and their

randomized versions. The randomization used is that proposed by Maslov and Sneppen [10]: edges are interchanged so that both the in- and out-degrees of network nodes remain unchanged.

In the case of multistability the two node motif with genes being mutually inhibitory and self-activating (double negative feedback loop with autoregulation) is found to be of great importance. Typically in the resulting GRNs, there is one such motif for two fixed points imposed, two for three fixed points imposed and three in case found four steady states. The randomized networks almost always do not have any motifs of this type (see [2] for exact frequencies). Interestingly, this simple network motif is found in various biological gene networks, with a good example being the genetic switch between lysogeny and lysis of the phage λ [12]. Clearly such a pair of genes acts as a bistable between situations with one gene being "on" and the other being "off". When embedded in the whole network this type of motif influence other genes in a downstream effect along the associated tree-like graph structure.

In the case of target phenotypes being periodic in time the bistable switch is not present, and four node motifs like bifan, diamond and "frustrated" loop appear and are highly overrepresented compared to randomized networks (see Fig. 1 for graphical representation). Again, biological gene networks have been found containing some of these motifs [1] the bifan motif being perhaps the most prominent. The function of this motifs treated separately can be understood only for the small network sizes. However, for networks with several genes (as in the discussed study [2]) it is necessary to consider how these motifs cooperate within the overall network, just like parts in a larger machine. More importantly, none of motifs overrepresented for periodic gene expression pattern imposed was found significant in the multiple fixed point scenario, and vice-versa.

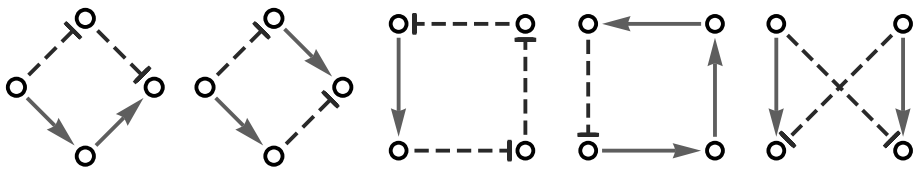

Fig. 1. Network motifs overrepresented in case of time periodic gene expression: incoherent diamonds (from the left: 1st and 2nd), frustrated four-node loops (3rd and 4th), incoherent bifan (5th). The arrows represent activatory (solid) and inhibitory (dashed) interactions.

Hence, we can conclude that different classes of motifs are observed for different types of functional capabilities of GRN. This result is very striking if we realize that no motif structures are incorporated inside the presented framework on any level. Instead motifs emerge from purely random background due to imposed functional patterns and selection pressure. Within the proposed computational framework it is also possible to impose gene expression patterns taken from experimental works (recently we have applied our model to cell cycle profiles of

two yeast species and mammals [15]). Specifically, the question of probability of observing certain interaction between selected genes can be addressed, so the model can be also used as a tool for network structure prediction.

References

1. Alon, U.: Network motifs: theory and experimental approaches. Nature Reviews Genetics 8, 450 (2007)
2. Burda, Z., Krzywicki, A., Martin, O.C., Zagorski, M.: Proc. Natl. Acad. Sci. U.S.A. 108, 17263 (2011)
3. Elowitz, M., Leibler, S.: A synthetic oscillatory network of transcriptional regulators. Nature 403, 335 (2000)
4. Gardner, T., Cantor, C., Collins, J.: Construction of a genetic toggle switch in *Escherichia coli*. Nature 403, 339 (2000)
5. Gerstein, M.B., et al.: Architecture of the human regulatory network derived from ENCODE data. Nature 489, 91 (2012)
6. Herrgard, M., Covert, M., Palsson, B.: Reconstruction of microbial transcriptional regulatory networks. Current Opinion in Biotechnology 15, 70 (2004)
7. Hu, Z., Killion, P., Iyer, V.: Genetic reconstruction of a functional transcriptional regulatory network. Nature Genetics 39, 683 (2007)
8. Lee, T., et al.: Transcriptional regulatory networks in *Saccharomyces cerevisiae*. Science 298, 799 (2002)
9. Ma, H., et al.: An extended transcriptional regulatory network of *Escherichia coli* and analysis of its hierarchical structure and network motifs. Nucleic Acids Research 32, 6643 (2004)
10. Maslov, S., Sneppen, K.: Specificity and stability in topology of protein networks. Science 296, 910 (2002)
11. Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., Teller, E.: Equation of state calculations by fast computing machines. Journal of Chemical Physics 21, 1087 (1953)
12. Ptashne, M.: A Genetic Switch: Phage λ Revisited. Cold Harbor Spring Laboratory Press, NY (2004)
13. Salgado, H., et al.: Regulondb (version 5.0): *Escherichia coli* k-12 transcriptional regulatory network, operon organization, and growth conditions. Nucleic Acids Research 34, D394 (2006)
14. Shen-Orr, S., Milo, R., Mangan, S., Alon, U.: Network motifs in the transcriptional regulation network of *Escherichia coli*. Nature Genetics 31, 64 (2002)
15. Zagorski, M., Krzywicki, A., Martin, O.C.: Edge usage, motifs and regulatory logic for cell cycling genetic networks. Phys. Rev. E 87, 012727 (2013)
16. Zhu, J., et al.: Integrating large-scale functional genomic data to dissect the complexity of yeast regulatory networks. Nature Genetics 40, 854 (2008)