

STOCHASTIC ANALYSIS OF NONLINEAR DEGRADATION OF DIMERIC PROTEINS IN GENETIC CIRCUITS

A. S. CARSTEA

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

(Received January 15, 2009)

Abstract. Stochastic effects on cooperative stability for simple genetic circuits is discussed. The full chemical equations for protein production are analysed using linear noise approximation and fractional deviation from the steady state is determined. Also the stochastic effects is discussed on effective rate equations for protein production.

1. INTRODUCTION

Protein production represents the most important role of genes in living organisms. This process is usually called gene expression. It is a wellknown fact that controlled proteolysis in which the degradation of the protein is influenced by the presence of another protein can play a regulatory role in genetic circuits. Another effect of proteolysis which does not involve a regulatory control but has impact on the gene expression is the so called ‘cooperative stability’ [1]. It is relying on the following fact; many proteins perform their physiological functions as dimers or higher order oligomers which are more stable than their monomeric components. It has been observed in *Saccharomyces cerevisiae* where the dimerisation of α_1 and α_2 reduced the degradation rate as much as 15-fold [3]. Also in the SOS response of *E. Coli*, UmuC degradation is reduced by oligomerisation with $Umu D_2'$ [2].

In this paper we are going to study the influence of intrinsic noise (which is present in genetic systems due to the low number of molecules involved in biochemical processes) on the function of a simple genetic motif – a self-activator, where the transcription factor exerts its biological function as a homodimer.

In deterministic models the cooperative stability has been studied for bistable systems (having LOW and HIGH concentrations of proteins for the two distinct states of the cell) and it was shown that it can widen the ration between these two

concentrations when compared with the ration of the mRNA levels in the two states. Moreover for oscillatory circuits the cooperative stability broadened the basin of parameter space supporting the circuits function. However rate equations in deterministic models neglect that chemical reaction networks are composed of species that evolve on the discrete space – jumping from some number of molecules to another as each reaction occurs. The emergent effect which is a deviation from the deterministic case is called intrinsic noise in the system (since it appears from the intrinsic dynamics not from external sources). In cellular systems with small number of reactant molecules the effect can be strong. For instance a system that has several possible states may be induced to spontaneous transition between them leading to stochastic switching [4]. In this direction it has been confirmed experimentally that the precision of gene expression is limited by the fundamental statistical mechanical $1/N$ scaling of relative fluctuation amplitude and abundance N and that population variability in the expression of single genes depends on the DNA-encoded parameters in accordance with relationships predicted by stochastic theory [5]. In addition there are some experimental and theoretical approaches which are beginning to clarify how noisy gene regulatory signals propagate through signaling pathways, transcriptional cascades and feedback loops [6].

In any case despite these significant progresses there is still a need to develop methods for studying how different motifs/networks behave in the presence of stochasticity. The mathematical treatment is in general a computationally daunting task and from numerical point of view it is in most cases prohibitively time consuming. Here we discuss the method of linear noise approximation applied on a slightly more complicated autoactivator motif involving dimerisation which is responsible for the cooperative stability and the interesting dynamics of the system. Also the effective stability approximation will be implemented to see how it is influenced by intrinsic noise.

The paper is organised as follows. In Sec. 1 the stochasticity of the single gene due to intrinsic noise is discussed together with linear noise approximation. In Sec. 2 the selfactivator is modeled and the steady-state time correlation functions are computed and in the Sec. 3 the effective monomer production equation is discussed also using linear noise approximation.

2. STOCHASTICITY OF THE SINGLE GENE

We start our discussion by considering how internal (intrinsic) perturbations affect the expression of the single gene. The simple autoactivator involves the synthesis of mRNA from a single gene template with the aid of a protein called transcription factor, the synthesis of the transcription factor from mRNA templates (which in turn triggers again the gene expression) and the decay of mRNA and

protein (transcription factor) molecules. The role of transcription factor is in general switching between active/inactive state and that why is called sometimes activator or repressor. The deterministic modelling equations for these facts are the following:

$$\frac{dm}{dt} = \alpha g_A(p) - \lambda_m m, \quad (1)$$

$$\frac{dp}{dt} = \nu m - \lambda_p p, \quad (2)$$

where m and p are the concentration of mRNA and transcription factor (TF), α is transcription rate of the promoter at full activation, $g(p)$ is the promoter activity function which depends on the TF concentration, λ_m is the degradation rate of mRNA, ν is translation rate of mRNA and λ_p is the transcription factor's turnover. The steady states can be computed by assuming that equation are stationary and they depend on the form of $g(p)$. For simplified models the function $g(p)$ is assumed to be constant and the steady states have the following forms:

$$m_0 = \frac{\alpha g}{\lambda_m}, \quad p_0 = \frac{\nu \alpha g}{\lambda_m \lambda_p},$$

which by definition are related to the average steady state number of molecules divided by system size (the cellular volume V_{cell})

$$m_0 = \frac{\langle m_0 \rangle}{V_{cell}}, \quad p_0 = \frac{\langle p_0 \rangle}{V_{cell}}.$$

The deterministic equations (1) and (2) can be cast in the universal dynamical system form

$$\frac{dx}{dt} = f(\mathbf{x}), \quad (3)$$

where \mathbf{x} is the chemical species vector (in the above case given by (m, p)).

As we said this deterministic model (3) is not good for genetic processes. In reality we have $\mathbf{n} = n_1, n_2, n_3, \dots$ molecules of different species. A chemical reaction makes a transition between the state characterised by \mathbf{n} to the state characterised by \mathbf{m} . Accordingly we have transition amplitudes and the full reaction is described by the chemical Master equation [7]:

$$\frac{\partial P(\mathbf{n}, t)}{\partial t} = \sum_{\mathbf{m}} W_{\mathbf{m} \rightarrow \mathbf{n}} P(\mathbf{m}, t) - W_{\mathbf{n} \rightarrow \mathbf{m}} P(\mathbf{n}, t), \quad (4)$$

where $P(\mathbf{n}, t)$ is the probability that the system is in state \mathbf{n} at the time t , $W_{\mathbf{m} \rightarrow \mathbf{n}}$ are the transition probabilities for reaction events moving the system in the state \mathbf{n} from state \mathbf{m} . The discrete transition probabilities can be expressed in terms of discrete operator $e^{k\hat{n}_i}$ which acts as:

$$e^{\pm k\hat{n}_i} f(\dots n_i \dots) = f(\dots n_i \pm k \dots).$$

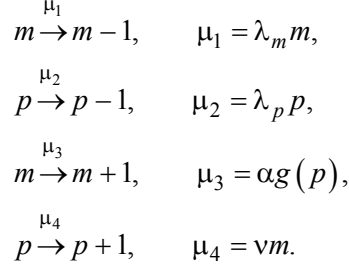
The first and second moments of $P(\mathbf{n}, t)$ with respect to a species i are the average number $\langle n_i(t) \rangle$ and the variance $\langle n_i^2 \rangle - \langle n_i \rangle^2 \equiv \sigma_i(t)$. The *intrinsic noise* is defined by

$$\eta_{\text{int}} = \frac{\sigma_i(t)}{\langle n_i(t) \rangle}.$$

Now we can write down the Master equation for our simplified model given by the equations (1) and (2). Since we have two type of chemicals (mRNA and proteins) we have the probability to depend on number of molecules \mathbf{m} for mRNA and \mathbf{p} for protein (here they are not concentrations as in (1) and (2)). So the form will be the following (g is considered constant):

$$\begin{aligned} \frac{\partial P(\mathbf{m}, \mathbf{p}, t)}{\partial t} = & g\alpha \left(P(\mathbf{m}-1, \mathbf{p}, t) \delta_{\mathbf{m},0} - P(\mathbf{m}, \mathbf{p}, t) \right) + \\ & + \lambda_m \left((\mathbf{m}+1) P(\mathbf{m}+1, \mathbf{p}, t) - \mathbf{m} P(\mathbf{m}, \mathbf{p}, t) \right) + \\ & + \nu \left(P(\mathbf{m}, \mathbf{p}-1, t) \delta_{\mathbf{p},0} - P(\mathbf{m}, \mathbf{p}, t) \right) + \\ & + \lambda_p \left((\mathbf{p}+1) P(\mathbf{m}, \mathbf{p}+1, t) - \mathbf{p} P(\mathbf{m}, \mathbf{p}, t) \right). \end{aligned}$$

So we have a linear differential discrete equation which can be solved using Laplace transform together with the constraint $\sum_{m,p} P(m, p, t) = 1$. However the main problem appears when g is no longer a constant. In order to cope with this situation we have to introduce some approximation methods, namely the linear noise approximation. But before this let us introduce the *propensity* and *stoichiometry* [8] which will make the connection between deterministic dynamics and stochastic one. The propensity tells how frequently a reaction occurs and the stoichiometry tells us how much the system is changed when the reaction is completed. In our model given by the equations (1) and (2) we have two reactions showing protein and mRNA degradation and two synthesis reactions for mRNA and protein as well. Accordingly we have:



Now our deterministic system will have the following form:

$$\frac{d}{dt} \begin{pmatrix} m \\ p \end{pmatrix} = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{pmatrix} = \mathbf{S}\mathbf{N},$$

where \mathbf{S} is the *stoichiometry* matrix and \mathbf{N} is the *propensity* matrix.

Now we can describe the linear noise approximation generally. We consider the chemical master equation (4). We want to approximate the discrete space (the number of molecules) which appears in the Master equation with the continuum space existent in the deterministic system (3) of course with calculation of the additional terms describing the fluctuations along macroscopic trajectory. More precisely the number of molecules $n_i(t)$ of a species i is approximated in terms of concentrations $x_i(t)$ and the term $\alpha_i(t)$ describing the deviation from $x_i(t)$ as a result of intrinsic noise. So we have:

$$n_i(t) = V_{cell} x_i(t) + \sqrt{V_{cell}} \alpha_i(t). \quad (5)$$

Also the step operator is approximated as:

$$\exp(\pm k \partial_{n_i}) = 1 \pm \frac{1}{\sqrt{V_{cell}}} \partial_i + \frac{k^2}{2V_{cell}} \partial_i^2,$$

where $\partial_i = \partial / (\partial \alpha_i)$.

We want to calculate the average fluctuations $\langle \alpha_i \rangle$ and the variances $\langle \alpha_i \alpha_j \rangle$ so we need a probability distribution for $\alpha = \alpha_1, \alpha_2, \dots, \alpha_d$ where d is the number of distinct species of the system. The probability distribution is introduced as $\Pi(\alpha, t) = V_{cell}^{d/2} P(\mathbf{n}, t)$. Now introducing (5) in (4) and expand in inverse powers of V_{cell} at the zero order we get

$$\frac{d\mathbf{x}}{dt} = f(\mathbf{x}),$$

and at the order $-1/2$ we get the Fokker-Planck equation:

$$\frac{\partial \Pi}{\partial t} = -\sum_{i,j} A_{ij} \partial_i (\alpha_j \Pi) + \frac{1}{2} \sum_{i,j} B_{ij} \partial_{ij} \Pi, \quad (6)$$

where

$$A_{ij}(t) = \frac{\partial f_i}{\partial x_j}, \quad B_{ij} = \mathbf{S} \cdot \text{diag}[\mathbf{N}] \cdot \mathbf{S}^T.$$

The matrices \mathbf{A} and \mathbf{B} are independent of the vector α which appears only linearly in the drift term. Accordingly the Fokker-Planck equation can be solved and the solution is a Gaussian

$$\Pi(\alpha, t) = \frac{1}{\sqrt{(2\pi)^d \det[\mathbf{C}]}} \exp\left[-\frac{1}{2} \alpha^T \mathbf{C}^{-1} \alpha\right],$$

where the variance matrix $C_{ij} = \langle \alpha_i \alpha_j \rangle$ obeys the following fluctuatio-dissipation differential equation:

$$\frac{d\mathbf{C}}{dt} = \mathbf{A}\mathbf{C} + \mathbf{C}\mathbf{A}^T + \mathbf{B}. \quad (7)$$

Now the statistics can be computed immediately from the differential equation for \mathbf{C} because the mean $\langle \alpha \rangle$ is zero for all time if initial state is known precisely. This can be seen from the equation of $\langle \alpha \rangle$ which is nothing but the linearised dynamical system (3)

$$\frac{\partial}{\partial t} \langle \alpha \rangle = \mathbf{A} \langle \alpha \rangle.$$

3. SELFACTIVATOR WITH COOPERATIVE STABILITY

Now we can introduce our model. In its simplified form it has been introduced in [1]

$$\begin{aligned} \frac{dm}{dt} &= \alpha g([TF]) - \lambda_m m, \\ \frac{dp}{dt} &= vm - (\lambda_1 p_1 + 2\lambda_2 p_2). \end{aligned} \quad (8)$$

In the TF-degradation term the total protein concentration $p = p_1 + 2p_2$ is shared into monomers and homodimers. Also in the case of cooperative stability ($\lambda_1 > \lambda_2$) then the TF-equation is nonlinear.

In our model we want to take into account the full process of dimerisation. If it is assumed to be rapid then $p_2 = p_1^2 / K_d$, where K_d is the dissociation constant at equilibrium which is assumed to be *constant*. However K_d is not constant because proteolysis and dilution modify the steady protein concentrations. The combined effect of all these processes gives the following generalised model of the self-activator:

$$\frac{dm}{dt} = \alpha g([TF]) - \lambda_m m, \quad (9)$$

$$\frac{dp_2}{dt} = \nu m - \lambda_1 p_1 - 2k_a p_1^2 + 2k_d p_2 \equiv f_2(p_1, p_2), \quad (10)$$

$$\frac{dp_1}{dt} = k_a p_1^2 - k_d p_2 - \lambda_2 p_2 \equiv f_1(p_1, p_2). \quad (11)$$

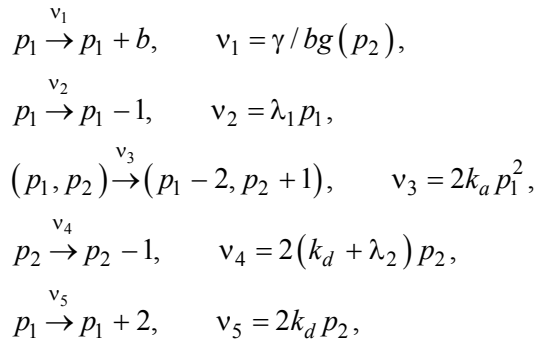
Because the mRNA turnover is very fast compared to the protein's we can consider the equation (9) to be stationary so we can solve it for m and introduce in (10). The final system will be:

$$\frac{dp_1}{dt} = \gamma g(p_2) - \lambda_1 p_1 - 2k_a p_1^2 + 2k_d p_2 \equiv f_1(p_1, p_2), \quad (12)$$

$$\frac{dp_2}{dt} = k_a p_1^2 - k_d p_2 - \lambda_2 p_2 \equiv f_2(p_1, p_2), \quad (13)$$

where $\gamma = \alpha \nu / \lambda_m$. Immediately one can see that if the dimerisation process is fast then $k_a p_1^2 = (k_d + \lambda_2) p_2$. In this case (10)+(11) gives precisely (8).

In order to study the stochastic effects on (12) and (13) we need the propensity and stoichiometry matrices. The elementary biochemical processes are the following:



where $b = v/\lambda_m$ is the stochastic bursting in protein production.

So the dynamical system (12), (13) is given by:

$$\frac{d}{dt} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} b & -1 & -2 & 0 & 2 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \end{pmatrix} = \mathbf{S}\mathbf{N}.$$

The diffusion matrix in the Fokker-Planck equation is the following:

$$\mathbf{B} = \mathbf{S} \cdot \text{diag}[\mathbf{N}] \cdot \mathbf{S}^T = \begin{pmatrix} b^2 v_1 + v_2 + 4v_3 + 4v_5 & -2v_3 \\ -2v_3 & v_5 + v_3 + v_4 \end{pmatrix}.$$

Now we have to evaluate the drift matrix \mathbf{A} at the stationary points of (12), (13) namely p_1^0, p_2^0 :

$$\mathbf{A} = \begin{pmatrix} \partial f_1 / \partial p_1 & \partial f_1 / \partial p_2 \\ \partial f_2 / \partial p_1 & \partial f_2 / \partial p_2 \end{pmatrix} = \begin{pmatrix} -\lambda_1 - 4k_a p_1^0 & \gamma g'(p_2^0) + 2k_d \\ 2k_a p_1^0 & -(k_d + \lambda_2) \end{pmatrix}.$$

Since these matrices are time independent the system (7) which gives the variance matrix \mathbf{C} is stationary so we have an algebraic system for correlation functions. The solutions can be written analytically but their expression is very long and not useful as they stand. Rather, we intend to see the value of dimer-protein noise. It attains the maximal value at the low stationary point where the concentration p_2^0 is small. Moreover at this point $g'[p_2]$ becomes practically zero and $g[p_2^0]$ is a constant, namely the inverse of the *fold activation*. Expanding in power series the $\langle \alpha_2^2 \rangle$ we get:

$$\begin{aligned} \langle \alpha_2^2 \rangle & \approx \frac{(k_d + \lambda_2)\lambda_1 + \lambda_1^2 + 2b\gamma k_a f^{-1}}{\lambda_1(k_d + \lambda_1 + \lambda_2)} p_2^0 + \mathcal{O}\left((p_2^0)^{3/2}\right), \\ \langle \alpha_1 \alpha_2 \rangle & \approx \frac{b\gamma f^{-1} k_a p_1^0}{\lambda_1(k_d + \lambda_1 + \lambda_2)} + \mathcal{O}\left((p_1^0)^2\right), \\ \langle \alpha_1^2 \rangle & \approx \frac{b\gamma f^{-1}}{2\lambda_1} + \left(\frac{1}{2} - \frac{2b\gamma k_a (\lambda_1 + \lambda_2)}{\lambda_1^2 (k_d + \lambda_1 + \lambda_2)} \right) p_2^0 + \mathcal{O}\left((p_1^0)^2\right). \end{aligned} \tag{14}$$

We can see a full dependence of fluctuations on degradation rates. With the above formulas we can compute the dimeric protein noise which is given by:

$$\eta_2 = \frac{\sqrt{\langle p_2^2 \rangle - \langle p_2 \rangle^2}}{p_2^0},$$

where

$$p_2 = p_2^0 + \frac{1}{\sqrt{V_{cell}}} \alpha_2.$$

Taking the average after making the square we get:

$$\eta_2 = \sqrt{\frac{(k_d + \lambda_2)\lambda_1 + \lambda_1^2 + 2b\gamma_1 k_a f^{-1}}{V_{cell} p_2^0 \lambda_1 (k_d + \lambda_1 + \lambda_2)}},$$

where k_a is the association rate, k_d dissociation rate.

In the same way we can compute the protein noise for monomers. The result is

$$\eta_1 = \sqrt{\frac{b\gamma f^{-1}}{2V_{cell} (p_1^0)^2 \lambda_1} + \frac{1}{V_{cell} p_1^0} \left(\frac{1}{2} - \frac{2b\gamma k_a (\lambda_1 + \lambda_2)}{\lambda_1^2 (k_d + \lambda_1 + \lambda_2)} \right)}.$$

From the above formulas one can see that contrary to the deterministic case there is no imposed relation between λ_1 and λ_2 which would decrease the noise. Rather, the noise decreases uniformly as *both* λ_1 and λ_2 increase. Accordingly the cooperative stability effect seems to be valid only for enlarging deterministic bistability parameter space. However the degradation rates should not be very small in order to apply deterministic methods.

4. VALIDITY OF EFFECTIVE RATE EQUATION

So far we have discussed the full set of rate equations in order to see precisely the distribution probability of fluctuations. But in many cases we are interested in some effective equations in order to extract easier information about, for instance, bistability behaviour at stochastic perturbations. In order to see this let us work with the initial form.

$$\frac{dm}{dt} = \alpha g([TF]) - \lambda_m m,$$

$$\frac{d}{dt}(p_1 + 2p_2) = vm - (\lambda_1 p_1 + 2\lambda_2 p_2). \quad (15)$$

At the rapid dimerisation we have $p_2 = k_a / (\lambda_2 + k_d) p_1^2$. Accordingly, assuming the mRNA turnover is small we end up with a single equation for the monomer production of the gene:

$$\frac{d}{dt} \left(p_1 + \frac{2k_a}{\lambda_2 + k_d} p_1^2 \right) = \gamma g(p_1^2) - \lambda_1 p_1 - \frac{2\lambda_2 k_a}{\lambda_2 + k_d} p_1^2,$$

which can be written as an *effective rate equation*:

$$\frac{dp_1}{dt} = \gamma G_A(p_1) - D_A(p_1), \quad (16)$$

where

$$G_A(p_1) = \frac{g(p_1^2)}{1 + 4k_a p_1 / (\lambda_2 + k_d)},$$

$$D_A(p_1) = \frac{(\lambda_2 + k_d)\lambda_1 p_1 + 2\lambda_2 k_a p_1^2}{\lambda_2 + k_d + 4k_a p_1}.$$

Also here we express $g(p_1)$ with the aid of f-maximum fold activation in the circuit and k_A equilibrium dissociation constant of the activator and its cognate binding site.

Applying here the linear noise approximation one gets immediately

$$\langle \alpha_1^2 \rangle = \frac{\gamma_1 G_A(p_1) - D_A(p_1)}{2D'_A(p_1) - 2G'_A(p_1)} \Big|_{p_1=p_1^0},$$

which differs substantially from the previous expression (14). Accordingly it is important to study the full system of rate equations in order to have a good view of the stochastic effects.

5. CONCLUSIONS

In this paper we analysed the stochastic effects on the protein production for a selfactivator with transcription factors as homodimers. Theoretical and experimental results show that proteins in this form are more stable and the properties of genetic circuits are strongly improved by including the so called

cooperative stability – different turnover rates for monomers and dimers. Here we studied the stochastic effects on the full rate equations treating in all generality the transcription, translation and dimerisation processes. Using linear noise approximation we computed the correlation functions for fluctuations in both monomer and dimer proteins for a general promoter activity function. It was shown that in the first order, for low stationary states the protein noise is independent on a specific relation between turnover rates. The only way which may decrease the noise is a bigger value for both of them. However we have to point out that we used the linear noise approximation which assumes from the beginning that the noise is *small* so the validity of our results is somehow limited. An example given for an effective rate equation has been treated in the last section where it was shown that the noise is different from the results emerging the full systems of rate equations. Effects on bistability will be treated in a future publication.

Acknowledgements. This work was fully supported by the CNCSIS Project 325, “Nonlinear Dynamics and Stochastic Effects in Gene Regulatory Networks” Romanian Ministry of Education and Research.

REFERENCES

1. N. Buchler, U. Gerland, T. Hwa, Proc. Nat. Acad. Sci USA, **102**, 9559 (2005).
2. M. Gonzales, R. Woodgate, BioEssays, **24**, 141 (2002).
3. P. R. Johnson et al. Cell, **94**, 217 (1998).
4. E. Aurell, K. Sneppen, Phys. Rev. Lett, **88**, 048101 (2002).
5. E. M. Ozbudak et al. Nature Genetics, **31**, 69 (2002).
6. A. Becksey, L. Serrano, Nature, **405**, 590 (2000).
7. N. G. van Kampen, Stochastic processes in physics and chemistry, Elsevier, 1992.
8. M. Scott, B. Ingalls, M. Kaern, Chaos, **16**, 026107 (2006),
M. Thattai, A. van Oudenaarden, Proc. Nat. Acad. Sci USA, **98**, 8614, (2005), J. M. Raser, E. K. O’Shea, Science, **309**, 2010 (2005).