BIOINFORMATICS ORIGINAL PAPER
Vol. 23 no. 23 2007, pages 3185–3192
doi:10.1093/bioinformatics/btm490

Systems biology

Steady-state expression of self-regulated genes

T. Fournier[1], J.P. Gabriel[1], C. Mazza[1,*], J. Pasquier[1], J.L. Galbete[2] and N. Mermod[2,*]

[1]Department of Mathematics, University of Fribourg, Chemin du Musée 23, CH-1700 Fribourg and [2]Institute of Biotechnology, University of Lausanne, CH-1015 Lausanne, Switzerland

Received on May 23, 2007; revised on August 29, 2007; accepted on September 24, 2007

Advance Access publication October 12, 2007

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Regulatory gene networks contain generic modules such as feedback loops that are essential for the regulation of many biological functions. The study of the stochastic mechanisms of gene regulation is instrumental for the understanding of how cells maintain their expression at levels commensurate with their biological role, as well as to engineer gene expression switches of appropriate behavior. The lack of precise knowledge on the steady-state distribution of gene expression requires the use of Gillespie algorithms and Monte-Carlo approximations.

Methodology: In this study, we provide new exact formulas and efficient numerical algorithms for computing/modeling the steady-state of a class of self-regulated genes, and we use it to model/compute the stochastic expression of a gene of interest in an engineered network introduced in mammalian cells. The behavior of the genetic network is then analyzed experimentally in living cells.

Results: Stochastic models often reveal counter-intuitive experimental behaviors, and we find that this genetic architecture displays a unimodal behavior in mammalian cells, which was unexpected given its known bimodal response in unicellular organisms. We provide a molecular rationale for this behavior, and we implement it in the mathematical picture to explain the experimental results obtained from this network.

Contact: christian.mazza@unifr.ch, nicolas.mermod@unil.ch

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 INTRODUCTION

The regulation of gene expression is at the center of many central biological processes comprising embryo development and cell differentiation, while improper regulation can result in diseases such as cancer. Genomic studies have revealed that genes form part of complex regulatory networks, whereby the protein determined by one gene may control the transcription of numerous other genes, which in turn may regulate the synthesis of further proteins. However, how cells usually maintain the expression of their genes at levels commensurate with their biological role and own survival is still poorly understood. Recently, a growing number of evidence was given for stochastic or probabilistic mechanisms of gene regulation, whereby a population of cells seemingly identical genetically and submitted to the same environment may not express their genes at the same level. Rather, gene expression was often found to oscillate around an average value. This fluctuation, or noise, has been attributed to rare regulatory events, a consequence of the small number of regulatory molecules controlling a given gene, or to global stochastic variations of the cell physiology (Paulsson, 2004 or 2005; Raser and O'Shea, 2004). In addition, gene expression may oscillate between various semi-stable states mediated by positive or negative feedback loops within regulatory networks. The promoters of the genes switch randomly between active states (on), during which transcription occurs at some rate, and inactive states (off), inducing molecular intrinsic noise. In this context, noisy signals may be instrumental in mediating cells or virus switching between distinct semi-stable gene expression and biological status (McAdams and Arkin, 1997; Weinberger *et al.*, 2005). Finally, particular gene network architectures may themselves contribute to amplify the fluctuations caused by noisy upstream signals. Experimental and theoretical studies show that positive feedback coupled to molecular noise generates stochastic gene expression, and induces variable phenotypes (Arkin *et al.*, 1998; Austin *et al.*, 2005; Becskei *et al.*, 2001; Blake and Collins, 2005; Guido *et al.*, 2006; Isaacs *et al.*, 2003; Raser and O'Shea, 2004; Weinberger *et al.*, 2005). Such self-regulated networks are currently developed in engineered modular systems. To comprehensively understand gene regulation, as well as to be able to genetically engineer gene expression switches of appropriate properties, it would be useful to describe the kinetics and steady-states of gene networks using mathematical models that would account for their noisy behavior (Guido *et al.*, 2006).

The usual approach considers a time continuous Markov chain, the Gillespie algorithm, which models the various chemical reactions involved in the system (Adalsteinsson *et al.*, 2004; Gillespie, 1977, 2001). Such systems contain reactions evolving at different time scales. The system can be divided in two parts, the fast and the slow components; one then focus on the slow reactions by assuming a quasi-equilibrium where fast reactions equilibriate instantaneously (Burrage *et al.*, 2004; Cao *et al.*, 2005b; Goutsias, 2005; Kepler and Elston, 2001). Typically, reactions that involve the synthesis of a protein from a gene are slow, of the order of 40 min in eukaryotes, while the association of proteins in

*To whom correspondence should be addressed.

multimeric forms such as dimers takes place within seconds. Here, we approach such systems by modeling auto-regulation within the context of self-regulated genes and for a network regulating the expression of therapeutic proteins.

The Gillespie algorithm simulates the time-evolution of the number of molecules of each species, and it is described by a chemical master equation (see e.g. 2). For large gene networks, the computation time can be overwhelming. Therefore, alternatives to the Gillespie algorithm have been designed to accelerate simulations, such as the *tau-leaping algorithm* (Gillespie, 2001), or diffusions approximations at quasi-equilibrium (Kepler and Elston, 2001). There are also a variety of fast, often multi-scale, algorithms approximating the time-evolution of gene networks (El Samad *et al*., 2005), while multi-scale acceleration algorithms can be used to simulate large gene networks (Burrage *et al*., 2004; Cao *et al*., 2005a; Chatterjee *et al*., 2007). For smaller gene networks, particular simulation algorithms have been developed according to the specificity of the setting (Kepler and Elston, 2001; Salis and Kaznessis, 2005; Salis *et al*., 2006; Shibata, 2003a, b, see also Erban *et al*., 2006 for a coarse-graining approach). However, most of these dynamics do not simulate the biological process described by the chemical master equation, but rather propose approximations.

In what follows, we first consider a fundamental module consisting in a self-regulated gene, which is a building block of many gene networks. Usually, the related steady-state distribution is obtained through Monte-Carlo simulations of many random trajectories. We present a novel strategy to determine analytically the steady-state. This avoids stochastic simulations, as it provides exact values for the steady-state mean and variance, which might be useful for reverse engineering problems. We next model a regulatory network of interest in gene therapy, using a semi-stochastic model involving coupled self-regulated genes. The analytical results on the self-regulated gene are then used to study the regulatory network in comparison with experimental measurements.

2 MODELING OF SELF-REGULATED GENE

The system is composed of a promoter and a gene that can oscillate between a transcribed (active) and a silent state, as schematized in Figure 1. As stated above, one source of molecular noise is the random nature of the states taken by the promoter (on/off). The stochastic or noisy behavior of transcription is well documented both in prokaryotes and in eukaryotes, including mammalian cells (Blake *et al*. 2003; Raj *et al*., 2006). From the mathematical point of view, stochastic models are more general and extend naturally deterministic ones, as the latter describe in most cases the average behavior of random models. Furthermore, stochastic models provide information on the variance and coefficient of variation of the related gene product, which cannot be obtained from deterministic models. Figure 1 shows polypeptides produced during the transcription and translation processes. Protein monomers react quickly to form dimers: we assume a quasi-equilibrium where fast reactions equilibrate instantaneously. For a global amount of n polypeptides, the proportion of dimers at quasi-equilibrium is a well-defined function of n, see,

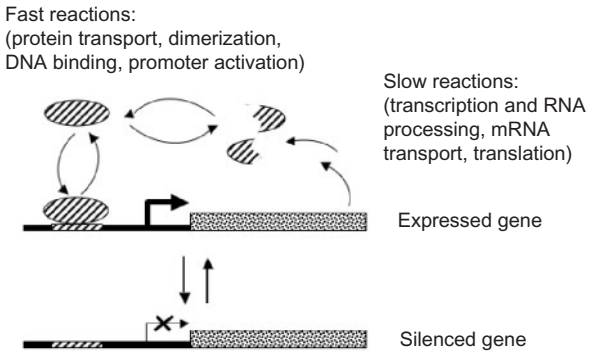

Fig. 1. Modeling of a self-regulated gene with a feed-back loop. Regulatory proteins are shown by ellipses while DNA elements are shown by rectangular boxes, i.e. wide boxes for protein-coding sequences and thin boxes for regulatory sequences. The gene can oscillate between an expressed ON state where all components of the transcription machinery are bound to the promoter and where RNA polymerase can initiate transcription, and a silent or OFF state. Expression leads to the formation of monomeric proteins in a series of slow processes, as indicated. A series of fast transitions lead to the binding of the activator protein to the promoter of its own gene, increasing the probability of observing the ON state. RNA polymerase and other proteins mediating transcription or translation are omitted for clarity.

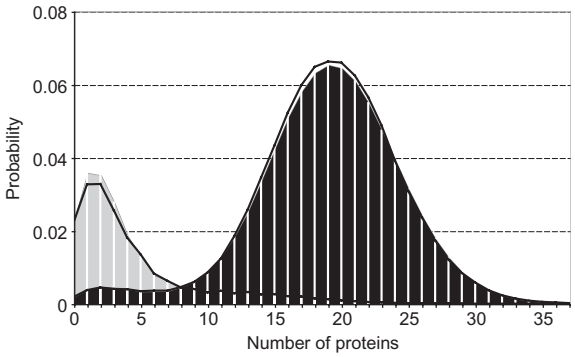

Fig. 2. Plot of the steady-state distribution related to a bistable self-regulated gene, for the off (light bars) and on (dark bars) regimes. The distribution is bimodal in both regimes, and has been computed using the formula (4). The curves are obtained from simulations based on the Gillespie algorithm.

e.g. the *Supporting Information*. Dimers can bind to specific DNA sequences near the promoter, and thereby enhance transcription, corresponding to a positive feedback loop. These binding events can be assumed to be fast with respect to events like protein formation. They are however included in some chain of events leading to the assembly of a multiprotein complex that ends with a state where the RNA polymerase initiates transcription. This will correspond to the on state $\mathcal{O}_1$. When these conditions are not satisfied, the promoter is off $\mathcal{O}_0$. The rates of transitions between these two states are functions of the proportion of dimers of the activator protein, and therefore are functions of n when the cell contains n polypeptides. These random events are usually modeled by

supposing that the probability that the promoter switches from the off to on state in a small time interval of length $h \approx 0$ is of order $g(n)h$ for n polypeptides, where the function g can be chosen according to the specificity of the setting. We shall consider in the next paragraph, a gene network where g is obtained by modeling noise sources. To be as general as possible, and to eventually allow negative feedback loops, we also assume that the probability of transition of the reverse reaction is given by some function $\kappa(n)$. Basal activity is introduced by supposing that $g(0)$ is positive, so that a transcription event can occur without protein dimers. The remaining involved chemical reactions are essentially protein monomers production and degradation, which are summarized in Figure 1. Transcription is stopped when the promoter is off, so that we assume that the probability $\mu_0 h$ that a protein is created during a small time interval of length h vanishes after some delay, with $\mu_0 \approx 0$. When the promoter is on, transcription is possible, and the probability that a transcription event occurs is of order μh. Degradation of protein dimers is summarized by the rate $\nu(n)$, for some function ν, which is usually linear as a function of n. The time evolution of the state of this self-regulated gene is described by a pair of time continuous stochastic process $N(t)$ and $Y(t)$, where $N(t)$ gives the number of proteins present in the cell at time t and where $Y(t)$ takes the values 0 and 1 corresponding to the off and on states of the promoter.

The reaction scheme is written as

$$\mathcal{O}_0 \underset{\kappa(n)}{\overset{g(n)}{\leftrightarrows}} \mathcal{O}_1, \ \mathcal{M} \xrightarrow{\nu} \emptyset, \quad \emptyset \xrightarrow{\mu_l} \mathcal{M}, \ l = 0, 1, \tag{1}$$

where $\mathcal{M}$ denotes protein monomers, and where μ_1 gives the production rate in the off state ($l = 0$) and in the on state ($l = 1$). Let $p_n^0(t) = P(N(t) = n, Y(t) = 0)$ and $p_n^1(t) = P(N(t) = n, Y(t) = 1)$ give the probability of having n proteins at time t when the states of the promoter are $\mathcal{O}_0$ and $\mathcal{O}_1$, respectively. We assume here that $0 \le n \le \Lambda$ for some integer Λ. The related Gillespie algorithm is given as a time-continuous Markov chain $\eta(t) = (N(t), Y(t))$ (Cao *et al.*, 2005b or Gillespie, 2001), where $N(t) \in \{0, 1, \ldots, \Lambda\}$ and $Y(t) \in \{0, 1\}$, with transition rates given by

$$\begin{aligned} &P((n, y), (n+1, y)) = \mu_y, \ P((n, y), (n-1, y)) = \nu(n), \\ &P((n, y), (n, 1-y)) = \kappa(n) \text{ when } y = 1, \text{ and,} \\ &P((n, y), (n, 1-y)) = g(n) \text{ when } y = 0. \end{aligned}$$

The chemical master equation associated to the reaction scheme (1) is then given by

$$\begin{aligned} \frac{\mathrm{d}p_n^s(t)}{\mathrm{d}t} &= \mu_s(p_{n-1}^s(t) - p_n^s(t)) + \nu(n+1)p_{n+1}^s(t) - \nu(n)p_n^s(t) \\ &\quad + (-1)^s(\kappa(n)p_n^1(t) - g(n)p_n^0(t)), \end{aligned} \tag{2}$$

where $s \in \{0, 1\}$ (Kepler and Elston, 2001; Shibata, 2003a or b). Even for this fundamental reaction scheme, the master Equation (2) cannot be solved explicitely.

The steady-state distribution π associated with (2) is obtained by letting $t \to \infty$: π is defined as

$$\pi_n(0) = \lim_{t \to \infty} p_n^0(t) \text{ and } \pi_n(1) = \lim_{t \to \infty} p_n^1(t),$$

and solves the linear system obtained from (2) by imposing $\mathrm{d}p_n^s/\mathrm{d}t = 0$:

$$\begin{aligned} 0 &= \mu_s(\pi_{n-1}(s) - \pi_n(s)) + \nu(n+1)\pi_{n+1}(s) - \nu(n)\pi_n(s)) \\ &\quad + (-1)^s(\kappa(n)\pi_n(1) - g(n)\pi_n(0)), \ s = 0, \ 1. \end{aligned}$$

The probability of observing n proteins at equilibrium is just $\pi_n(0) + \pi_n(1)$. The steady-state distribution has been approached using generating functions in Hornos *et al.* (2005) for self-regulated genes with g constant and $\kappa(n)$ linear; in this case, the limiting marginal probability of finding n proteins at steady-state is provided, and is expressed using hypergeometric functions. However, there is a lack of analytical results in the general case. Again, one can use Monte-Carlo methods based on approximating Markov chains to perform stochastic simulations. Our topic consists in providing the exact formula for π for the single gene described by (1) and (2). This will play a crucial rôle when considering the regulatory gene network of the next sections. In what follows, we give the exact formula for the steady-state, in the general case. The invariant measure $\pi_n(i)$ of the Gillespie algorithm is related to the steady-state distribution $\tilde{\pi}_n(i)$ of the associated discrete time jump chain: $\tilde{\pi}_n(i)$ is proportional to $\pi_n(i)\ q_{n,i}$, where $q_{n,i}$ is the sum of the transition rates from state (n,i). For $0 < n < \Lambda$, consider the matrices

$$Q_{n+1} = \begin{pmatrix} \frac{\nu(n+1)}{d_{n+1}} & 0 \\ 0 & \frac{\nu(n+1)}{c_{n+1}} \end{pmatrix}, \ R_n = \begin{pmatrix} 0 & \frac{g(n)}{d_n} \\ \frac{\kappa(n)}{c_n} & 0 \end{pmatrix},$$

and

$$P_{n-1} = \begin{pmatrix} \frac{\mu_0}{d_{n-1}} & 0 \\ 0 & \frac{\mu_1}{c_{n-1}} \end{pmatrix},$$

where $c_n = \nu(n) + \mu_1 + \kappa(n)$, $n < \Lambda$, $c_\Lambda = \nu(\Lambda) + \kappa(\Lambda)$, and $d_n = \nu(n) + \mu_0 + g(n)$ for $n < \Lambda$, $d_\Lambda = \nu(\Lambda) + g(\Lambda)$. The invariance of $\tilde{\pi}$ gives the relation $\tilde{\pi}_n = \tilde{\pi}_{n+1}Q_{n+1} + \tilde{\pi}_n R_n + \tilde{\pi}_{n-1}P_{n-1}$. The idea is to look for matrices α_n such that $\tilde{\pi}_n = \tilde{\pi}_{n+1}\alpha_n$. Plugging this relation in the above gives $\tilde{\pi}_n = \tilde{\pi}_{n+1}Q_{n+1}\gamma_{n-1}$, where the matrices γ_n are defined by $\gamma_{n-1} = (\mathrm{id} - R_n - \alpha_{n-1}P_{n-1})^{-1}$. One can then find the matrices α_n iteratively by considering the matrix valued continued fraction

$$\alpha_n = Q_{n+1}(\mathrm{id} - R_n - \alpha_{n-1}P_{n-1})^{-1}. \tag{3}$$

This provides an algorithm for computing exactly the steady-state distribution in the general case. Set $w_\Lambda := (\kappa(\Lambda)/c_\Lambda, 1)$. Then $\tilde{\pi}_n = (\tilde{\pi}_n(0), \tilde{\pi}_n(1))$ is given by

$$\begin{aligned} \tilde{\pi}_n &= w_\Lambda \alpha_{\Lambda-1}\alpha_{\Lambda-2}\cdots\alpha_n/\tilde{Z}_\Lambda, \\ \tilde{\pi}_\Lambda &= w_\Lambda/\tilde{Z}_\Lambda, \end{aligned}$$

where $\tilde{Z}_\Lambda$ is the normalization constant

$$\tilde{Z}_\Lambda = w_\Lambda\Big(\sum_{j=0}^{\Lambda-1} \alpha_{\Lambda-1}\cdots\alpha_j\Big)\mathbf{1} + w_\Lambda\mathbf{1}, \text{ and } \mathbf{1} = (1, 1)'.$$

The steady-state $\pi_n = (\pi_n(0), \pi_n(1))$ of the Gillespie algorithm is given by

$$\pi_n = \left(\frac{\tilde{\pi}_n(0)}{d_n Z_\Lambda}, \frac{\tilde{\pi}_n(1)}{c_n Z_\Lambda}\right), \tag{4}$$

where Z_Λ is the normalization constant

$$Z_\Lambda = \sum_{n=0}^{\Lambda} \left(\frac{\tilde{\pi}_n(0)}{d_n} + \frac{\tilde{\pi}_n(1)}{c_n}\right).$$

The complete proof of (4) is provided in the Appendix. When $\mu_0 = 0$ and $\mu_1 = \mu$, it turns out that one can solve explicitely (3). The solution is given by (see the Appendix)

$$\alpha_n = \frac{\nu(n+1)}{\mu} \begin{pmatrix} \frac{\kappa(n)+\mu}{d_{n+1}} & \frac{c_n}{d_{n+1}} \\ \frac{\kappa(n)}{c_{n+1}} & \frac{c_n}{c_{n+1}} \end{pmatrix}$$

Formula (4) must be used with care numerically since, when Λ is large, both the numerator and denominator rapidly diverge. It can be improved with the following normalisation algorithm. Let $\|\mathbf{w}\| := |w_1| + |w_2|$.

Normalization algorithm

(STEP 1): Define $\tilde{\nu}_n$ for $n = \Lambda - 1$ to 0 as

$$\tilde{\nu}_\Lambda := \frac{w_\Lambda}{\|w_\Lambda\|}, \quad \text{and} \quad \tilde{\nu}_n := \frac{\tilde{\nu}_{n+1}\alpha_n}{\|\tilde{\nu}_{n+1}\alpha_n\|}.$$

(STEP 2): Given the $\tilde{\nu}_n$, define $\nu_0 = \tilde{\nu}_0$ and, for $n = 1$ to Λ, set

$$\nu_n := \frac{\tilde{\nu}_n}{\|\tilde{\nu}_n\alpha_{n-1}\| \cdot \|\tilde{\nu}_{n-1}\alpha_{n-2}\| \cdots \|\tilde{\nu}_1\alpha_0\|}.$$

Finally the steady-state is obtained as

$$\tilde{\pi}_n = \frac{\nu_n}{V_\Lambda}, \qquad \text{where} \quad V_\Lambda := \sum_{i=0}^{\Lambda} \nu_i \cdot \mathbf{1}.$$

3 A REGULATORY NETWORK FOR EFFICIENT CONTROL OF TRANSGENE EXPRESSION

A more elaborate gene network consists of three genes. A first gene encodes a transcriptional repressor. Because this gene is expressed from an unregulated promoter, it mediates a stable number of repressor. This repressor binds to and inhibits the promoters of the two other genes, coding for a transactivator protein and for a quantifiable or a therapeutic protein, respectively (Fig. 3A). The activity of the repressor is inhibited by doxycycline, a small antibiotic molecule that acts as a ligand of the repressor and thereby controls its activity. Addition of the antibiotic will inhibit the repressor and relieve repression, allowing low levels of expression of the regulated genes and synthesis of some transactivator protein. This, in turn, allows further activation of the two regulated genes, in a positive feedback loop (Fig. 3B). When introduced in mammalian cells, this behaves as a signal amplifier and as a potent genetic switch, where the expression of a therapeutic gene can be controlled to vary from almost undetectable to very high levels in response to the addition of the antibiotic to the cells (Imhof *et al.*, 2000).

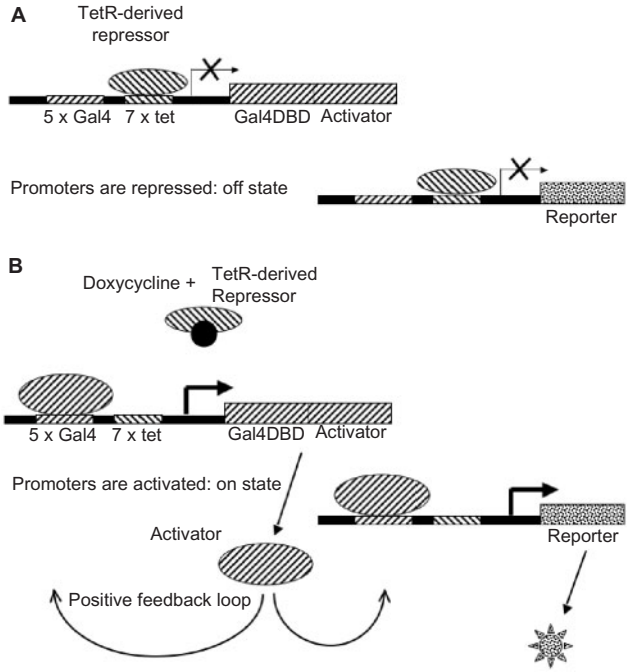

Fig. 3. Schematic representation of the regulatory network. Elements are as described in the legend to Figure 1, and activating and repressing elements are shaded with forward and backward stripes, respectively. The doxycycline inducer is represented by a filled circle, while the GFP or therapeutic protein is represented by a star. In the OFF-state, constitutively produced repressors, consisting of a fusion of the bacterial tetracycline repressor (TetR) to a eukaryotic repressor protein, bind to seven operator elements and thereby repress the promoters of both the reporter gene and of the activator. In the ON-state, doxycycline prevents the repressor from binding to DNA. Consequently, the activator, a fusion of the GAL4 DNA binding domain (GAL4DBD) with the VP16 activator is produced at low levels and it binds in its own promoter as well as that of the transgene. The autoregulatory positive feedback loop eventually leads to the build up of the activator concentration and to maximal activation of the reporter gene.

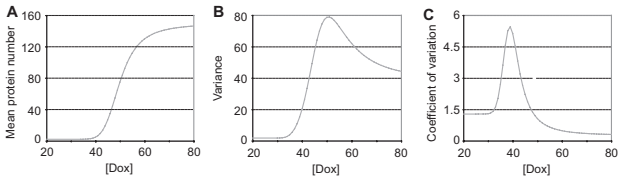

Fig. 4. These plots represent the mean and variance related to a self-regulated gene, here the transactivator contained in the network, and are obtained using (4).

Our formula is used to compute the steady-state mean and variance of transactivator proteins, as provided in Figure 4, as function of the number [Dox] of doxycycline molecules. One sees clearly the peak of variances observed previously in experimental data (Pedraza and van Oudenaarden, 2005).

3.1 A semi-stochastic model: inclusion of delays

The modelization of the time-evolution of the network is more involved. We propose a semi-stochastic model with time delays,

generalizing models considered recently in this setting by Goutsias and Kim (2006) and Bratsun *et al.* (2005), which allow the study of complex gene networks. We extend the models proposed in Goutsias and Kim (2006) by including stochastic signals related to promoters. We illustrate first these methods in the context of a self-regulated gene with some positive feedback rate $g(n)$ and no negative feedback, i.e. we assume that $\kappa(n)=\kappa$ is constant. The idea developed in Goutsias and Kim (2006) consists in replacing the feedback rate $g(N(t))$ by $c(t)=g(\mathbb{E}(N(t-\theta)))$, where θ is a biologically meaningful time delay, and where $\mathbb{E}$ denotes statistical average. In mammalian cells, proteins diffuse through random non-homogeneous environments like the nuclear and cytoplasmic spaces. Here, e.g. transactivator proteins diffuse and move eventually to the promoters associated with the activator and therapeutic genes: θ models then the average diffusion time needed to reach these promoters.

As explained in the next paragraph, our experimental results suggest that the complex structure of mammalian chromatin might act as a noise-filtering device that allows graded response from stochastic events; considering $\mathbb{E}(N(t-\theta))$ instead of $N(t)$ for feedback involving regulatory protein thus models this filtering process. The models proposed in Goutsias and Kim (2006) however do not involve the state of the promoter $Y(t)\in\{0,1\}$, but focus on $N(t)$, with transition rates of the form $P(n,\, n+1)=\mu g(\mathbb{E}(N(t-\theta)))/(g(\mathbb{E}(N(t-\theta)))+\kappa)$ and $P(n,\, n-1)=\nu(n)$, that is assume a quasi-equilibrium with fast assembly of the initiation complex. Assuming convergence of $\mathbb{E}(N(t))$ as $t\to\infty$, setting $g_\infty=g(\mathbb{E}(N(\infty)))$, the limiting process is a birth and death process with birth rate $\mu g_\infty/(g_\infty+\kappa)$ and death rate $\nu(n)$. When $\nu(n)=\nu n$, the related steady-state is then Poisson of parameter $\lambda=\frac{\mu}{\nu}g_\infty/(g_\infty+\kappa)$. This model is however not completely satisfactory since experimental data often exhibit peak of variances typical for bistable systems or for genetic switches (see, e.g. the data given in Pedraza and van Oudenaarden, 2005), while the mean is increasing as a function of the number of inducer molecules. This kind of behavior cannot be obtained with Poisson distributions. We therefore extend a mean field model proposed in Bratsun *et al.* (2005), which preserves the stochasticity of promoter fluctuations, and call this new model semi-stochastic. In fact, the dynamics related to these transition rates are different in nature of the preceding: the promoter is located near the upstream part of the gene, and the fact that the promoter is on just enhances the attachment of RNA polymerase. The related functions $c(t)$ satisfies a time-delayed differential equation, which can be obtained from the chemical master equation (see the Supplementary Material). The semi-stochastic transition rates corresponding to the related time non-homogeneous Markov chain are then given by

$$P((n,y),(n+1,y))=\mu_y,\;\; P((n,y),(n-1,y))=\nu(n),$$
$$P((n,y),(n,1-y))=\kappa \text{ when } y=1,$$
$$P((n,y),(n,1-y))=c(t) \text{ when } y=0.$$

Under some assumptions, the limiting distribution of the pair $(N(t),Y(t))$ is given by the steady-state distribution $\pi^{c(\infty)}$ of a self-regulated gene with the above transition rates where $c(t)$ is replaced by $c(\infty)$. The related steady-state distribution can then be computed using (4). In this limiting case, there is no feedback since the feedback rate $g(n)$ is given by $c(\infty)$; the results of Kepler and Elston (2001) imply then that the related stochastic dynamics is monostable, i.e. the steady-state distribution is unimodal, at least in the large steady-state gene-product level.

The time-evolution of the network is modeled here using a semi-stochastic model. The various molecular interactions between, e.g. the repressor, its doxycycline ligands, etc. as well as the mathematical properties of the networks are given in the Supplementary Material. The process of interest is

$$\eta(t)=(N(t),\, Y(t),\, X(t),\, Z(t)),$$

where $N(t)$ and $X(t)$ denote the number of activator and therapeutical proteins present in the cell at time t, and where $Y(t),\, Z(t)=0,\, 1$ denote the state of the associated promoters. The therapeutic network contains one positive feedback loop, and transactivator proteins enhance the transcription of the therapeutic gene, see Figure 3; these two reactions are here considered using two functions $c(t)$ and $\hat{c}(t)$, which satisfy time-delayed differential equations, see the Supplementary Material. As $t\to\infty$, assuming convergence, the limiting steady-state distribution is a product measure $\pi^{c(\infty)}\otimes\pi^{\hat{c}(\infty)}$, that is the probability of observing the network in state $(n,\, y,\, x,\, z)$ after a long time is given by $\pi_n^{c(\infty)}(y)\pi_x^{\hat{c}(\infty)}(z)$. Here again both $\pi^{c(\infty)}$ and $\pi^{\hat{c}(\infty)}$ are computed using (4). Notice that the interactions between the two genes i.e. between the two processes $(N(t),Y(t))$ and $(X(t),Z(t))$ are contained implicitly in the differential system related to $c(t)$ and $\hat{c}(t)$. Again, these steady-state distributions are related to self-regulated genes with no feedback, so that the related stochastic dynamics are monostable in the large steady-state gene-product level (Kepler and Elston, 2001).

4 EXPERIMENTAL RESULTS

The behavior of the genetic network was analyzed experimentally in living cells. Three plasmids were used, one encoding the constitutively expressed doxycycline-regulated transcriptional repressor, while the other two plasmids encode the conditionally expressed transcriptional activator or the reporter protein as shown in Figure 3 (Imhof *et al.*, 2000). The enhanced green fluorescent protein (EGFP) was chosen as a reporter because it can be easily detected by flow cytometry, which allows the analysis of its distribution in many individual cells within large populations.

The three plasmids were introduced in CHO cells by stable transfection together with a fourth plasmid encoding a protein conferring resistance to an antibiotic. After transfection, the cells were cultured in the presence of the antibiotic for 2 weeks, allowing the selection of cells having incorporated the plasmids stably in their own genome. Flow cytometric analysis and cell sorting was performed on the polyclonal populations to isolate cells displaying regulated expression of EGFP in response to the addition of doxycycline in the growth medium. A cell clone displaying regulated EGFP expression was selected for further characterization.

In each experiment, the level of EGFP fluorescence was quantified in 50 000 individual cells as a response to the addition of doxycycline in the growth medium, and steady-state

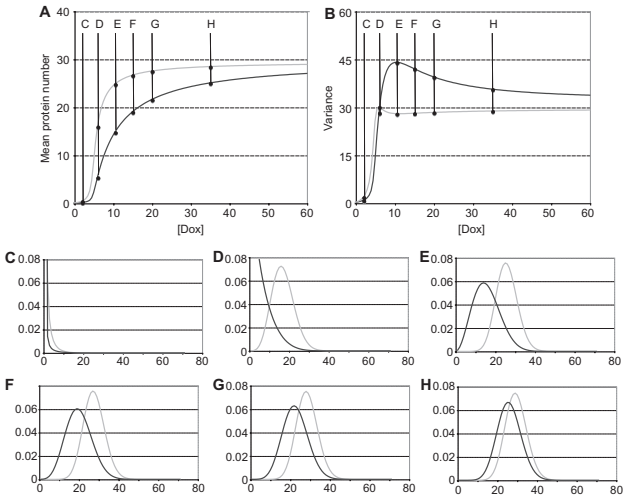

Fig. 5. Plots of the evolution of the steady-state means (**A**) and variances (**B**) related to transactivator (light) and therapeutic (dark) proteins, using a semi-stochastic model, as function of the number [Dox] of doxycycline molecules. (**C–H**) show the steady-state distributions in various regimes. One sees the switch between low (C-D), intermediate (E) and high (**G–H**) expression levels. Our simulations indicate that steady-state distributions are generically unimodal.

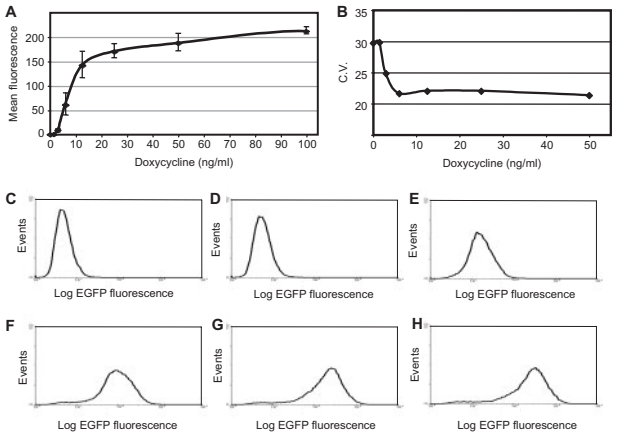

Fig. 6. Assay of the regulation of EGFP expression in live cells. A monoclonal population of CHO cells containing the EGFP gene expressed from the network stably integrated in its genome was treated with different concentrations of the Dox inducer. (**A**) Mean fluorescence and standard deviations of the means as a response to the concentration of dox were determined from three independent cell populations. (**B**) Coefficient of variation of the fuorescence levels were determined from individual cells within the same population, as shown in panels (**C–H**) as a function of the concentration of Dox inducer. (**C–H**) Steady-state E-GFP fluorescence profiles of cell populations treated for 24 h with increasing concentrations of the inducer.

conditions were insured by prior time course experiments which indicated that equilibrium was reached 24 h after the addition of doxycycline (data not shown). In the absence of the doxycycline inducer, where the repressor is most active, average fluorescence was low as expected, and cannot be distinguished from that of a population of cells that do not contain the EGFP gene (Figs 6A and C, and data not shown). Thus, this profile corresponds to the low levels of natural background fluorescence of CHO cells. Gene expression followed a sharp sigmoid dose-response typical of strongly cooperative systems and reached a maximum value at 100 ng/ml of dox (Fig. 6A and H).

Sigmoid induction curves resulting from cooperative genetic systems have been either associated with a bimodal response (also referred to as bistable, see Fig. 2) or with gradual or unimodal transcriptional activation (Figs 5 and 6). Upon addition of increasing amounts of the inducer, a gradual increase of fluorescence was observed, leading to a mostly unimodal transition mode (Fig. 6C–H), in agreement with the probabilistic modeling of the behavior of this genetic network. Mathematically, one can obtain bistable behaviors (see, e.g. Fig. 2), but our simulations indicate that the semi-stochastic model yield generically unimodal steady-state distributions, in accordance with the experimental results.

5 DISCUSSION

Gene expression noise has been associated to stochastic fluctuations of transcription or translation in various cellular systems. For instance, the gene expression noise strength has been found to remain essentially unaffected upon transcriptional induction in bacterial cells (Guido *et al*., 2006; Ozbudak *et al*., 2002) while it shows a bell-curve behavior in yeast cells (Blake and Collins 2005). This has led to the proposal that noise may be intrinsically different in prokaryotes and eukaryotes, and that a noisy behavior may result from the multiple rounds of transcriptional initiation in eukaryotic cells (Blake *et al*., 2003; Blake and Collins, 2005 or Raj *et al*., 2006). However, in many cases the molecular causes of the observed stochastic behavior remain poorly understood. In the case of mammalian cells, this may be a consequence of chromatin structure and/or promoter switching, leading to stochastic transcription (Chubb *et al*., 2006; Raj *et al*.,2006; Sato *et al*., 2004; Weinberger *et al*., 2005). However, other processes may also behave in a random fashion, including alternative splicing, mRNA transport or translation. Here, we develop a model where only the first probabilistic step is taken into consideration. This stems from the fact that the regulatory network studied here is controlled by doxycycline, which acts to regulate promoter switching rates and to mediate the synthesis of an activator of limiting concentration, which is a known cause of stochastic transcription in mammalian cells (Weinberger *et al*., 2005). The good correlation between experimental and computational results provides support to the validity of our model and of these assumptions.

Sotiropoulos and Kaznessis (2007) have considered other types of networks architectures involving the stochastic transcription of tetracycline-regulated genes. Their work focused on the time evolution of the amount of gene product using stochastic simulations, while the semi-stochastic model developed here provides analytical expressions for the mean and variances of the number of transactivator and therapeutical proteins at steady-state. These and other approaches may form the basis of unified stochastic models towards the exact descriptions of the kinetic and steady-state properties of regulatory gene networks.

Here, we find that the noise shows a peak or plateau at low [Dox] concentration followed by a decrease upon addition of the inducer when represented as the CV of the fluorescence of individual cells in a monoclonal population (Fig. 6B), which is reminiscent of Figure 5. This behavior of mammalian cells thus contrasts that of unicellular eukaryotes (Blake *et al*., 2006). Because yeast and mammalian cells share multiple rounds of initiation and pulsatile transcription (Raj *et al*., 2006), our results indicate that other differences must contribute to distinct noise propagation in these cell types.

A notable difference between the yeast and mammalian cell systems is that a small proportion of the mammalian cells never induce transgene expression, even after prolonged treatment with the inducer, which is suggestive of a bimodal distribution at the steady-state (Fig. 6G and H and data not shown). This behavior of the mammalian cells may be attributed to the silencing of transgenes integrated at a chromosomal locus that adopts a non-permissive chromatin structure (Sato *et al*., 2004). Indeed, coordinate silencing of numerous transgene copies integrated at one chromosomal locus has been implicated in a noisy gene expression pattern in cultured mammalian cells (Raj *et al*., 2006), and mutations that inhibit chromatin remodeling result in increased noise and in heterogeneous expression in yeast (Raser and O'Shea, 2004).

Work performed in bacteria or yeast have revealed that stable bimodal distribution of individual cells can be attributed either to the noisy expression of the regulatory proteins that bind the promoter under study (Blake and Collins, 2005; Blake *et al*., 2006), to slow stochastic transitions rates of the promoter between the inactive and active states (Raser and O'Shea, 2004), or to positive feed-back loops in genetic networks (Kaern *et al*., 2005). Thus, the network studied here displays several of the properties shown to provoke a bimodal distribution of the reporter protein concentration in yeast or bacterial cells (Becskei *et al*., 2001; Isaacs *et al*., 2003). In yeast, an autoactivated gene behaves as a bistable system where cells can only oscillate between an expressing and a non-expressing states and where the inducer acts to change the distributions of the cells in the two subpopulations (Becskei *et al*., 2001). Thus, a bistable behavior may be intuitively expected for the genetic network studied here in mammalian cells. However, we find that this architecture displays an essentially unimodal and gradual increase in expression in response to varying doses of the inducer (Fig. 6C–H), in agreement with the probabilistic modeling.

Yeast and mammalian cells share many transcription factors and chromatin constituents. Thus, differences in the behavior of genetic systems must result from the fine-tuning and interplay of these elements, such as the relative importance of chromatin remodeling versus promoter binding by activators or repressors, and the regulation of initiation complex assembly. The essentially unimodal distribution observed here may be linked to the complex structure of mammalian chromatin, and to its prominent role in the silencing of chromosomal genes in mammalian cells. The many steps and relatively slow transitions between the non-permissive and the permissive states of chromatin in mammalian cells may dampen the noise that stems from the stochastic binding of a low number of activator proteins to the promoter and from noise amplification resulting from the gene auto-activation feedback. In this context, chromatin may act as a noise-filtering device that allows graded response from stochastic events. A semi-stochastic model where the feedback rates $g(N(t))$ are replaced by $g(\mathbb{E}(N(t-\theta))$ as used here is well adapted to this setting. Overall, our results indicate that a noisy behavior and a strong autocatalytic feedback do not necessarily lead to a bistable behavior and to stochastic transitions between the two stable states (Fig. 2), but they rather suggest the conclusion that propagation of noise differs between different cell types and genetic systems.

ACKNOWLEDGEMENTS

The authors would like to thank the anonymous reviewers for their valuable and critical remarks. Experimental work was supported by the Swiss Commission for Technology and Innovation, Selexis SA, the Swiss Foundation for research on muscular dystrophies and by the Etat de Vaud.

Conflict of Interest: none declared.

REFERENCES

Adalsteinsson,D. *et al*. (2004) Biochemical Network Stochastic Simulator (BioNetS): software for stochastic modeling of biochemical networks. *BMC Bioinformatics*, **5**, 24.

Arkin,A. *et al*. (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected escherichia coli cells. *Genetics*, **149**, 1633–1648.

Austin,D.W. *et al*. (2005) Gene network shaping of inherent noise spectra. *Nature*, **439**, 608–611.

Becskei,A. *et al*. (2001) Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J*., **20**, 2528–2535.

Blake,W.J. and Collins,J.J. (2005) And the noise played on: stochastic gene expression and HIV-1 infection. *Cell*, **122**, 147–149.

Blake,W.J. *et al*. (2003) Noise in eukaryotic gene expression. *Nature*, **422**, 633–637.

Blake,W.J. *et al*. (2006) Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell*, **24**, 853–865.

Bolthausen,E. and Goldsheid,I. (2000) Recurrence and transience of random walks in random environments on a strip. *Commun. Math. Phys*., **214**, 429–447.

Bratsun,D. *et al*. (2005) Delay-induced stochastic oscillations in gene regulation. *Proc. Natl. Acad. Sci. USA*, **102**, 14593–14598. Comparative Study.

Burrage,K. *et al*. (2004) A multi-scaled approach for simulating chemical reaction systems. *Prog. Biophys. Mol. Biol*., **85**, 217–234. Comparative Study.

Cao,Y. *et al*. (2005a) Multiscale stochastic simulation algorithm with stochastic partial equilibrium assumption for chemically reacting systems. *J. Comput. Phys*., **206**, 395–411.

Cao,Y. *et al*. (2005b) The slow-scale stochastic simulation algorithm. *J. Chem. Phys*., **122**, 14116.

Chatterjee *et al*. (2007) An overview of spatial microscopic and accelerated kinetic monte carlo methods. *J. Comput. Aid. Mater. Des*., **14**, 253–308.

Chubb,J.R. *et al*. (2006) Transcriptional pulsing of a developmental gene. *Curr. Biol*., **16**, 1018–1025.

El Samad,H. *et al*. (2005) Stochastic modelling of gene regulatory networks. *Int. J. Robust Nonlinear Control*, **15**, 691–711.

Erban,R. *et al*. (2006) Gene regulatory networks: a coarse-grained, equation-free approach to multiscale computation. *J. Chem. Phys*., **124**, 084106.

Gillespie,D.T. (1977) Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem*., **81**, 2340–2361.

Gillespie,D.T. (2001) Approximate accelerated stochastic simulation of chemically reacting systems. *J. Chem. Phys*., **115**, 1716–1733.

Goutsias,J. (2005) Quasiequilibrium approximation of fast reaction kinetics in stochastic biochemical systems. *J. Chem. Phys*., **122**, 184102.

Goutsias,J. and Kim,S. (2006) Stochastic transcriptional regulatory systems with time delays: a mean-field approximation. *J. Comput. Biol*., **13**, 1049–1076.

Guido,N.J. *et al*. (2006) A bottom-up approach to gene regulation. *Nature*, **439**, 856–860.

Hornos,J.E.M. *et al.* (2005) Self-regulating gene: an exact solution. *Phys. Rev. E Stat. Nonlin. Soft. Matter. Phys.*, **72** (5 Pt 1), 051907.
Imhof,M.O. *et al.* (2000) A regulatory network for the efficient control of transgene expression. *J. Gene Med.*, **2**, 107–116.
Isaacs,F.J. *et al.* (2003) Prediction and measurement of an autoregulatory genetic module. *Proc. Natl. Acad. Sci. USA*, **100**, 7714–7719.
Kaern,M. *et al.* (2005) Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.*, **6**, 451–464.
Kepler,T.B. and Elston,T.C. (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, **81**, 3116–3136.
Mcadams,H. H. and Arkin,A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. USA*, **94**, 814–819.
Ozbudak,E.M. *et al.* (2002) Regulation of noise in the expression of a single gene. *Nat. Genet.*, **31**, 69–73.
Paulsson,J. (2004) Summing up the noise in gene networks. *Nature*, **427**, 415–418.
Paulsson,J. (2005) Models of stochastic gene expression. *Phys. Life Rev.*, **2**, 157–175.
Pedraza,J.M. and van Oudenaarden,A. (2005) Noise propagation in gene networks. *Science*, **307**, 1965–1969.
Raj,A. *et al.* (2006) Stochastic mrna synthesis in mammalian cells. *PLoS Biol.*, **4**, 1707–1719.
Raser,J.M. and O'Shea,E.K. (2004) Control of stochasticity in eukaryotic gene expression. *Science*, **304**, 1811–1814.
Salis,H. and Kaznessis,Y. (2005) Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.*, **122**, 054103.
Salis,H. *et al.* (2006) Multiscale hy3s: Hybrid stochastic simulation for supercomputers. *BMC Bioinformatics*, **7**, 93.
Sato,N. *et al.* (2004) Fluctuation of chromatin unfolding associated with variation in the level of gene expression. *Genes Cells*, **9**, 619–630.
Shibata,T. (2003a) Fluctuating reaction rates and their application to problems of gene expression. *Phys. Rev. E*, **67**, 061906.
Shibata,T. (2003b) Reducing the master equations for noisy chemical reactions. *The J. Chem. Phys.*, **119**, 6629–6634.
Sotiropoulos,V. and Kaznessis,Y.N. (2007) Synthetic tetracycline-inducible regulatory networks: computer-aided design of dynamic phenotypes. *BMC Syst. Biol.*, **1**, 7.
Weinberger,L.S. *et al.* (2005) Stochastic gene expression in a lentiviral positive-feedback loop: Hiv-1 tat fluctuations drive phenotypic diversity. *Cell*, **122**, 169–182.

APPENDIX: PROOF OF FORMULA (4)

Notice that $q_{n,0}=d_n$ and $q_{n,1}=c_n$. The proof of (2.4) is based on methods presented in Bolthausen and Goldsheid (2000), but here we solve explicitely the matrix valued continued fractions (3). We shall also need the following matrices, defined similarly at the boundaries $n=0$ and $n=\Lambda$:

$$\hat{R}_0=\begin{pmatrix}0 & 1\\ \frac{\kappa(0)}{\kappa(0)+\mu} & 0\end{pmatrix},\quad Q_1=\begin{pmatrix}\frac{\nu(1)}{d_1} & 0\\ 0 & \frac{\nu(1)}{c_1}\end{pmatrix},$$

$$P_0=\begin{pmatrix}0 & 0\\ 0 & \frac{\mu}{c_0}\end{pmatrix},\quad \hat{R}_\Lambda=\begin{pmatrix}0 & \frac{g(\Lambda)}{\nu(\Lambda)+g(\Lambda)}\\ \frac{\kappa(\Lambda)}{\kappa(\Lambda)+\nu(\Lambda)} & 0\end{pmatrix},$$

$$\hat{Q}_\Lambda=\begin{pmatrix}\frac{\nu(\Lambda)}{\nu(\Lambda)+g(\Lambda)} & 0\\ 0 & \frac{\nu(\Lambda)}{\nu(\Lambda)+\kappa(\Lambda)}\end{pmatrix}.$$

Notice that the matrices $\hat{R}_0+P_0$ and $\hat{R}_\Lambda+\hat{Q}_\Lambda$ are stochastic. Let $\tilde{\pi}=(\tilde{\pi}_n)_{0\le n\le\Lambda}$ be the steady-state, where we recall that $\tilde{\pi}_n$ is the row vector $\tilde{\pi}_n=(\tilde{\pi}_n(0),\tilde{\pi}_n(1))$.

Of course, this definition is inductive, and the inverse of $\mathrm{id}-R_n-\alpha_{n-1}P_{n-1}$ is not necessarily defined. We shall see in the sequel that these matrices are well defined; this implies that $\alpha_n=Q_{n+1}\gamma_{n-1}$. To start the induction, one needs to define α_0: this is obtained by using the invariance of $\tilde{\pi}$ at the left boundary $n=0$. This yields $\tilde{\pi}_0=\tilde{\pi}_0\hat{R}_0+\tilde{\pi}_1Q_1$, that is to $\tilde{\pi}_0=\tilde{\pi}_1Q_1(\mathrm{id}-\hat{R}_0)^{-1}$. One gets thus the natural candidate $\alpha_0:=Q_1(\mathrm{id}-\hat{R}_0)^{-1}$.

Lemma: Suppose that the matrices α_n and γ_n are well defined. Then the matrices $\theta_{n+1}=\gamma_nP_{n+1}$, $0\le n<\Lambda-1$, are stochastic.

Proof. The proof proceeds by induction. When $n=0$, we first check that $\theta_1=\gamma_0\ P_1$ is stochastic, with $\gamma_0^{-1}=\mathrm{id}-R_1-\alpha_0P_0$. Then $\theta_1\ \mathbf{1}=\mathbf{1}$ is equivalent to $(P_1+R_1)\mathbf{1}=\mathbf{1}-\alpha_0\ P_0\ \mathbf{1}$, and therefore to $0=Q_1\mathbf{1}-\alpha_0P_0\mathbf{1}=Q_1(\mathbf{1}-(\mathrm{id}-\hat{R}_0)^{-1}P_0\mathbf{1})$, where we use the fact that $P_1+Q_1+R_1$ is stochastic. But Q_1 is diagonal, and the last statement is equivalent to $((\mathrm{id}-\hat{R}_0)\mathbf{1}=P_0\mathbf{1}$, which holds true since $\hat{R}_0+P_0$ is stochastic. The induction step is obtained in the same way.

Suppose that we have already obtained the sequence α_k, $k=0,\ldots,\ n-1$. Then we can obtain α_n by setting $\alpha_n=Q_{n+1}\gamma_{n-1}$. The above Lemma gives that $\gamma_{n-1}\ P_n$ is stochastic. But, taking into account the specific form of the involved matrices, one can write

$$\gamma_{n-1}P_n=\begin{pmatrix}0 & \frac{\mu}{c_n}\gamma_{n-1}(1,2)\\ 0 & \frac{\mu}{c_n}\gamma_{n-1}(2,2)\end{pmatrix},$$

where $\gamma_{n-1}(i,\,j)$ denotes the $(i,\,j)$ entry of the matrix γ_{n-1}. Thus $\gamma_{n-1}(1,2)=\gamma_{n-1}(2,2)=\frac{c_n}{\mu}$. Finally, $\alpha_n=Q_{n+1}\ \gamma_{n-1}$ shows that $\alpha_n(1,2)=\frac{\nu(n+1)c_n}{d_{n+1}\mu}$, and $\alpha_n(2,2)=\frac{\nu(n+1)c_n}{\mu c_{n+1}}$. Let $A_{n-1}=\gamma_{n-1}^{-1}$, with

$$A_{n-1}=\begin{pmatrix}1 & -\frac{g(n)}{d_n}-\frac{\mu\alpha_{n-1}(1,2)}{c_{n-1}}\\ -\frac{\kappa(n)}{c_n} & 1-\frac{\mu\alpha_{n-1}(2,2)}{c_{n-1}}\end{pmatrix},$$

of determinant $|A_{n-1}|$ given by $|A_{n-1}|=\frac{\mu}{c_n}$. It follows that

$$\alpha_n=\frac{c_n}{\mu}\begin{pmatrix}\frac{\nu(n+1)}{d_{n+1}} & 0\\ 0 & \frac{\nu(n+1)}{c_{n+1}}\end{pmatrix}\begin{pmatrix}\frac{\kappa(n)+\mu}{c_n} & 1\\ \frac{\kappa(n)}{c_n} & 1\end{pmatrix}$$

$$=\frac{c_n}{\mu}\begin{pmatrix}\frac{(\kappa(n)+\mu)\nu(n+1)}{c_nd_{n+1}} & \frac{\nu(n+1)}{d_{n+1}}\\ \frac{\kappa(n)\nu(n+1)}{c_nc_{n+1}} & \frac{\nu(n+1)}{c_{n+1}}\end{pmatrix},$$

as required.