Available online at www.sciencedirect.com

ScienceDirect

IFAC Papers Online CONFERENCE PAPER ARCHIVE

IFAC-PapersOnLine 49-24 (2016) 001–005

Stability of stochastic gene regulatory networks using entropy methods ⋆

M. Pájaro * **A. A. Alonso** ** **J. A. Carrillo** *** **C. Vázquez** ****

* *Process Engineering Group, IIM-CSIC. Spanish Council for Scientific Research. Eduardo Cabello 6, 36208 Vigo, Spain (e-mail: mpajaro@iim.csic.es).*

** *Process Engineering Group, IIM-CSIC. Spanish Council for Scientific Research. Eduardo Cabello 6, 36208 Vigo, Spain (e-mail: antonio@iim.csic.es)*

*** *Department of Mathematics, Imperial College London, SW7 2AZ, UK (e-mail: carrillo@imperial.ac.uk)*

**** *Departament of Mathematics, Faculty of Informatics. Campus de Elviña, s/n 15071 - A Coruña, Spain (e-mail: carlosv@udc.es)*

Abstract: The study of self regulated gene expression networks must be modelled using chemical master equations. However, its solution is not available in the most cases. In this work, we derive a partial integral differential model as the continuous counterpart of one master equation with jump process. This model allows us to reproduce numerically the dynamic behaviour of the protein distribution whose steady state admits an analytical solution. To study the convergence to the equilibrium, we test the applicability of entropy methods. Using these techniques we find numerical evidences of exponential stability. The derivation and methods presented can be of the help to extend the applicability of this model to more complex gene regulatory networks including more than one protein.

Keywords: Gene circuits, mathematical models, stochastic systems, entropy methods, stability, dynamics.

1. INTRODUCTION

The study of the DNA expression (transcription into messenger RNA and translation into proteins) and their regulation becomes essential to predict the response of cells to environmental signals. The regulatory mechanism normally takes place under the union of proteins to the DNA binding sites that inhibit or activate its expression. Typically, the number of molecules involved in the regulation mechanism is small, thus making gene expression a truly stochastic process (Gillespie, 2007; Kepler and Elston, 2001).

The chemical master equation (CME) is at the basis of dynamic reaction network modelling (Kepler and Elston, 2001; Paulsson, 2005; Mackey et al., 2011; Sherman and Cohen, 2014) as the method which incorporates the underlying stochastic behaviour. However, the CME solution cannot be obtained in most cases, due to the large (even infinite) number of coupled equations. Although computationally very involved, extensive stochastic simulations via SSA (Gillespie, 1976) are typically the approach adopted to reproduce the CME dynamics. Alternatives are CME approximations, such as, moment methods (Engblom, 2006), finite state projection (Munsky and Khammash, 2006) or hybrid models (Jahnke, 2011). Unfortunately, those methods are only able to approximate the CME solution in quite particular situations.

In case of gene self regulatory networks, the obstacles to the solution of the CME can be overcome by the 1D partial integral differential equation (PIDE) model proposed by Friedman et al. (2006). To the best of our knowledge, a rigorous deduction of the PIDE model from the CME has not been reported yet. Here we show that, under protein production in bursts (Friedman et al., 2006; Shahrezaei and Swain, 2008; Dar et al., 2012), the PIDE model can be deduced as the continuous counterpart of one CME with jump processes. Using this 1D PIDE model, we can both reproduce the dynamics of one protein distribution and obtain an analytical solution for the steady state.

In addition, we make use of an entropy method (Michel et al., 2005; Cáceres et al., 2011; Carrillo et al., 2011) to study stability of steady state solutions of the PIDE system. In particular, we test the applicability of entropy methods show asymptotic stability and give numerical evidences of the exponential rate of convergence.

The contribution is structured as follows: In Section 2 we discuss the gene regulatory network and its corresponding CME and PIDE dynamic descriptions. The entropy methods together with results on asymptotic and exponential stability are presented in Section 3. We end up with some conclusions and future work.

⋆ This work has been partially supported by Ministerio de Economia y Competitividad AGL2012-39951-C02-01. M. Pájaro acknowledge support from grants BES-2013-063112 and EEBB-I-16-10540.

10.1016/j.ifacol.2016.10.741

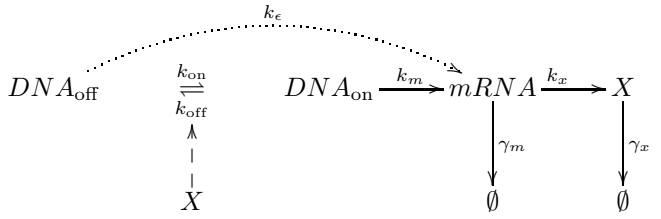

Fig. 1. Schematic representation of the transcription-translation mechanism under study. The promoter associated with the gene of interest is assumed to switch between active (DNA_{on}) and inactive (DNA_{off}) states, with rate constants k_{on} and k_{off} per unit time, respectively. In this study, the transition is assumed to be controlled by a feedback mechanism induced by the binding/unbinding of a given number of X-protein molecules, what makes the network self-regulated. Transcription of messenger RNA ($mRNA$) from the active DNA form, and translation into protein X are assumed to occur at rates (per unit time) k_m and k_x, respectively. k_ε is the rate constant associated with transcriptional leakage. Both $mRNA$ and X-protein degradation are assumed to occur by first order processes with rate constants γ_m and γ_x, respectively.

2. SELF REGULATORY GENETIC SYSTEMS

The genetic system under study consists of a transcription-translation network involving a single gene that expresses a protein X which regulates its own production. The representative biochemical steps, including protein and $mRNA$ degradation, are depicted in Fig. 1. We represent also a basal transcription level from the inactive promoter which takes place at a rate constant k_ε lower than k_m, (Friedman et al., 2006; Ochab-Marcinek and Tabaka, 2015; Huang et al., 2015).

Typically, the self regulation mechanism is described by one input function of the form (Friedman et al., 2006; Ochab-Marcinek and Tabaka, 2015; Pájaro et al., 2015):

$$c(x) = [1 - \rho(x)] + \rho(x)\varepsilon\,, \tag{1}$$

with x representing protein level, $\varepsilon = \frac{k_\varepsilon}{k_m} \in (0,\,1)$ the transcriptional leakage constant and $\rho(x)$ a Hill type function (Alon, 2007) that relates x to the fraction of DNA_{off}:

$$\rho(x) = \frac{x^H}{x^H + K^H}. \tag{2}$$

where $K = \frac{k_{off}}{k_{on}}$ is an equilibrium constant and H the Hill coefficient, proportional to the number of protein molecules bonded to the promoter. Its values can be positive or negative depending on whether the circuit represses or promotes protein production, thus resulting into a negative or positive feedback, respectively.

2.1 Continuous formulation deduction

In the following, we consider gene self regulatory networks where the degradation rate of $mRNA$ is much faster than the corresponding to protein, so that $\gamma_m/\gamma_x \gg 1$. Such condition is verified in many gene regulatory networks, both in prokaryotic and eukaryotic organisms (Shahrezaei and Swain, 2008; Dar et al., 2012), and results in protein being produced in bursts. As suggested in Friedman et al. (2006); Elgart et al. (2011), the burst size (denoted by $b = \frac{k_x}{\gamma_m}$) is typically modelled by an exponential distribution. The conditional probability for protein level to jump from a state y to x after a burst is proportional to:

$$\omega(x-y) = \frac{1}{b}\exp\left[\frac{-(x-y)}{b}\right] \tag{3}$$

This burst behaviour in protein production can be modelled by the superposition of jumps from lower states as it is depicted in Fig. 2. We define $g_i^n : \mathbb{N} \to [0\,1]$ as the transition probability for a jump going from a lower state i into a state n, assuming that the size of the jump follows the expression (3). Furthermore, the transition probability is proportional to the messenger RNA production rate, so that, g_i^n is defined as:

$$g_i^n := k_m c(i)\omega(n-i). \tag{4}$$

Let $P : \mathbb{R}_+ \times \mathbb{N} \to [0\,1]$, be the probability of having n proteins at time t. The time evolution of $P(t,n)$ is given by a chemical master equation (Gardiner, 2009; Van Kampen, 2007) with jumps that reads:

$$\begin{aligned}\frac{\partial P(t,n)}{\partial t} = &\sum_{i=0}^{n-1} g_i^n P(t,i) - \sum_{i=n+1}^{\infty} g_n^i P(t,n) \\ &+ r_{n+1}P(t,n+1) - r_n P(t,n),\end{aligned} \tag{5}$$

where $r_n = \gamma_x n$ represents the degradation transition probability. In order to obtain a continuous version of (5) we define $p : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$, as the continuous protein probability distribution, and add and subtract $g_n^n P(t,n)$ at the right hand side of (5) to get:

$$\sum_{i=0}^{n-1} g_i^n P(t,i) - \sum_{i=n+1}^{\infty} g_n^i P(t,n) = \sum_{i=0}^{n} g_i^n P(t,i) - \sum_{i=n}^{\infty} g_n^i P(t,n). \tag{6}$$

Next, approximating the summations at the right hand side of the last equation by integrals and substituting in (5) we obtain:

$$\begin{aligned}\frac{\partial p(t,x)}{\partial t} = &\int_0^x g_y^x p(t,y)\,\mathrm{d}y - \int_x^\infty g_x^y p(t,x)\,\mathrm{d}y \\ &+ r_{x+1}p(t,x+1) - r_x p(t,x),\end{aligned} \tag{7}$$

where the integer indexes n and i are substituted by real x and y respectively. Note that the second term at right hand side in (7) reduces to:

$$\int_x^\infty g_x^y p(t,x)\,\mathrm{d}y = k_m c(x) p(t,x) \int_x^\infty \omega(y-x)\,\mathrm{d}y, \tag{8}$$

with $\int_x^\infty \omega(y-x)\,\mathrm{d}y = 1$. Employing the Taylor theorem to approximate the third term at right hand side in (7) to the first order, we also get:

$$r_{x+1}p(t,x+1) \approx r_x p(t,x) + \frac{\partial\,[r_x p(t,x)]}{\partial x} \tag{9}$$

Finally, replacing the last expressions (8)-(9) in (7) and using a dimensionless time, $\tau = \gamma_x t$, associated with the time scale of protein degradation, we obtain the temporal evolution of the probability distribution $p(\tau, x)$, which reads as: