Romanian Reports in Physics, Vol. 62, No. 1, P. 169–178, 2010

TIME DELAY EFFECTS ON OSCILLATORY SIGNALS IN GENETIC REPRESSILATORS WITH TWO PROMOTERS

A. S. CARSTEA, A. T. GRECU

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

(Received June 9, 2009)

Abstract. We investigate the dynamics of 2 promoters controlled genes in a repressing periodic (plasmid) cascade. It is shown that for plasmids having these type of genes the protein production is propagating in an oscillatory way depending on the delay time involved between transcription and translation. Biological implications are discussed.

1. INTRODUCTION

Biological systems are organised into big blocks of networks that control the precise regulation of biochemical reactions within the cell and its interaction with the outside medium. But the complexity of these networks and nonlinear intracellular and intercellular interactions makes difficult the examination of dynamical processes. This fact has motivated researchers to engineer synthetic gene networks to see the nonlinear effects in isolation. Thus, starting with the cascade repressilator, many other circuits have been studied and constructed experimentally, like oscillators, toggle switches, negative autoregulation circuits, logic gates and pulse generating nets [4, 15]. Transcription cascades are defined by a set of transcription factors that regulate each other sequentially. The first step transcription factor activates or represses the second step one, which in turn, activates or represses the third and so on. Cascades are oftenly appearing in nature. For example in the case of *E. Coli* and *Saccharomices cerevisiae*, regulatory networks contain transcriptional cascades with two or more stages [7]. Cascades help programs of succesive gene expression as observed in the formation of flagella in *E. Coli*, sporulation in budding yeast, or regulatory pathways in bacterial cell cycles [8]. In multicellular organisms like *Drosophila* and sea urichin some developmental programs require temporal ordering of events controlled by cascade processes [12].

However most of the studies so far have been focused on simple bacterial genes whose expression is controlled by only one promoter. In this paper we intend to make a small step further and see the dynamics of cascades made out of genes controlled by two promoters. More precisely, we intend to study a repressilator, i.e. a cascade of repressors having two promoters and taking into account the time delay which appear between transcription and translation (because the genes are more complex, these type of temporal delays should be taken into account). In order to see more effects we consider that each product of the gene is also a selfactivator for it (otherwise the results would have been just trivial generalisations of the ones existing already in literature [5]).

Genes with two promoters are quite often encountered in biological world in specially in eukariotic cells. However even in bacterial cells, like *Pseudomonas Aeruginosa* [9] and *Ersinia entherocolitica* [13], the rpoE operons and araP are controlled by two ar more promoters. However here we are not intending to study a specific cascade but only to give an exactly solvable model which shows some particularities which can be useful for future engineered circuits. What we are interested in is the time dependence of protein production. The cascade can be closed (plasmid-type structure) which is more amenable to be constructed synthetically. Writting the rate equatuions for protein production on every gene and taking into account that binding probability of RNA polymerase (RNAp) is controlled by two promoters and also the delay, we end up with a partial differential delay equation. Remarkably this equation can be solved exactly in terms of elliptic functions.

2. TRANSCRIPTION OF GENES WITH TWO PROMOTERS

To start our discussion let us remind how transcription and translation processes work in the simplest case of a gene controlled by a single promoter. The transcription and translation involve the synthesis of mRNA from a single DNA-encoded gene template (by means of sticking RNAp to the promoter) then the synthesis of protein from mRNA template and then decay of mRNA and protein molecules. This is, of course, a strongly simplified image of the gene expression. In addition, the DNA flips between the transcriptionally active and inactive states at rates that depend on the concentrations of transcription factors which bind to the promoter region of the gene. The regulatory aspect of gene expression comes from the fact that all the transcription factors are proteins expressed by some genes which in turn activate or repress other/same genes.

The simplest mathematical model of the trasncription-translation process for a single gene is based on the mean-field equations:

$$\frac{\mathrm{d}}{\mathrm{d}t} m = \alpha - \lambda_m m, \tag{1}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} p = \nu m - \mu - \lambda_p p. \tag{2}$$

Here m and p are the concentrations of mRNA and produced protein respectively. Also α is the mRNA synthesis rate (typical 1/min), λ_m^{-1} is the mRNA life time (typical 5min), ν is the protein synthesis (typical 10nM/min) and λ_p^{-1} is the protein life time (typical 30 min)[14]. The term μ in the equation (2) is related to the fact that a certain amount of mRNA entering into translation process is not coded [11]. However there are many things which are not contained in this model. For instance time-delay from transcription initiation to mRNA completion, time-delay related to the transport to ribosomes, and time-delay in protein synthesis. Also fluctuations over scale of cell doubling time are important. In any case the model is valid for big time scales ("big" means beyond cell-cycle doubling time). In this big scale the mRNA synthesis can be considered fast so the equation (1) in the system is stationary i.e. dm/dt = 0. Accordingly, the gene expression is given by:

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\nu\alpha}{\lambda_m} - \mu - \lambda_p p. \tag{3}$$

In bacteria the protein synthesis rate α is controlled by the amount of time RNAp spends bound to the promoter so is no longer a constant. It can be measured by the binding probability of the RNAp to the promoter region, binding which is amplified or reduced by the influence of other proteins called transcription factors (TF). In the case of activation [1] this can be mathematically modelled as:

$$\alpha_A(\rho, q) = \frac{Z_{on}}{Z_{on} + Z_{off}} = \frac{\rho(1+\omega q)}{1+\rho+q+\omega\rho q},$$

where Z_{on}/Z_{off} are partition functions over all states of TF-binding for the promoter bound/not bound by the RNAp. Here $\rho = [RNAp]/K_\rho$, $q = [TF]/K$, $\omega = \exp\left(-G_{TF-RNAp}/k_B T\right)$, where in the square brakets we put the concentrations of RNAp and TF, K_ρ, K are the dissociation constants between the RNAp, TF and the respective operator sequence in the regulatory region and $\Delta G_{TF-RNAp}$ is the free energy interaction between RNAp and TF. Because the concentration of RNAp is small the promoter activity is given by a truncated Taylor expansion in the first order (and rescaling with $\omega\rho$):

$$\alpha_A(\rho, q) = \rho \frac{1+\omega q}{1+q} + \mathcal{O}(\rho^2) \propto \frac{\omega^{-1}+q}{1+q}. \tag{4}$$

In bacteria, oftenly the produced proteins are transcription factors themselves and they may act sinergically. Taking into account the delay the transcription translation process is modelled more generally as:

$$\frac{dm}{dt} = \alpha_{A,R}(\rho, q_1, q_2, \ldots, q_N) - \lambda_m m, \tag{5}$$

$$\frac{dp}{dt} = \nu m(t-\tau) - \mu - \lambda_p p, \tag{6}$$

where τ is the time delay between transcription and translation, $q_1,\ldots q_N$ are transcription factors and p is the protein produced by the gene (which can be one in the set $q_1,\ldots q_N$). We have to mention that the formula (4) works only when transcription factor acts as a monomer (i.e one operator site in the regulatory region). In the case of multiple operator sites (and allosteric polymerisations of transcription factors before entering into regulatory regions) an effective formula is used [3]:

$$\alpha(\rho, q) = \frac{a + bq^n}{1+q^n},$$

where $a < b$ and n is the Hill coefficient which is a positive number.

All we have discussed above for the activation is valid on the repressing case but the promoter activity becomes a decreasing function and the formula (4) turns into:

$$\alpha_R = \frac{1+\omega^{-1}q}{1+q}. \tag{7}$$

Also the effective formula has the same form provided $a > b$.

There are many cases where many transcription factors act on the promoter to help or break the transcription. In this case one has to take into account the partition functions for all the possibilities of $RNAp - TF_1 - TF_2 - \ldots - TF_m$ couplings in order to get the mathematical expression for promoter activity function [2].

In the case of two promoters, which is the case we are interested in, the promoter activity is given by the equilibrium probability that the RNAp binds to at least one of the promoters. Imposing that there is no interaction between the promoters (which means that transcription factors do not simultaneously interact with both polymerases in the unlikely case that both promoters are occupied) the transcription rate is given by [2]:

$$\alpha(\rho, p) = \frac{Z_{on}^{(1)} Z_{off}^{(2)} + Z_{off}^{(1)} Z_{on}^{(2)} + Z_{on}^{(1)} Z_{on}^{(2)}}{Z_{off}^{(1)} Z_{off}^{(2)} + Z_{on}^{(1)} Z_{off}^{(2)} + Z_{off}^{(1)} Z_{on}^{(2)} + Z_{on}^{(1)} Z_{on}^{(2)}}.$$

Here $Z_{on}^{(i)}$ and $Z_{off}^{(i)}$ are partition functions over all states of TF binding when promoter i is bound and not bound by the RNAp respectively. Again in the case of activation they can be written as:

$$Z_{on}^{(i)} = \rho + \rho\omega_i q_i,$$

$$Z_{off}^{(i)} = 1 + q_i,$$

where $g_i = [TF]/K_i$ is the TF affinity coresponding to the regulatory region of the i-promoter, and $\omega_i = \exp\left(-G_{TF-RNAp_i}/k_B T\right)$ is related to the free energy of interaction between TF and RNAp in the regulatory region of the i-th promoter. One can see immediately that in the limit of small ρ (RNAp affinity) then in the leading order we have:

$$\alpha_A(\rho, p) = \rho\frac{1+\omega_1 q_1}{1+q_1} + \rho\frac{1+\omega_2 q_2}{1+q_2} + \mathcal{O}\left(\rho^2\right), \tag{8}$$

so in fact we have the sum of the two contributions of the *same* protein which binds on the two promoters. Had we had many proteins we would have taken into account all possibilities of coupling among them and promoters.

3. THE MODEL

Now we can formulate our model. Consider a gene network cascade where each gene is controlled by two promoters of σ^{54}-type and the protein is a self activator and repressor for the next gene in the cascade. The genes are indexed with integers *n*. Let us motivate why we have chosen promoters of type σ^{54}. This type of sigma factor is used at promoters that have many activation sites (some of them being quite far allowing DNA looping for activation). They act somehow as eukariotic enhancers (i.e. sets of contiguous regulatory binding sites) so the transcription mechanism here is believed to be a hibrid one between bacterial and eukariotic. Moreover, activators interact with σ^{54}-polymerases in a more specialised way. Namely there is a single site on polymerase that must be contacted in order to trigger the conformational change, which lead to open complex [10]. Accordingly, in our model only *one* type of protein touches every promoter. This

simplifies considerably the mathematical expression of the promoter activity α which becomes precisely the sum of the contributions of every protein. The promoter activity functions for the two promoters are considered for simplicity to have the same parameters ω. The cascade's underlying network mechanism is the following: the protein p_n produced by the the gene (n) activates the gene (n) and represses the gene ($n + 1$). This scenario works for every gene (i.e. every n).

Now from the one gene model (5), (6) we have the equation of the network for the gene labelled by n;

$$\frac{\mathrm{d}m_n}{\mathrm{d}t} = \alpha_{A,R}\left(\rho, q_n, q_{n+1}\right) - \lambda_n m_n, \tag{9}$$

$$\frac{\mathrm{d}p_n}{\mathrm{d}t} = \nu m_n\left(t-\tau\right) - \mu - \delta_n p_n. \tag{10}$$

Now in order to see the effect of the two proteins p_n and p_{n+1} on the expression of α we have to take into account all the possible interactions. But, as we pointed out above the σ^{54} promoters allows only one protein to be attached on the regulatory region. Accordingly we shall have exactly the situation described by the equation and the expresion of α_{AR} in (6) will be a sum of the form

$$\alpha_{AR}\left(\rho, q_n, q_{n+1}\right) \propto \frac{\omega_n^{-1} + q_n}{1+q_n} + \frac{1+\omega_n^{-1}q_{n+1}}{1+q_{n+1}}.$$

Assuming stationarity of the mRNA production in equation (6) we end up with the following partial differential-delay equation (after we normalize $p_n -> p_n / K_n \equiv q_n$)

$$\frac{\mathrm{d}q_n\left(t+\tau\right)}{\mathrm{d}t} = \rho\frac{\omega_n^{-1} + q_n\left(t\right)}{1+q_n\left(t\right)q_{n+1}\left(t\right)} + \rho\frac{1+\omega^{-1}q_{n+1}\left(t\right)}{1+q_{n+1}\left(t\right)} - \mu - \delta_n q_n\left(t\right). \tag{11}$$

In the next section we are going to discuss this equation and biological implications.

4. RESULTS

First of all we put the equation in a more tractable form by the following substitution:

$$u_n = \frac{1}{1+q_n}.$$

Since q_n is always positive this will impose restrictions on the values of u_n to be in the interval (0,1). In the new variable our equation will be:

$$\frac{\dot{u}_n}{u_n^2} = (1-1/\omega)\left[u_n(t-\tau) - u_{n+1}(t-\tau)\right] - (1-1/\omega) + \mu + \delta\left(\frac{1}{u_n} - 1\right).$$

Since μ and $1/\omega$ are small we can write the equation (after scaling time) as a *perturbed* differential-discrete-delay one in the form:

$$\dot{u}_n = u_n^2\left[u_n(t-\tau) - u_{n+1}(t-\tau)\right] + \text{perturbation}. \tag{12}$$

For simplicity we analyse the unperturbed equation as an *evolution* one for the protein production $q(n, t)$ of the whole periodic gene network. Our discuss will be focused on how an initial protein distribution localised on some genes evolves in time. In order to see this we take the travelling-wave ansatze $q(n, t) = q(n–vt)$ where v is the speed of travelling wave. This evolving distribution will be called proteomic signal or proteomic wave. Because we have a plasmid shape with an arbitrary number of genes the solution will be a periodic nonlinear wave

Calling $\xi = n - vt$ we will have:

$$-vu(\xi)' = u(\xi)^2\left[u(\xi + v\tau) - u(\xi + 1 + v\tau)\right].$$

This equation admits solution for $v = -1/2\tau$. Indeed we have:

$$\frac{1}{2\tau}u(\xi)' = u(\xi)^2\left[u(\xi - 1/2) - u(\xi + 1/2)\right].$$

This is nothing but semidiscrete Korteweg de Vries equation [6] in the travelling wave ansatze. Its solution is well known [17] and is given in terms of Jacobi-theta 4 function:

$$u(\xi) = u\left(n + \frac{t}{2\tau}\right) = \frac{1}{\sqrt{2\tau}} \frac{\Theta_4\left(\frac{p^\pi}{N}\left(n - 1 + \frac{t}{2\tau}\right)\middle|B\right)\Theta_4\left(\frac{p^\pi}{N}\left(n + 1 + \frac{t}{2\tau}\right)\middle|B\right)}{\Theta_4^2\left(\frac{p^\pi}{N}\left(n + \frac{t}{2\tau}\right)\middle|B\right)},$$

where B is the matrix period, N is the number of genes (which is periodicity) and p is any integer number. Here we have a constraint on it namely, $2N\tau\Theta_2\left(\frac{\pi}{N}\middle|B\right) = \Theta_1'(0|B)$ which gives a transcendental equation for B. This constraint is nothing but the dispersion relation of the periodic KdV soliton which here must be related to the time delay. In order to have a real positive solution B should be positive and smaller than 1. For negative values the solution is singular (which means u may go to infinity and the protein concentration goes to zero).

Also, because the dissociation constant in the expression of q_n is big and the protein concentration is small we can consider the linear limit namely the following equation (which is nothing but the expansion up to order two)

$$\frac{\mathrm{d}}{\mathrm{d}t}q_n\left(t+\tau\right)=\left(\omega^{-1}+1-\mu\right)+q_n\left(1-\omega^{-1}-\delta\right)-q_{n+1}\left(1-\omega^{-1}\right). \tag{13}$$

We can search for periodic solutions of the form:

$$q_n\left(t\right)=A+\cos\left(kn+\Omega t\right)$$

such that $q_n\left(t\right)>0$. We find the following transcendental dispresion relation:

$$i\Omega \mathrm{e}^{i\Omega\tau}=\left(1+\omega^{-1}-\delta\right)-\left(1-\omega^{-1}\right)\mathrm{e}^{ik},$$

$$A=\frac{1+\omega^{-1}-1}{\delta}.$$

Condition for positivity reads as $1+\omega^{-1}-\mu>2\delta$ and is satisfied for a large class of proteins since they are stable and δ is very small. Accordingly in the linear limit we have an oscillating wave which propagates along the plasmid.

5. DISCUSSIONS AND CONCLUSION

The crucial thing here is to check that indeed the solution is in the interval (0, 1). Fortunately, the equation is invariant to the following two scalings $t\to b^2t, u\left(n,t\right)\to bu\left(\xi\right)$ so we can put in front of u any number b (such that $bu < 1$) provided we change time scale with b^2. This time scale changing means modification of the speed. Accordingly the speed of propagation is proportional with the amplitude (a wellknown phenomenon in soliton theory) and this amplitude is given by the initial condition. Anyway in our case since the velocity is fixed and the independent variable is fixed the only chance is that the factor $1/\sqrt{2\tau}$ to keep the maximum value smaller than 1. Accordingly we must have a compatibility with the value of the corresponding period matrix B (which also depends on τ). So the existence of such solutions is depending crucially on the value of time delay.

The solution is indeed a periodic travelling wave which can be interpreted in two ways. First it is a propagating signal which activates and represses in cascade the genes along the plasmid. The signal is faster or slower depending on the initial condition (namely the activity of the triggering gene) and also the time delay τ. So indeed a bigger time delay will slow down the sequential genetic activity of the plasmid. Second, this periodic travelling wave can be interpreted as describing an oscillatory behaviour of every gene in the plasmid with variable frequency

depending again on the signal speed. In both cases the delay plays an extremely important role controlling the full activation-repression activity of the palsmid. Of course our solution is in fact a solution of the nonperturbed equation. Adding perturbation will modify the stability of the signal and indeed for some specific values of the constants the signal propagation can die out or even cannot be launched at all. The stability of the travelling signal in the nonperturbed case is not at all automatic as well even though it is a solution of a completely integrable equation (semidiscrete KdV). However the rigurous study of it is beyond the scope of this paper. From this considerations we can made some assumptions related to the biological relevance. It is known that gene cascades offer a temporal programme of gene expression that can be used for scheduling protein synthesis activities sequentially [16]. In our model we ecounter precise this fact. The genes are turned off succesively as the nonlinear wave propagates. Also it has been suggested that cascades exhibit a certain capability, the so called "low pass filtering" which means ignoring rapid variations of protein concentrations or environmental conditions and respond only to longer-lasting ones. This may very well happen in our model since the equation is a prototype of integrable system and the solitary waves are supossed to be stable (although to our knowledge there is no proof so far in the periodic case). As a conclusion we can say that in the plasmids made out of genes with two promoters some nice dynamics may appear both as a result of nonlinearities and time delay. In this simplified model we computed exactly a periodic travelling wave solution for the proteomic signal and showed that the speed depends both on the amplitude and time delay.

Acknowledgement. This work was supported by the PN-2, project IDEI, Nr. 325, CNCSIS, Romanian Ministery of Education.

REFERENCES

1. L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, R. Phillips, *Transcriptional by the numbers: Models*, Curr. Opp. Genetics, Develop., **15**, 116–124 (2005).
2. N. E. Buchler, U. Gerland, T. Hwa, *On schemes of combinatorial transcriptional log*ic, Proc. Natl. Acad. Sci. USA, **100**, 5136–5141 (2003).
3. N. E. Buchler, U. Gerland, T. Hwa, *Nonlinear protein degradation and function of genetic circuits*, Proc. Natl. Acad. Sci. USA, **102**, 9559–9564 (2005).
4. T. S. Gardner, R. C. Cantor, J.J. Collins, *Construction of a genetic toggle switch in Escherichia Coli*, Nature, **403**, 339–342 (2000);
   M. B. Elowitz, S. Leibler, *A synthetic oscillatory network of transcriptional regulators*, Nature, **403**, 335–338 (2000);
   S. Kalir, U. Alon, *Using a quantitative blueprint to reprogram the dynamics of the agella gene network*, Cell, **117**, 713–720 (2004).
5. B. Grammaticos, A. S. Carstea, A. Ramani, *On the dynamics of a gene regulatory network*, J. Phys. A: Math. Gen., **39**, 2965–2971 (2006);
   A. S. Carstea, *Proteomic waves in networks of transcriptional regulators*, to appear in Math. Comp. oin Simulation.

6. R. Hirota, *Proceedings of Symmetries and Integrability of Difference Equations*, SIDE III CRM, 2000.
7. S. Hooshangi, S. Thiberge, R. Weiss, *Ultrasensitivity and noise propagation in a synthetic transcriptional cascade*, Proc. Natl. Acad. Sci.USA, **102**, 3581–3586 (2005).
8. S. Kalir et al., *Ordering genes in a agella pathway by analysis of expression kinetics from living bacteria*, Science, **292**, 2080–2083 (2001);
S. Chu et al., *The transcriptional program of sporulation in budding yeast*, Science, **282**, 699–705 (1998);
H. H. McAdams, L. Shapiro, *A bacterial cell-cycle regulatory network operating in time and space*, Science, **301**, 1874–1877 (2003).
9. M. E. Maxson, A. J. Darwin, *Multiple promoters control expression of the Yersinia enterocolitica phage-shock-protein A (pspA) operon*, Microbiol., **152**, 1001–1010 (2006); Cells, Proc. Natl. Acad. Sci. USA, **103**, 7304–7309 (2006).
10. M. Ptashne, A. Gann, *Genes and Signals*, Cold Spring Harbor Laboratory Press, 2001; M. Ptashne, *Genetic Switch*; *Phage Lambda Revisited*, Cold Spring Harbor Laboratory Press, 2004;
R. Lutz, H. Bujard, *Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements*, Nucleic Acids Res., **25**, 1203–1210 (1997).
11. C. Reiss, *Nonlinear Excitations in Biomolecules*, Eds. M. Peyrard, Springer, 1994.
12. N. Rosenfeld, U. Alon, *Response Delays and the Structure of Transcription Networks*, J. Molec. Biol, **329**, 645–654 (2003);
S. Shen-Orr, R. Milo, S. Mangan, U. Alon, *Network motifs in the transcriptional regulation network of Escherichia coli*, Nat. Genet., **31**, 64–68 (2002).
13. D. G. Storey et. al., *Multiple promoters control the regulation of the Pseudomonas aeruginosa regA gene*, Molec. Microbiol., **4**, 499–503 (1990).
14. S. Wickner, M. R. Maurizi, S. Gottesman, *Posttranslational quality control: folding, refolding, and degrading proteins*, Science, **286**, 1888–1893 (1999);
J. A. Bernstein, et al., *Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color uorescent DNA microarrays*, Proc. Natl. Acad. Sci. USA, **99**, 9697–9702 (2002).
15. D. M. Wolf, A. P. Arkin, *Motifs, modules and games in bacteria*, Curr. Op. Microbiology, **6**, 125–134 (2003).
16. A. Zaslaver, et al., *Just-in-time transcription program in metabolic pathways*, Nat. Genet., **36**, 486–491 (2004).
17. J. Zagrodzinski, *Quasiperiodic solutions of discretised Korteweg de Vries equation*, Phys. Rev., E, **51**, 2655–2572 (1995).