Physica D 239 (2010) 967–971

Contents lists available at ScienceDirect

Physica D

journal homepage: www.elsevier.com/locate/physd

Proteomic signals in modular transcriptional cascades: A discrete time and cellular automaton approach

A.S. Carstea *, A.T. Grecu, D. Grecu

Department of Theoretical Physics, Institute of Physics and Nuclear Engineering, 407 Atomistilor, Magurele, 077125 Bucharest, Romania

ARTICLE INFO

Article history:
Received 6 January 2010
Accepted 1 February 2010
Available online 7 February 2010
Communicated by A.C. Newell

Keywords:
Gene networks
Discrete equations
Kink wave
Ultradiscretisation
Cellular automaton

ABSTRACT

A transcriptional modular network cascade with general activation/repression coupling is analysed. It is shown that in certain conditions (the Hill coefficient is unity) it is equivalent to a cascade made of the same type of gene, and moreover the dynamics is linearisable. The protein production is shown to be a travelling kink wave on the cascade showing a successive gene expression along it. Also a cellular automaton description for both open and closed (plasmid) cascades is provided.

1. Introduction

Genes are the basic building blocks for any living organism. The main role of gene interactions is regulating each other's activity, and this leads to the desired response. Typically a gene is subject to the regulatory effect of other genes. The main task of theoretical studies on genetic circuits thus far has been on the combinatorial control of transcriptional initiation, which to a large extent determines the connectivity of the network. It is thus extremely important to study, and understand, the dynamics of the gene regulatory networks in general. The activity of a gene is regulated by other genes through the production of transcription factor (TF) proteins. Physically, this is accomplished through the interaction of these transcription factor proteins with the RNA polymerase complex in the regulatory region of the gene. The gene-code segment of the DNA chain is read by the RNA polymerase, which binds to DNA and moves along activated by transcription factor proteins and gives rise to the RNA messenger. This transports the respective code to the ribosomal machine, and proteins are produced according to codon sequences. Among these proteins we also have, of course, constituents of the transcription factor. Cellular regulation is realised by a very complex network of interactions and processes. The links in this huge network involve tens of thousands of biochemical reactions. It is, therefore, very important to find procedures for simplifying the description and facilitating quantitative analysis. One is thus led to the concept of a motif which represents some basic subnetwork that is classifiable on the basis of function, architecture dynamics, etc. At a bigger scale, the motifs can be seen as being organised in modules having their specific role. The main task of synthetic molecular biology is to create small synthetic genetic networks, inserting them in different living cells in order to regulate and control their behaviour. Recent experimental regulatory motifs include a toggle switch, a repressor oscillator, logic gates, and pulse generators [1]. Engineered gene networks are supposed to have tremendous applications in biotechnology and medicine. Such engineered biological devices will engage in simple computations and cell–cell communications to diagnose diseases. The main challenge in synthetic biology is creating and tuning gene networks to desired functions. Tackling the problem on simpler systems such bacterial cells will give many useful insights, while delivery of such devices in higher organisms involves the interesting problem of cell-specific recognition. The above-mentioned regulatory motifs involve small numbers of interacting genes, and the theoretical study is focused on stochastic/differential kinetic rate equations for mRNA and protein production. The complexity comes from feedback activator–repressor loops, nonlinear response functions, and the presence of noise or other stochastic effects. These properties are all ultimately responsible for the interesting *physics* of the motifs. Fixed points, oscillations and stochastically controlled switching between different states are the main theoretical results obtained so far (using essentially linear analysis). For example, switches (theoretically modelled by stochastic differential equations with noise-controlled bistability) proved to be instrumental in understanding the dynamics of Phage lambda cross-repressive feedback loops, quorum sensing, kinase pathways in Xenopus oocytes or synthetic switches. Oscillators have

* Corresponding author.
E-mail address: carstea@gmail.com (A.S. Carstea).

0167-2789/$ – see front matter © 2010 Elsevier B.V. All rights reserved.
doi:10.1016/j.physd.2010.02.002

applications in cell cycles, glycolysis, cytosolic Ca2+, and synthetic oscillators [2]. In this paper we are going to extend the systems described above to ones possessing a large number of genes, namely ensembles of regulatory motifs and modules in interaction. We are going to discuss the simplest case of a modular cascade of activators/repressors and show both the dynamics and the modular nature of it. This is quite important since motifs do not work in isolation; complex processes such as growth, cell cycle, developmental programs, motility and pathogenic processes are controlled by motifs connected in elaborate hierarchical and feedback structures. So far the interaction between simple regulatory motifs has been little investigated in the literature from a theoretical point of view. Some recent work includes the effect of intercell signalling on a population of nonidentical and noisy repressilators coupled by quorum sensing [3]. Another result is the experimental control of gene expression using pulse-generating networks [4].

The idea of proteomic signal *propagation* as a discrete nonlinear wave along a cascade (open or closed) has been given in [5], where we studied an alternating activator–repressor network of genes, every one expressed by two promoters. Here we proved that an initial protein distribution strongly localised on a certain number of genes can propagate along the cascade in a *solitonic* way, i.e., successive gene expression. Cascades are found in many biological systems. For instance, *Escherichia Coli* and *Saccharomices Cerevisiae* contain transcriptional cascades with many stages [6]. Protein kinase pathways exhibit cascade dynamics which regulate activities in cells [7]. They are the basic networks which support programs of successive gene expressions as observed in flagella formations of *E. Coli*, developmental programs in *Drosophila* [8], etc. Also we studied simpler cascades [9] and computed travelling wave solutions (as kinks) taking into account the time delay between the transcription and translation.

In this paper we propose a model of multigene cascade with simple activator or repressor regulatory loops. We are going to generalise the approach discussed in [5] and [9] to a cascade of *groups* of different genes (modular multi-gene cascade) and show that every such module behaves as a *single* gene characterised by some effective parameters but producing proteins at a *slower rate*. Also we show that the whole dynamics is integrable in the sense that the equations are linearisable by means of a Cole–Hopf-type transform, and support travelling kink solutions. Throughout the paper we use (as in [9]) the discrete time approach, which is more amenable that other approaches to analytical treatment of time-delay effects (often encountered in modelling transcription and translation). Moreover, we shall discuss the so-called ultradiscrete approach [10], which allows a simpler description of any nonlinear discrete dynamical system by the so-called *tropical* or ultradiscrete limiting procedure. This one rigorously and algorithmically turns every discrete equation (with positive definite variables and parameters) into a generalised cellular automaton by replacing in a specific way all the nonlinearities with the *simplest* nonlinear function (namely the absolute value or $f(x) = |x|$), and thus the dynamics is given by sequences of integer numbers. We shall also present new solutions using this procedure.

2. The model

The dynamics of gene regulatory networks can be modeled by chemical rate equations corresponding to each gene. These equations represent, of course in a very simplified way, the transcription and translation mechanisms. For the transcription mechanism one must first describe the probability of the RNAp molecule (together with transcription factors) binding to the gene promoter region (localised at the beginning of the gene). This probability is usually called 'promoter activity function'. It represents in a certain way the efficiency of transcription due to the interaction between RNAp and the transcription factors involved (activators and/or repressors) [11].

In a thermodynamical description [12], this probability is given by the function

$$g_a(p) = \frac{1 + \omega p/k_a}{1 + p/k_a}$$

for activation, and

$$g_r(p) = \frac{1}{1 + p/k_r} + \Lambda$$

when a gene is repressed. The Boltzmann weight ω is given by $\omega = e^{-E/kT}$ corresponding to the activator and RNAp interaction; $k_{a,r}$ are the dissociation constants between the activator/repressor protein and the respective operator sequence in the regulatory region. Finally, Λ is the effect of the promoter leakage on the repression. The values of k_a and k_r are tunable and are in the interval (1, 1000) nM [13]. Also, $\omega = 100$ [14], and in the case of activation for small values of k_a the second term in the numerator dominates. Also, the promoter leakage has very small values in the range $(10^{-3}, 10^{-1})$ [13], and is neglected.

Sometimes, in order to take into account possible allosteric polymerisations of transcription factors, an effective simplified form is used:

$$g_{a,r}(p) = \frac{\alpha + \beta p^q}{1 + p^q} \tag{2.1}$$

where q is the Hill coefficient, taking positive values which are small (throughout the paper we will take $q = 1$) and $\alpha \geq 0$, $\beta \geq 0$. The case of $\alpha = 0$ corresponds precisely to the limit of $g_a(p)$ when ω is big. Now we can write down the transcription process equation for one gene activated/repressed by one transcription factor (which is the simplest genetic circuit):

$$\frac{d}{dt} m(t) = g_{a,r}(p(t)) - \lambda_m m(t) \tag{2.2}$$

where $m(t)$ is the time-dependent concentration of transcription-emerging mRNA and λ_m its degradation constant. The next step in the model is to consider the protein production from this mRNA, which enters into the ribosomal machine (the translation process). A linear form is usually adopted for the ribosomal efficiency as a function of $m(t)$. As a consequence, the translation process is described by the following differential-delay equation:

$$\frac{d}{dt} p(t) = \nu m(t - \tau) - \lambda_p p(t) \tag{2.3}$$

where τ is the delay between transcription and translation and λ_p is the protein degradation rate. In the formulation of the model we do not make any distinction between protein and transcription factor. This is motivated by the fact that, in many cases, as for instance in bacteria, the two coincide. The value of ν corresponds to the protein synthesis at full rate of activation and has a large span of values (1–100 nM/min) [11]. The two differential equations (2.2) and (2.3) constitute the kinetic rate equations which govern the dynamics of a single gene in interaction with the protein it produces (which acts as a transcription factor).

As we pointed out in the introduction, genes do not work in isolation. The model we shall consider here consists in a cascade of genes, where the protein produced by any gene is a transcription factor for the next gene in the network, acting on it in an activating or repressing way. This is practically the simplest cascade. In our model, genes are *not identical* as in the previous paper [9], but they are grouped together, every group having the same number of different genes. Accordingly this model can be viewed as a cascade of identical genetic groups (modules), or a multigene cascade. Related to this model we are going to study the following facts:

- This multigene cascade is equivalent to a single gene cascade characterised by some effective parameters. This is valid only