Pis'ma v ZhETF, vol. 93, iss. 1, pp. 43 – 47 © 2011 January 10

Are the Complex Distributed Genetic Networks Inherently Oscillatory and Chaotic?

V. P. Zhdanov[1)]

Department of Applied Physics, Chalmers University of Technology, S-412 96 Göteborg, Sweden

Boreskov Institute of Catalysis RAS, 630090 Novosibirsk, Russia

Submitted 16 November 2010

Complex mRNA-protein networks with distributed regulation of gene transcription are often mimicked by random Boolean networks. According to this mapping, the gene-expression kinetics are expected to depend strongly on the network connectivity, m. Specifically, the kinetics are predicted to be oscillatory or chaotic for $m > 2$. We analyze this problem in more detail by using an accurate model explicitly describing the mRNA-protein interplay with randomly organized transcriptional regulation. Contrary to the expectations, our mean-field calculations and Monte Carlo simulations performed for $m = 1$, 2 and 4 show that oscillations and chaos are lacking irrespective of m.

Complex genetic networks (CGNs) include many genes transcribed to mRNAs and regulated by proteins synthesized via mRNA translation. In addition, CGNs may include non-coding RNAs (ncRNAs). The understanding of CGNs is still limited [1 – 4] (for simple genetic networks, see reviews [5]). In general, the behaviour of CGNs depends on the details of regulation. Qualitatively, one can distinguish two globally different types of regulation and many more specific situations somewhere between. On one side, there are networks with hierarchical (or, in other words, "autocratic") regulation of downstream genes by upstream genes. Such networks are typically organized in the shape of a pyramid, with each row of genes linked to genes directly beneath it. On the other side, there are distributed (or, in other words, "democratic") networks relying on mutual regulation of genes. The former networks have been studied experimentally in more detail because their structure and function are often more evident. The latter networks gained empirical support only recently, partly because their characterization involves studies of genome-wide dynamical processes [1].

In the corresponding theoretical studies, focused primarily on the mRNA-protein unterplay, CGNs are usually treated by employing the Boolean models (see, e.g., reviews [2, 3], recent articles [6], and references therein). In this coarse-grained approach, mRNAs and proteins are not described explicitly. Each gene is instead represented by a node which can be either active or inactive. Each node is connected with one or a few other nodes. The time is changed discretely. During each step, the states of all the nodes are changed (usually simultaneously) depending on the states of other nodes. The distributed and hierarchial CGNs are mimicked by using random and scale-free topologies, respectively.

The generic Boolean model of CGNs with distributed regulation implies that each node is governed by m randomly chosen nodes and the Bolean function determining this regulation is random. One of the most interesting and conceptually important predictions, gained by studying such random Boolean networks, is that there is a critical value of m, $m_c = 2$ [7]. At $m \leq m_c$, the model typically predicts short circles which can be interpreted as a unique stable steady state. For $m > m_c$, one can observe long circles with the length increasing with m. The latter behaviour can be identified with oscillations or chaos. In reality, genes often have a few regulatory sites. Thus, one could expect that CGNs with distributed regulation would often be oscillatory or chaotic.

Genetic networks can also be analyzed by using the kinetic models explicitly describing gene regulation and mRNA and protein populations in the framework of the mean-field (MF) approximation or Monte Carlo (MC) simulations [2, 3, 5]. These two approaches are obviously more accurate compared to the Boolean formalism. The numerous examples of their use are, however, primarily focused on the interplay of a few genes. In this communication, we present the MF and MC calculations illustrating the likely behavior of the mRNA-protein networks including many genes with distributed regulation chosen at random. To specify the meaning of "many" in our context, we recall, for example, that in eukaryotes the number of protein-coding genes is about $3 \cdot 10^4$. This huge array of genes can be divided into many groups (see, e.g., networks shown in Ref. [8]) so that the protein-mediated interactions inside groups are

1)e-mail: zhdanov@chalmers.se

intensive while the interactions between groups are apparently sporadic. Such groups often contain from 20 to 40 genes. Taking into account that the connectivity can be appreciable, $m \geq 1$, only inside these groups, we aim our model to single groups and accordingly consider that a network is large if the number of genes, n, is between 20 and 40. In the Boolean models, the critical connectivity was studied in larger networks, e.g., with $n \simeq 100$ in Ref. [7]. To illustrate biologically interesting situations and to compare with the predictions of the Boolean models, we present results for $n = 30$ and 100.

Our model includes standard steps of transcription of n genes to mRNAs, mRNA translation to proteins (P_i; $1 \leq i \leq n$), and mRNA and protein degradation,

$$\mathrm{Gene}_i \to \mathrm{Gene}_i + \mathrm{mRNA}_i, \tag{1}$$

$$\mathrm{mRNA}_i \to \mathrm{mRNA}_i + \mathrm{P}_i, \tag{2}$$

$$\mathrm{mRNA}_i \to \emptyset, \tag{3}$$

$$\mathrm{P}_i \to \emptyset. \tag{4}$$

The corresponding MF kinetic equations for the mRNA and protein populations in a cell are represented as

$$dN_i/dt = w_i - k_i N_i, \tag{5}$$

$$dn_i/dt = v_i N_i - \kappa_i n_i, \tag{6}$$

where w_i is the transcription rate, v_i is the translation rate constant, and k_i, and κ_i are the degradation rate constants.

To mimic distributed networks, we should specify the regulation of the mRNA synthesis by proteins or, in other words, the dependence of the rates of the mRNA synthesis, w_i, on the protein populations (the other parameters are considered to be constant). In analogy with the random Boolean networks, we consider that (i) each gene is governed by m randomly chosen genes (including self-regulation), (ii) the type (positive or negative) of each channel of regulation is determined at random, and (iii) the regulation channels operate independently. To illustrate the effect of the connectivity on the kinetics of gene expression, we have analyzed the networks with $m = 1$, 2 and 4.

Physically, m represents a number of regulatory sites of each gene. The independent operation of the regulation channels practically means that the protein attachment to a regulatory site occurs independently of the state of other sites. In this case, the dependence of w_i on the protein populations can be factorized. In particular, w_i can be represented for $m = 1$, 2 and 4, respectively, as

$$w_i = w_i^b + w_i^\circ F_1(n_1), \tag{7}$$

$$w_i = w_i^b + w_i^\circ F_1(n_1) F_2(n_2), \tag{8}$$

$$w_i = w_i^b + w_i^\circ F_1(n_1) F_2(n_2) F_3(n_3) F_4(n_4), \tag{9}$$

where w_i^b is the so-called basal synthesis rate, w_i° is the maximum value of the regulated rate, and $F_1(n_1)$, $F_2(n_2)$, $F_3(n_3)$ and $F_4(n_4)$ are the dimensionless functions (≤ 1) dependent on the protein populations, n_1, n_2, n_3 and n_4 (the locally used subscripts 1, 2, 3 and 4 correspond to proteins regulating gene i). Each function describes either positive or negative regulation, respectively, as

$$F_j(n_j) = n_j/(K_j + n_j), \tag{10}$$

and

$$F_j(n_j) = K_j/(K_j + n_j), \tag{11}$$

where K_j is the protein association constant. Physically, $K_j/(K_j + n_j)$ and $n_j/(K_j + n_j)$ represent the probabilities that a regulatory site is free or occupied by a protein, respectively. The use of these expressions implies that the protein attachment to and detachment

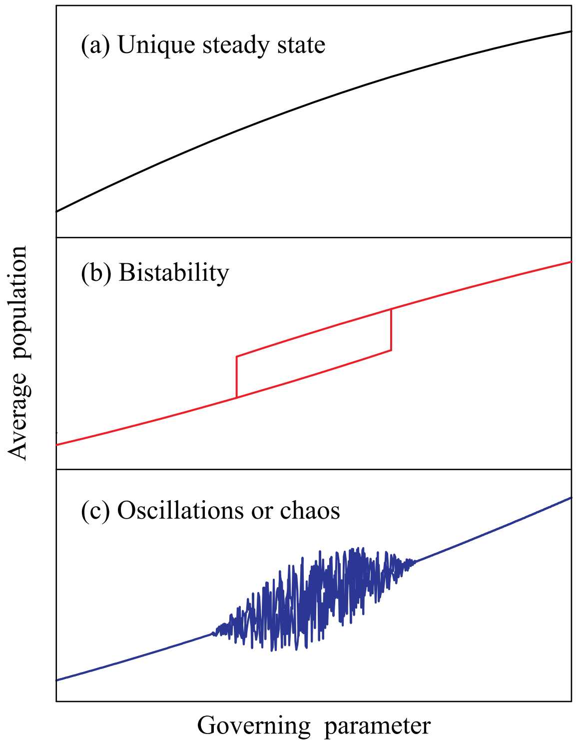

Fig.1. Schematic dependence of the average mRNA or protein population on a governing parameter in the cases of (a) a unique and stable steady state, (b) bistablity, and (c) oscillations or chaos

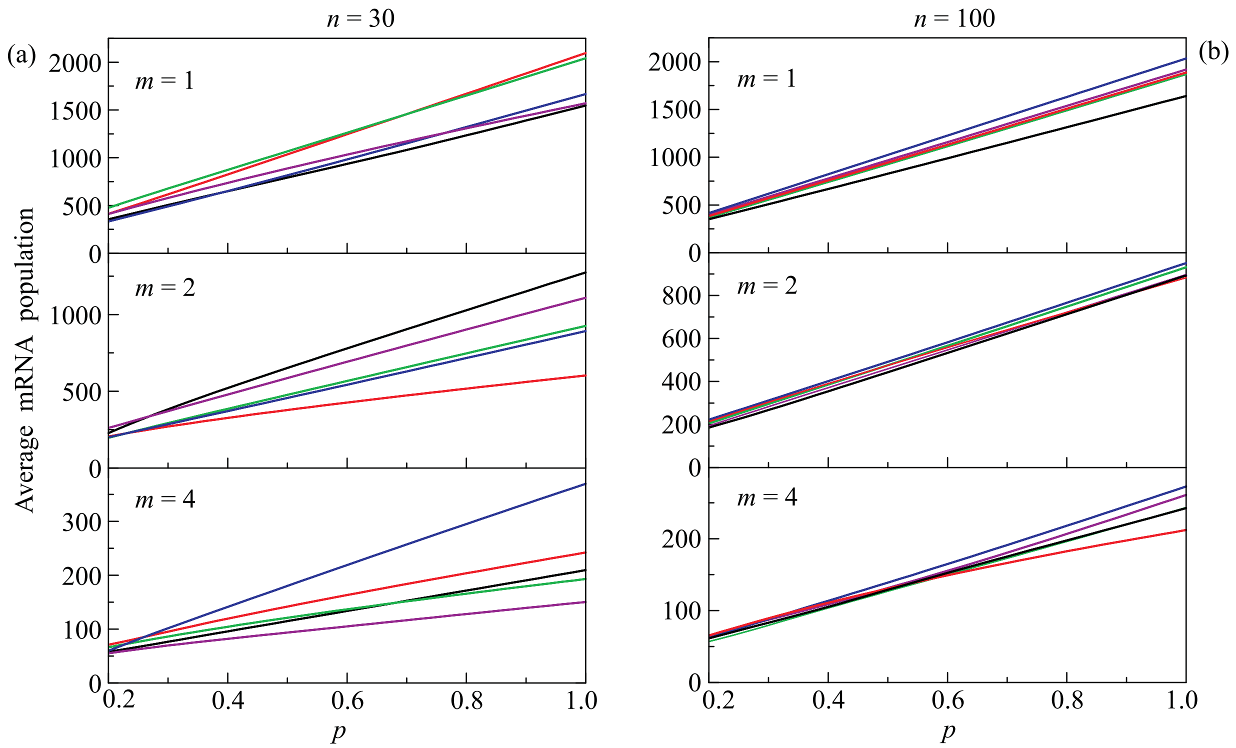

Fig.2. Average mRNA population as a function of p for $n = 30$ (a) and 100 (b). The average protein population (not shown) is about two times higher. Each panel (for $m = 1$, 2 and 4) exhibits the results obtained in five runs of MF calculations with different sets of parameters

from regulatory sites is rapid and these steps are close to equilibrium. This condition is standard in the theory of gene regulation [9].

Steps (1)–(4) and Eqs. (5)–(11) define our model. To specify the corresponding model parameters, we notice that the mRNA and protein degradation usually occurs on the time scale between a few minutes and one hour, the corresponding rate constants are accordingly in the range 0.01–0.1 min^{-1}, and the protein populations are often comparable or somewhat higher than the corresponding mRNA populations [10]. For these reasons, the values of the mRNA and protein degradation rate constants are chosen at random in the range from 0.05 to 0.15 min^{-1}, while the rate constants of the protein synthesis are selected at random in the range from 0.1 to 0.3 min^{-1}. In this case, on average, the protein population is about two times higher than the mRNA population.

The parameters determining the transcription rates have been chosen following three guidelines. (i) The average mRNA and protein populations should be biologically reasonable. (ii) The average protein population should be comparable with the average value of K_j. This condition is used in order to have appreciable non-linearities in the transcriptional regulation [Eqs. (7)–(11)] in the hope that it will result in kinetic oscillations and/or chaos. (iii) The choice of the parameters should naturally allow us to show the type of the kinetics predicted by a model. Practically, this means that we should formally introduce a suitable governing parameter in order to show the kinetics as a function of this parameter.

Following guideline (iii), we use representation $w_i^\circ = pu_i^\circ$, where p ($0.2 \le p \le 1$) is the dimensionless governing parameter, and u_i° are the parameters selected [according to guideline (i)] at random in the range from 200 to 500 min^{-1}. The basal synthesis rates, w_i^b, are considered to be low and chosen at random between 1 and 2 min^{-1}. The protein association constants, K_j, are selected [according to guideline (ii)] at random in the ranges from 700 to 1400 for $m = 1$, from 200 to 500 for $m = 2$, and from 100 to 200 for $m = 4$.

For each run of MF calculations, we choose the parameters as described above, set $p = 1$, and define the initial conditions for Eqs. (5) and (6) as $N_i(0) = (w_i^b + w_i^\circ)/k_i$ and $n_i(0) = v_i N_i(0)/\kappa_i$ (these values represent the maximum possible steady-state mRNA and protein populations; the results of calculations are independent of the details of the choice of the initial conditions). Then, Eqs. (5) and (6) are integrated during sequential relatively long time intervals, $\Delta t = 1000$ min. The reactant populations obtained in the end of a time interval are used as the initial conditions for the integration of Eqs. (5) and (6) during a subsequent time interval. After each integration interval, p is decremented by $\Delta p = 0.01$. After reaching a minimum value of p (0.2), p is incremented after each integration interval by Δp up to reaching $p = 1$. The duration of time intervals was proved to be sufficient in order to reach steady states corresponding to chosen parameters. For each m, we have performed 100 runs of calculations with variation of p and different sets of kinetic parameters.

The MF kinetics under consideration can be illustrated by showing the mRNA or protein populations, obtained in the ends of integration intervals, as a function of p. In such graphs, the dependence of a population on p will represent a line without intersections (Fig.1a) provided that there is a unique and stable steady state. In the case of bistability, the dependence of a population on p will contain a loop corresponding to hysteresis (Fig.1b). Oscillations or chaos will be represented by a region (or regions) completely covered by data points (Fig.1c).

Referring to the Boolean networks and looking at Fig.1, one could expect to observe case (a) for $m = 1$, cases (a) and (b) for $m = 2$, and case (c) for $m = 4$. Our MF calculations for $n = 30$ (Fig.2a) show, however, that except a few percent of bistable kinetics observed for $m = 4$ we have case (a) irrespective of the value of m, i.e., the model does not predict oscillation and chaos. Our MC calculations, performed by using the standard Gillespie algorithm [11] (for the specifics related to gene transcription, see Ref. [12]), confirm this conclusion (see, e.g., Fig.3). The results of our calculations for $n = 100$ are nearly identical to those for $n = 30$ [cf., e.g., Fig. 2(a) and (b)].

As already noted in the introduction, CGNs may include ncRNAs. Due to their abilities to silence genes via pairing with various target mRNAs and subsequent preventing their translation or facilitating degradation of the mRNA-ncRNA complex (see, e.g., review [13] and analysis [14] of the effect on ncRNA on the populations of mRNAs in the absence of the protein-mediated feedbacks), ncRNAs can play a role of a global regulator. Complementing steps (1)–(4) of our present model by the formation and degradation of ncRNA and its association with mRNAs, we have found that, although the bistable kinetics are observed more often (especially for $m = 4$), oscillations and chaos are lacking.

The difference in the behaviour of the random Boolean networks and the corresponding MF or MC networks including mRNAs and proteins explicitly seems to be related to a prescribed discreteness of the changes of the states of the nodes in the former case. In reality, the

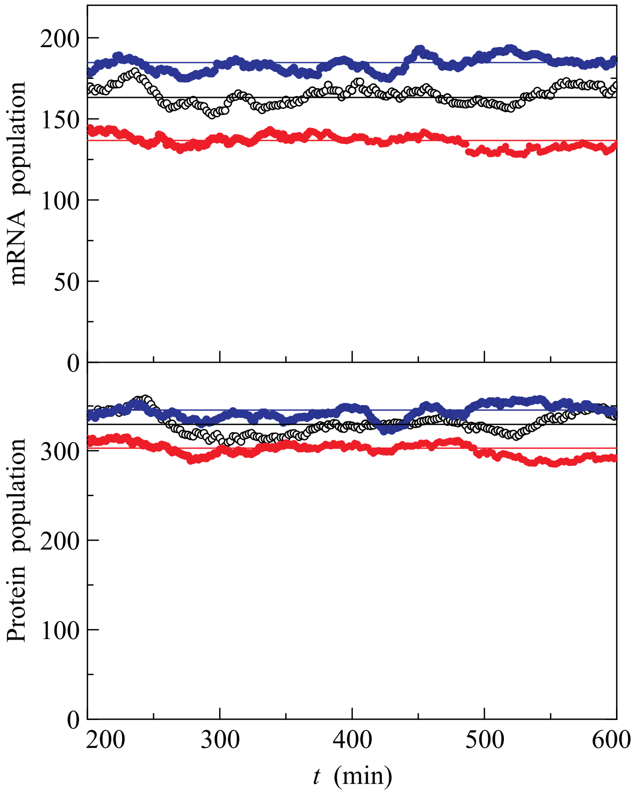

Fig.3. Average mRNA and protein populations as a function of time during three MC runs performed for $n = 30$, $m = 4$ and $p = 0.6$. The distance between data points is 2 min. The straight solid lines correspond to the MF equations

protein-mediated regulation is gradual [see, e.g., Eqs. (7)–(11)]. Discrete changes are related to bistability. Numerous models focused on small genetic networks indicate that in order to realize bistability, one should use more complex (non-random) regulation including, for example, positive regulation with two regulatory sites for protein association. Our analysis of complex networks with $m = 4$ and this type of regulation [this case can be described, e.g., by replacing expression (11) by $w_i = w_i^b + w_i^\circ F_1(n_1)^2 F_2(n_2)^2$] indicate that one can often observe bistability and sometimes oscillations but not chaos.

Two additional indirect arguments in favour of our conclusions come from the analysis of the conditions of the appearance of oscillations in small mRNA-protein networks and from experimental studies of global protein expression during the growth of cells. (i) In small mRNA-protein networks, kinetic oscillations are known to occur under rather special conditions including, e.g., the delay steps [15]. Realization of such conditions in the case when the kinetic parameters are chosen at random is unlikely. (ii) Oscillatory or chaotic kinetics are sensitive to kinetic parameters. If such kinetics of gene expression were typical, one could expect complex dependence of the protein concentrations on time during the growth of cells. The experiments indicate, however, that the concentrations of the majority of proteins are independent of the phases of the cell cycle [16].

In summary, our MF and MC calculations clearly indicate that oscillations and chaos are not inherent to CGNs with distributed regulation. Such networks are common in cells, and accordingly our conclusion is conceptually important from various perspectives. For example, the choice of the cell fate during differentiation is often believed to be controlled by kinetic chaos in gene expression. Our analysis shows that the conditions of realization of this scenario is much narrower than one could expect.

1. Y. Bar-Yam, D. Harmon, and B. de Bivort, Science **323** 1016 (2009).
2. D. Kulasiri, L. K. Nguyen, S. Samarasinghe, and Z. Xie, Curr. Bioinf. **3**, 197 (2008).
3. M. C. Lagomarsino, B. Bassetti, G. Castellani, and D. Remondini, Molec. Biosyst. **5**, 335 (2009).
4. D. F. T. Veiga, B. Dutta, and G. Balazsi, Molec. Biosyst. **6**, 469 (2010).
5. M. Kaern, T. C. Elston, W. J. Blake, and J. J. Collins, Nature Rev. Genet. **6**, 451 (2005); J. Paulsson, Phys. Life Rev. **2**, 157 (2005); J. Rausenberger, C. Fleck, J. Timmer, and M. Kollmann, Progr. Biophys. Molec. Biol. **100**, 57 (2009); A. S. Ribeiro, Math. Biosci. **223**, 1 (2010).
6. M. Chaves, E. D. Sontag, and R. Albert, IEE Proc. Syst. Biol. **153**, 154 (2006); M. Andrecut and S. A. Kauffman, Phys. Lett. A **372**, 4757 (2008); A. Esmaeili and C. Jacob, Biosyst. **98**, 127 (2009); D.M. Wittmann, C. Marr, and F. J. Theis, J. Theor. Biol. **266**, 436 (2010).
7. B. Derrida and Y. Pomeau, Europhys. Lett. **1**, 45 (1986).
8. X. Zhu, M. Gerstein, and M. Snyder, Genes. Dev. **21**, 1010 (2007).
9. D. Michel, Progr. Biophys. Molec. Biol. **102**, 16 (2010).
10. E. Yang et al., Genome Res. **13**, 1863 (2003); A. Beyer, J. Hollunder, H.-P. Nasheuer, and T. Wilhelm, Molec. Cell. Proteom. **3**, 1083 (2004); R. de Sousa Abreu, L. O. Penalva, E. M. Marcotte, and C. Vogel, Mol. Biosyst. **5**, 1512 (2009).
11. D. T. Gillespie, J. Phys. Chem. **81**, 2340 (1977).
12. V. P. Zhdanov, Chem. Phys. Lett. **424**, 394 (2006).
13. D. P. Bartel, Cell **136**, 215 (2009).
14. V. P. Zhdanov, Molec. Biosyst. **5**, 638 (2009).
15. G. Tiana, S. Krishna, S. Pigolotti et al., Phys. Biol. **4**, R1 (2007); V. P. Zhdanov, Biophys. Rev. Lett. **5**, 89 (2010).
16. E. Lundberg et al., J. Proteomics **71**, 448 (2008).