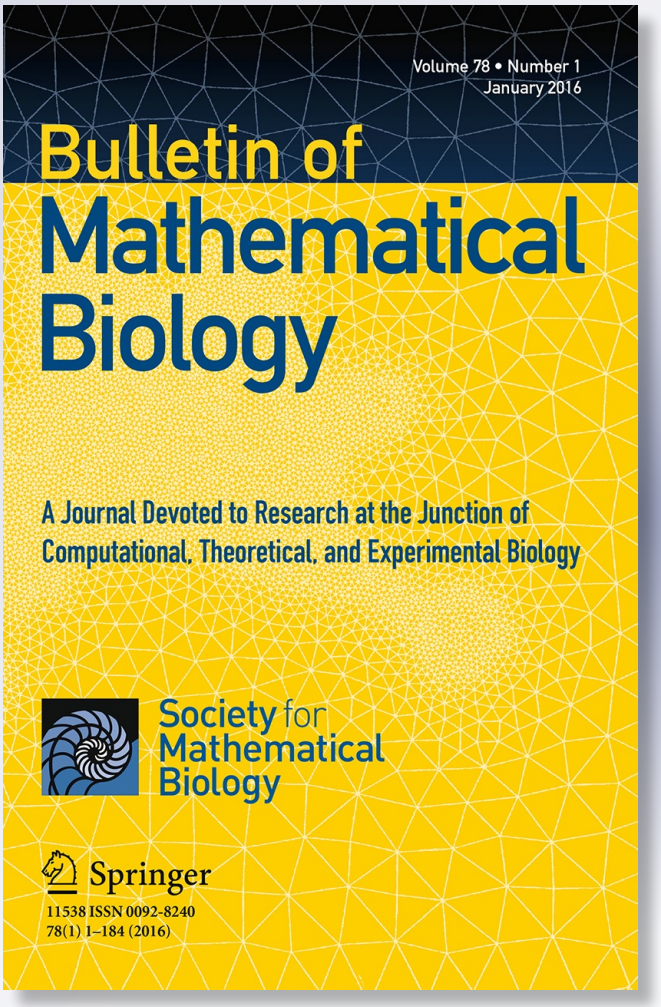
Protein Synthesis Driven by Dynamical Stochastic Transcription

Guilherme C. P. Innocentini, Michael Forger, Ovidiu Radulescu & Fernando Antoneli

Bulletin of Mathematical Biology
A Journal Devoted to Research at the Junction of Computational, Theoretical and Experimental Biology Official Journal of The Society for Mathematical Biology

ISSN 0092-8240
Volume 78
Number 1

Bull Math Biol (2016) 78:110-131
DOI 10.1007/s11538-015-0131-3



Springer

Bull Math Biol (2016) 78:110–131
DOI 10.1007/s11538-015-0131-3

Society for Mathematical Biology

ORIGINAL ARTICLE

Protein Synthesis Driven by Dynamical Stochastic Transcription

Guilherme C. P. Innocentini[1,2] · **Michael Forger**[1] · **Ovidiu Radulescu**[2] · **Fernando Antoneli**[3]

Received: 12 November 2014 / Accepted: 18 November 2015 / Published online: 15 December 2015

Abstract In this manuscript, we propose a mathematical framework to couple transcription and translation in which mRNA production is described by a set of master equations, while the dynamics of protein density is governed by a random differential equation. The coupling between the two processes is given by a stochastic perturbation whose statistics satisfies the master equations. In this approach, from the knowledge of the analytical time-dependent distribution of mRNA number, we are able to calculate the dynamics of the probability density of the protein population.

Keywords Gene expression · Stochasticity · Exact solutions · Dynamics

Mathematics Subject Classification 92B05

✉ Guilherme C. P. Innocentini
ginnocentini@gmail.com

Michael Forger
forger@ime.usp.br

Ovidiu Radulescu
ovidiu.radulescu@univ-montp2.fr

Fernando Antoneli
fernando.antoneli@unifesp.br

1 Departamento de Matemática Aplicada, Instituto de Matemática e Estatística, Universidade de São Paulo, Rua do Matão, 1010, Cidade Universitária, São Paulo, SP 05508-090, Brazil

2 DIMNP, UMR 5235, Université de Montpellier 2, Pl. E. Bataillon, Bat. 24, 34095 Montpellier Cedex 5, France

3 Laboratório de Genômica Evolutiva e Biocomplexidade & DIS, Escola Paulista de Medicina, Universidade Federal de São Paulo, Rua Pedro de Toledo, 669, 4th floor, São Paulo, SP 04039-032, Brazil

1 Introduction

Stochasticity in biological processes, in particular of gene expression, has been studied, both experimentally and theoretically, at least since the pioneering work of Delbrück (1940). Recent advances in experimental methods have enabled direct observation of stochastic features of gene expression, such as temporal fluctuations in individual cells or steady-state variations across a cell population (Elowitz et al. 2002; Ozbudak et al. 2002; Blake et al. 2003; Raser and O'Shea 2004; Golding et al. 2005; Cai et al. 2006; Yu et al. 2006), and data acquisition has experienced a huge improvement in the last decade. However, theoretical models have not yet been developed to the point of providing a comprehensive quantitative description for the dynamics of gene expression. The stationary regime has been exhaustively discussed in the literature, but studies on time-dependent probability distributions are still scarce (Iyer-Biswas et al. 2009; Shahrezaei and Swain 2008; Ramos et al. 2011).

In this paper, our main goal is to present and discuss a stochastic description for mRNA-protein dynamics. More precisely, we propose and solve a hybrid model for stochastic gene expression, consisting of a master equation (ME) coupled to a random differential equation (RDE). The ME describes the production of messenger RNA (mRNA) molecules triggered by a gene with various levels of promoter activity. The RDE governs the dynamics of protein synthesis: it is a linear ordinary differential equation randomly perturbed by the Markov jump process underlying the ME. The ME part of the model is a particular case of a Markov process in a "random environment" (Cogburn and Torrez 1981), composed by a birth-and-death process and a two-state Markovian switching process, in continuous time; see Peccoud and Ycart (1995) for the interpretation in the context of gene expression. Several variations of this type of model have been employed for the study of gene expression and have been extensively discussed in the literature (Kepler and Elston 2001; Pirone and Elston 2004; Hornos et al. 2005; Paulsson 2005). The particular form of the ME part used in this paper is the one analyzed in Innocentini and Hornos (2007), Innocentini et al. (2013).

The motivations for such an approach can be justified on mathematical as well as biological grounds. From a mathematical point of view, the RDE employed here resembles a Langevin equation, with one crucial difference: the driving stochastic process is not a singular *delta*-like noise, but rather a non-singular, well-behaved stationary stochastic process. Non-white noise-driven Langevin-like equations have been widely discussed in the literature under different names, such as *colored noise* (Kampen 2007) or *real noise* (Arnold 1998). And the mathematical advantage in dealing with RDEs is that one does not need a sophisticated theory of integration in order to solve them. As a matter of fact, RDEs are solved by Riemann integration of ordinary differential equations, sample path by sample path—hence the term "random differential equation" instead of the more familiar term "stochastic differential equation," which is reserved for differential equations associated with a stochastic integration theory (Arnold 1998).

Besides the mathematical benefit, there is a biological motivation in modeling mRNA transcription by a ME and protein synthesis by a RDE, thus supposing that the transcription product should be treated as a discrete random variable (number of

mRNA molecules), while the translation product should be treated as a continuous random variable (density of protein molecules). The reason behind this distinction is the large gap, typically of several orders of magnitude, between mRNA numbers and protein numbers in the cell. The hybrid model we propose here attempts to incorporate the discrepancy between mRNA and protein molecules (which concerns not only their typical numbers but also their typical lifetimes) from the very beginning, instead of assuming that it can be ignored. That is why, in conformity with procedures already adopted implicitly in some of the literature but rarely spelled out [one exception is Friedman et al. (2006)], we suggest to model protein number by a continuous probability density rather than a discrete probability distribution.

Admittedly, this amounts to a change of paradigm, but as will be shown here, the resulting simplifications are so substantial that they allow us to solve the resulting model without constraints on the values of the parameters. Furthermore, this approach allows us to evaluate probability densities even for very high protein numbers, with no extra effort.

2 Model for Transcription and Translation

Let us describe our model in more detail. Gene transcription is described by a pair of master equations, corresponding to two states $\{1, 2\}$ of promoter activity, for a birth-and-death process coupled by a telegraph-like process encoding the switch between promoter states (generalization to a higher number of promoter states will be left to future work):

$$\begin{aligned}\frac{\mathrm{d}\phi_n^1}{\mathrm{d}t} &= k_1[\phi_{n-1}^1 - \phi_n^1] + \rho[(n+1)\phi_{n+1}^1 - n\phi_n^1] - h\phi_n^1 + f\phi_n^2,\\ \frac{\mathrm{d}\phi_n^2}{\mathrm{d}t} &= k_2[\phi_{n-1}^2 - \phi_n^2] + \rho[(n+1)\phi_{n+1}^2 - n\phi_n^2] + h\phi_n^1 - f\phi_n^2.\end{aligned} \tag{1}$$

The discrete random variable n stands for the number of mRNA molecules in the cell and $\phi_n^j(t)$ is the probability for finding the gene in state number j ($j = 1$ or 2) with n mRNA molecules in the cell, at time t; the resulting total probability will be denoted by $\phi_n(t) = \phi_n^1(t) + \phi_n^2(t)$. Production of mRNA is controlled by the rates k_1 and k_2, while its degradation is taken into account by the rate ρ which is independent of the activity level of the promoter. The switch between the two states is controlled by the rates h and f. The master equations (1) are mathematically equivalent (up to a change of notation) to the master equations (4)–(5) of Kepler and Elston (2001), since they describe the same underlying reactions (birth, death and switching between the two states). However, the discrete random variable in Kepler and Elston (2001) stands for protein number, so the state variables and their interpretation are different. Protein synthesis/degradation is governed by an RDE of the form

$$\frac{\mathrm{d}}{\mathrm{d}t}m_t = -Am_t + Bn_t, \tag{2}$$

where m is a continuous random variable representing the protein number density in the cell, A and B are the protein degradation and synthesis rates, respectively, and n is as before, but now with time dependence following a stochastic Markov jump process where $n_{t+\Delta t} = n_t \pm 1$ with probability $(k_1 + k_2)\Delta t$ for $+1$ and $\rho n_t \Delta t$ for -1 (and $n_{t+\Delta t} = n_t$ with remaining probability): this is consistent with the time evolution of the total probability distribution ϕ_n that follows from Eq. (1). With the assumption that A and B are constant, our model focuses on the effects of the stochasticity of the transcription process and neglects the protein production/decay noise.

3 Solutions of the Model

A complete description of n_t is achieved by obtaining the time-dependent solutions of the master equations (1), and this is what we do in the following. However, before dealing with the master equations, let us first redefine the parameter space and introduce the biological quantities of the model, as in Innocentini and Hornos (2007), namely: the efficiency parameters $N_1 = k_1/\rho$ and $N_2 = k_2/\rho$, the switching parameter $\epsilon = (h + f)/\rho$ and the occupancy probabilities $p_1 = f/(h + f)$ and $p_2 = h/(h + f)$. Using the generating function technique (Kampen 2007), the coupled master equations are transformed into a set of PDEs (partial differential equations) for the functions $\phi^1(z,t) = \sum_{n=0}^{\infty} \phi_n^1(t) z^n$ and $\phi^2(z,t) = \sum_{n=0}^{\infty} \phi_n^2(t) z^n$:

$$\begin{aligned} \frac{1}{\rho}\frac{\partial \phi^1}{\partial t} &= (z-1)\left[N_1 \phi^1 - \frac{\partial \phi^1}{\partial z}\right] - \epsilon p_2 \phi^1 + \epsilon p_1 \phi^2, \\ \frac{1}{\rho}\frac{\partial \phi^2}{\partial t} &= (z-1)\left[N_2 \phi^2 - \frac{\partial \phi^2}{\partial z}\right] + \epsilon p_2 \phi^1 - \epsilon p_1 \phi^2. \end{aligned} \tag{3}$$

The probability distributions are obtained from the generating functions using

$$\begin{aligned} \phi_n^1(t) &= \frac{1}{n!}\frac{\partial^n \phi^1(z,t)}{\partial z^n}\bigg|_{z=0}, \\ \phi_n^2(t) &= \frac{1}{n!}\frac{\partial^n \phi^1(z,t)}{\partial z^n}\bigg|_{z=0}, \end{aligned} \tag{4}$$

Introducing a new set of variables through the transformations $\mu = (z-1)\mathrm{e}^{-\rho t}$ and $\nu = z - 1$, Eq. (3) assumes the form

$$\begin{aligned} -\nu\frac{\partial \phi^1}{\partial \nu} + \nu N_1 \phi^1 - \epsilon p_2 \phi^1 + \epsilon p_1 \phi^2 = 0, \\ -\nu\frac{\partial \phi^2}{\partial \nu} + \nu N_2 \phi^2 + \epsilon p_2 \phi^1 - \epsilon p_1 \phi^2 = 0, \end{aligned} \tag{5}$$

i.e., this transformation reduces the original set of PDEs to a set of ODEs (ordinary differential equations), which have already been solved in Innocentini and Hornos (2007); a similar transformation with the same purpose has been used in Ramos et al. (2011),

Iyer-Biswas et al. (2009). Following Innocentini and Hornos (2007), the solutions of Eq. (5) are:

$$\phi^1(\mu,\nu) = F(\mu)\, p_1 \mathrm{e}^{N_1\nu}\, \mathrm{M}(a, b+1, \eta) - G(\mu)(1-b)\eta^{-b}\mathrm{e}^{N_1\nu}\, \mathrm{M}(a-b, 1-b, \eta), \tag{6a}$$

$$\phi^2(\mu,\nu) = F(\mu)\, p_2 \mathrm{e}^{N_1\nu}\, \mathrm{M}(a+1, b+1, \eta) + G(\mu)(1-b)\eta^{-b}\mathrm{e}^{N_1\nu}\, \mathrm{M}(1+a-b, 1-b, \eta), \tag{6b}$$

where F and G are arbitrary functions that must be determined from the initial conditions, where we note that $t = 0$ corresponds to $\nu = \mu$. The symbol M stands for the Kummer M function (Abramowitz and Stegun 1964) with parameters $a = \epsilon p_2$, $b = \epsilon$ and $\eta = (N_2 - N_1)\nu$.

In order to determine F and G we will use matrix and vector notation to rewrite the solutions of Eq. (6) as $\vec{\phi}(\mu,\nu) = U(\nu)\vec{F}(\mu)$, where $\vec{\phi} = (\phi^1, \phi^2)^{\mathrm{T}}$ and $\vec{F} = (F, G)^{\mathrm{T}}$ (where $.^{\mathrm{T}}$ means matrix transposition); then the entries of the matrix $U(\nu)$ are

$$\begin{aligned} U_{1,1} &= p_1\, \mathrm{e}^{N_1\nu}\, \mathrm{M}(a, b+1, \eta), \\ U_{1,2} &= -(1-b)\eta^{-b}\, \mathrm{e}^{N_1\nu}\, \mathrm{M}(a-b, 1-b, \eta), \\ U_{2,1} &= p_2\, \mathrm{e}^{N_1\nu}\, \mathrm{M}(a+1, b+1, \eta), \\ U_{2,2} &= (1-b)\eta^{-b}\, \mathrm{e}^{N_1\nu}\, \mathrm{M}(1+a-b, 1-b, \eta). \end{aligned} \tag{7}$$

Inverting the relation $\vec{\phi}(\mu,\nu) = U(\nu)\vec{F}(\mu)$ gives $\vec{F}(\mu) = U(\nu)^{-1}\vec{\phi}(\mu,\nu)$, and setting $\nu = \mu$, we obtain an expression for $\vec{F}(\mu)$ in terms of the initial conditions. Thus, we have to compute the inverse of the matrix $U(\nu)$, which requires calculating its determinant. At a first glance, it might appear difficult to find a compact formula for that, since it involves products of Kummer functions. Fortunately, the well-known relations for Kummer functions, especially the one concerning the Wronskian [relations 13.1.20 in Abramowitz and Stegun (1964)], allow us to obtain a simple expression for this determinant:

$$\det(U(\nu)) = \frac{\mathrm{e}^{-(N_1+N_2)\nu}\eta^{\epsilon}}{1-\epsilon}. \tag{8}$$

Putting everything together, we obtain the time-dependent probability distributions that solve Eq. (1) and will serve as input to solve Eq. (2).

Considering any given perturbation n_t as input, the ODE (2) governing the protein dynamics is easily solved by applying the standard integral formula from the theory of ODEs. Introducing the dimensionless parameters $\tau = \rho\, t$, $\alpha = A/\rho$ and $\beta = B/\rho$, the solution reads

$$m_\tau = m_0\, \mathrm{e}^{-\alpha\tau} + \beta\, \mathrm{e}^{-\alpha\tau} \int_0^\tau n_{\tau'}\, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau', \tag{9}$$

where the integral is an ordinary Riemann integral (applied to the product of a step function by an exponential function) and $m_0 = m(0)$. In the present case, where both n_τ and m_τ are stochastic processes, we can interpret this formula as an operator

that maps the process n_τ (for mRNA number) to the process m_τ (for protein number density), sample by sample.

Recalling that the ultimate goal is to compute the probability density of the protein population, say $\mathcal{P}(\tau, m)$, the traditional method consists in randomly generating stochastic processes n_τ for mRNA number, applying the previous integral formula to produce corresponding stochastic processes m_τ for protein number density and looking at the resulting statistics. Here, and this is perhaps the central point of the present paper, we propose a different procedure: since the solution of Eq. (1) has already provided us with a probability distribution for mRNA number, it suffices to take its push-forward, in the sense of measure theory, under the operator defined by solving Eq. (9) to directly obtain the corresponding probability distribution for protein number density, without having to resort to random process generation. To describe how to compute the push-forward, let us consider the integral on the rhs of Eq. (9). Dividing the interval $[0, \tau]$ in p subintervals, we have:

$$\int_0^\tau n_{\tau'} \, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau' = \sum_{q=0}^{p-1} \int_{\tau_q}^{\tau_{q+1}} n_{\tau'} \, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau', \tag{10}$$

where $\tau_0 = 0$ and $\tau_p = \tau$. If the partition is sufficiently fine (i.e., for p sufficiently large), the function n_τ will be constant on each subinterval and the integral can be performed explicitly:

$$m_\tau = m_0 \, \mathrm{e}^{-\alpha\tau} + \frac{\beta}{\alpha} \mathrm{e}^{-\alpha\tau} \sum_{q=0}^{p-1} n_{\tau_q} (\mathrm{e}^{\alpha\tau_{q+1}} - \mathrm{e}^{\alpha\tau_q}). \tag{11}$$

Otherwise, i.e., for smaller values of p, Eq. (11) provides only a "rectangular" or "piecewise constant" approximation of the integral in Eq. (10) since it amounts to replacing, on each of the subintervals $[\tau_q, \tau_{q+1}]$, the step function $n_{\tau'}$ by a constant (here chosen to be its value at the left endpoint):

$$\int_{\tau_q}^{\tau_{q+1}} n_{\tau'} \, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau' \approx n_{\tau_q} \int_{\tau_q}^{\tau_{q+1}} \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau' = \left. \frac{n_{\tau_q}}{\alpha} \right|_{\tau_q}^{\tau_{q+1}}. \tag{12}$$

Of course, a "trapezoidal" or "piecewise linear" approximation is more precise: it consists in replacing this expression by

$$\int_{\tau_q}^{\tau_{q+1}} n_{\tau'} \, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau' \approx \int_{\tau_q}^{\tau_{q+1}} (a\tau' + b) \, \mathrm{e}^{\alpha\tau'} \mathrm{d}\tau' = \frac{1}{\alpha} \left(a\tau' + b - \frac{a}{\alpha} \right) \mathrm{e}^{\alpha\tau'} \Big|_{\tau_q}^{\tau_{q+1}}, \tag{13}$$

where a and b are determined by solving the equations $n_{\tau_q} = a\tau_q + b$ and $n_{\tau_{q+1}} = a\tau_{q+1} + b$.

In order to obtain a sample path for the process m_τ using these formulas, it suffices to represent a sample path for the process n_τ by the "shrunk" numerical sequence $(n_0, \ldots, n_{p-1})$, the only modification being that we must now allow consecutive

numbers to differ by more than ± 1. Finally, to make our sample space finite, we also introduce a cutoff L and impose that all n_q should be $\leqslant L$. For instance, by choosing L so large that the probability of $n_q > L$ is smaller than 10^{-20}, say, we can certainly neglect all values higher than L and restrict the set of possible values for n_q to the finite set $\{0, 1, \ldots, L-1, L\}$; then the space of sequences has $(L+1)^p$ elements.

Now, Eq. (11) provides a map from this space of sequences $(n_0, \ldots, n_{p-1})$ to that of numbers m_τ. Using this mapping, we define the push-forward probability on the set of possible values of m_τ by

$$\mathbf{P}\big(m_\tau = m_\tau(n_0, \ldots, n_{p-1})\big) = \Phi(n_0; \ldots; n_{p-1}), \tag{14}$$

where

$$\Phi(n_0; \ldots; n_{p-1}) = \Phi^1(n_0; \ldots; n_{p-1}) + \Phi^2(n_0; \ldots; n_{p-1}) \tag{15}$$

is the total joint probability distribution for finding n_q mRNA molecules at times $\tau_q (q = 0, \ldots, p-1)$, whereas $\Phi^1(n_0; \ldots; n_{p-1})$ and $\Phi^2(n_0; \ldots; n_{p-1})$ encode the joint probability distributions for finding n_q mRNA molecules at times τ_q $(q = 0, \ldots, p-1)$ with the gene in promoter state 1 and 2, respectively. In general, such joint probability distributions are difficult to obtain, but in our case, the mRNA process governed by the master equations (1) is Markovian and therefore we can compute the joint probabilities in terms of conditional probabilities, according to the iterated Chapman- Kolmogorov equation:

$$\begin{aligned}
\Phi^1(n_0; \ldots; n_{p-1}) = & \sum_{j_0, \ldots, j_{p-2}=1}^{2} \Phi(n_{p-1}, \tau_{p-1}, 1 | n_{p-2}, \tau_{p-2}, j_{p-2}) \ldots \\
& \ldots \Phi(n_1, \tau_1, j_1 | n_0, \tau_0, j_0)\, \phi_{n_0}^{j_0}(\tau_0), \\
\Phi^2(n_0; \ldots; n_{p-1}) = & \sum_{j_0, \ldots, j_{p-2}=1}^{2} \Phi(n_{p-1}, \tau_{p-1}, 2 | n_{p-2}, \tau_{p-2}, j_{p-2}) \ldots \\
& \ldots \Phi(n_1, \tau_1, j_1 | n_0, \tau_0, j_0)\, \phi_{n_0}^{j_0}(\tau_0),
\end{aligned} \tag{16}$$

where, as before, $\phi_{n_0}^{j_0}(\tau_0)$ is the probability to find the gene in the state j_0 and with n_0 mRNA molecules in the cell, at time τ_0. The quantity $\Phi(n_{q'}, \tau_{q'}, j' | n_q, \tau_q, j)$ is the conditional probability of finding $n_{q'}$ mRNA molecules at time $\tau_{q'}$ and with the gene in state j' provided there were n_q mRNA molecules at time τ_q and with the gene in state j, where $\tau_q < \tau_{q'}$, $q, q' = 1, \ldots, p-1$ and $j, j' = 1, 2$. These conditional probabilities can be obtained from the solutions of the master equations (6). To this end, one has to take as initial condition the generating function encoding the information that, at time τ_q, the system has exactly n_q particles and with probability 1 is in one of the two promoter states, say 1 or 2. Such a generating function has one component equal to 0, whereas the other is given by $(1+\mu)^{n_q}$, i.e.,

$$(\phi^1(\mu, \nu), \phi^2(\mu, \nu)) = ([1+\mu]^q, 0) \quad \text{with} \quad \nu = \mu \tag{17}$$

for promoter in state 1 and

$$(\phi^1(\mu,\nu),\phi^2(\mu,\nu)) = (0,[1+\mu]^q) \quad \text{with} \quad \nu=\mu \tag{18}$$

for promoter in state 2. In the (z,τ) variables, the non-vanishing component takes the form

$$\left(1+(z-1)\mathrm{e}^{-(\tau-\tau_q)}\right)^{n_q} \tag{19}$$

since here the initial time is τ_q, rather than 0.

Regarding the validity of Eq. (14), it is important to note that according to the general definition of the push-forward of probabilities, one should really take the sum of the probabilities corresponding to all sequences $(n_0,\ldots,n_{p-1})$ producing the same value of m_τ. However, Eq. (11) implies that, generically, any two different sequences will give different values (more precisely, this will be the case if the intermediate times $\tau_1,\ldots,\tau_{p-1}$ are chosen such that the differences of exponentials $\mathrm{e}^{\alpha\tau_{q+1}}-\mathrm{e}^{\alpha\tau_q}$, $q=0,\ldots,p-1$, are linearly independent over the integers).

For the sake of greater clarity, and to illustrate how the conditional probabilities are obtained from the explicit solution (6) of the master equations with the appropriate initial conditions (see Eqs. (17), (18) and (19) above), let us consider the simplest example: $p=2$ and $L=1$. Here, the sample space has four elements, namely, (0, 0), (0, 1), (1, 0) and (1, 1), and in general each of these sequences will produce a different number m_τ. Therefore, the probability assigned to each of these values m_τ is equal to the joint probability assigned to the corresponding sequence (n_0,n_1), summed over the two possible promoter states,

$$\mathbf{P}(m_\tau = m_\tau(n_0,n_1)) = \Phi^1(n_0;n_1)+\Phi^2(n_0;n_1). \tag{20}$$

Specializing Eq. (16) to the case $p=2$, we see that these joint probabilities are

$$\begin{aligned}\Phi^1(n_0;n_1) &= \Phi(n_1,\tau_1,1|n_0,\tau_0,1)\,\phi^1_{n_0}(\tau_0)+\Phi(n_1,\tau_1,1|n_0,\tau_0,2)\,\phi^2_{n_0}(\tau_0),\\ \Phi^2(n_0;n_1) &= \Phi(n_1,\tau_1,2|n_0,\tau_0,1)\,\phi^1_{n_0}(\tau_0)+\Phi(n_1,\tau_1,2|n_0,\tau_0,2)\,\phi^2_{n_0}(\tau_0),\end{aligned} \tag{21}$$

where, as before, the conditional probabilities $\Phi(n_1,\tau_1,j_1|n_0,\tau_0,j_0)$ take into account the promoter states. To exemplify how these are obtained from the solutions of the master equations, let us, by way of example, focus on the conditional probability $\Phi(n_1=5,\tau_1,j_1=1|n_0=10,\tau_0,j_0=1)$. This means that we are considering the situation where, at time τ_0, the system has 10 mRNA molecules and the promoter is found in the state 1, corresponding to the initial condition

$$(\phi^1(\mu,\nu),\phi^2(\mu,\nu)) = ([1+\mu]^{10},0) \quad \text{with} \quad \nu=\mu, \tag{22}$$

or in the (z,τ) variables,

$$(\phi^1(z,\tau_1),\phi^2(z,\tau_1)) = \left(\left[1+(z-1)\,\mathrm{e}^{-(\tau_1-\tau_0)}\right]^{10},0\right). \tag{23}$$

Using Eq. (22) to determine the vector

$$\vec{F}(\mu) = U(\nu)^{-1}\vec{\phi}(\mu, \nu)\Big|_{\nu=\mu}, \tag{24}$$

substituting the entries of this vector in Eq. (6a) for ϕ^1 and returning to the variables (z, τ), we arrive at the generating function, let's say $\Psi(z, \tau)$, of the conditional probabilities $\Phi(n_1, \tau_1, j_1 = 1|n_0 = 10, \tau_0, j_0 = 1)$, from which the conditional probability under consideration can be obtained by taking derivatives, as follows:

$$\Phi(n_1 = 5, \tau_1, j_1 = 1|n_0 = 10, \tau_0, j_0 = 1) = \frac{1}{5!}\frac{\partial^5\Psi(z, \tau)}{\partial z^5}\Big|_{z=0}. \tag{25}$$

When the system at initial time τ_0 is in promoter state 2 rather than 1, we have to switch the two components in the vector of Eqs. (22) and (23), use the entries of this vector to determine $\vec{F}$, according to Eq. (24), and again apply Eq. (6a) for ϕ^1 to obtain the generating function for the conditional probability $\Phi(n_1 = 5, \tau_1, j_1 = 1|n_0 = 10, \tau_0, j_0 = 2)$. And finally, to compute the conditional probabilities $\Phi(n_1 = 5, \tau_1, j_1 = 2|n_0 = 10, \tau_0, j_0)$, with $j_0 = 1$ or 2, we proceed in the same way, the only difference being that instead of using Eq. (6a) for ϕ^1 we use Eq. (6b) for ϕ^2.

From Eq. (14), the probability density for protein number is obtained as the limit

$$\mathcal{P}(\tau, m) = \lim_{L,p\to\infty} \mathbf{P}(m_\tau(n_0, \ldots, n_{p-1})), \tag{26}$$

where $\tau_{q+1} - \tau_q \to 0$ as $p \to \infty$ in such a way that the product $p\,(\tau_{q+1} - \tau_q)$ remains finite. The computational implementation of this limit is obtained by approximating the probability density by a histogram.

Finally, to consider arbitrarily long times, we take advantage of the fact that Eq. (2) is autonomous and hence its solutions have a composition property, namely:

$$m_{\tau,\tau} = \mathrm{id}, \qquad m_{\tau,\tau'} \circ m_{\tau',\tau''} = m_{\tau,\tau''}. \tag{27}$$

These formulas are obtained from the general solution of the initial value problem with $m(\tau') = m_{\tau'}$ $(\tau' < \tau)$,

$$m_{\tau,\tau'} = m_{\tau'}\,\mathrm{e}^{-\alpha(\tau-\tau')} + \beta\int_{\tau'}^{\tau} n_{\tau''}\,\mathrm{e}^{-\alpha(\tau-\tau'')}\mathrm{d}\tau,'' \tag{28}$$

which defines a family of transformations acting on the set of initial conditions. By iteration, it follows that the solution may be written as $m_\tau = m_{\tau_p,\tau_{p-1}} \circ \cdots \circ m_{\tau_1,\tau_0}$, where $\{\tau_0 = 0, \ldots, \tau_p = \tau\}$ is any subdivision of the time interval $[0, \tau]$ and each m_{τ_{q+1},τ_q} is given by Eq. (28), with the initial condition $m(\tau_q) = m_{\tau_q}$ having probability density $\mathcal{P}(\tau_q, m)$, for $q = 0, \ldots, p - 1$.

4 Moments of mRNA Number and Protein Number Distribution

The time-dependent mRNA moments can be obtained directly from the solutions of Eq. (5) given in Eq. (6), by transforming back to the original (z, τ) variables and taking derivatives of these generating functions with respect to the variable z at $z = 1$:

$$\langle n_\tau^{(r)} \rangle_j = \left(z \frac{\partial}{\partial z} \right)^r \phi^j(z, \tau) \Big|_{z=1} \quad (j = 1, 2). \tag{29}$$

Alternatively, we can view each of these moments as the solution of its own system of ordinary differential equations, obtained by applying the operator $(z\,\partial/\partial z)^r|_{z=1}$ directly to the system of partial differential equations (3), rather than its solutions. This is the procedure we shall adopt in what follows, for the first two moments.

As a preliminary step, we note that taking $r = 0$ [which amounts to simply evaluating Eq. (3) at $z = 1$] gives, for the promoter state occupancy probabilities

$$\pi_j(\tau) = \sum_{n \geq 0} \phi_n^j(\tau) = \phi^j(\tau, z = 1) \quad (j = 1, 2), \tag{30}$$

the following system of differential equations,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\tau} \pi_1 &= -\epsilon p_2\, \pi_1 + \epsilon p_1\, \pi_2, \\ \frac{\mathrm{d}}{\mathrm{d}\tau} \pi_2 &= \epsilon p_2\, \pi_1 - \epsilon p_1\, \pi_2. \end{aligned} \tag{31}$$

Its solution is immediate,

$$\begin{aligned} \pi_1(\tau) &= p_1 + (\pi_1(0) - p_1)\, \mathrm{e}^{-\epsilon\tau}, \\ \pi_2(\tau) &= p_2 + (\pi_2(0) - p_2)\, \mathrm{e}^{-\epsilon\tau}, \end{aligned} \tag{32}$$

provided we take into account that $p_1 + p_2 = 1$: this will imply that the constraint $\pi_1(\tau) + \pi_2(\tau) = 1$ is conserved (it holds for all τ provided it holds for the initial condition, i.e., for $\tau = 0$) and allow us to interpret the coefficients p_j as the asymptotic promoter state occupancy probabilities:

$$p_j = \lim_{\tau \to \infty} \pi_j(\tau) \quad (j = 1, 2). \tag{33}$$

4.1 Mean Values

Considering the case $r = 1$, we apply the operator $(z\,\partial/\partial z)$ to Eq. (3) and evaluate at $z = 1$ to obtain, for the mean partial mRNA numbers

$$\langle n_\tau^{(1)} \rangle_j = \sum_{n \geq 0} n\, \phi_n^j(\tau) = \left(z \frac{\partial}{\partial z} \right) \phi^j(z, \tau) \Big|_{z=1} \quad (j = 1, 2), \tag{34}$$

the following system of differential equations,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(1)}\rangle_1 &= -(1+\epsilon p_2)\langle n_\tau^{(1)}\rangle_1 + \epsilon p_1\langle n_\tau^{(1)}\rangle_2 + N_1\pi_1(\tau),\\ \frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(1)}\rangle_2 &= -(1+\epsilon p_1)\langle n_\tau^{(1)}\rangle_2 + \epsilon p_2\langle n_\tau^{(1)}\rangle_1 + N_2\pi_2(\tau).\end{aligned} \tag{35}$$

The corresponding differential equation for the mean total mRNA number

$$\langle n_\tau^{(1)}\rangle = \langle n_\tau^{(1)}\rangle_1 + \langle n_\tau^{(1)}\rangle_2 \tag{36}$$

is obtained by summing over j:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(1)}\rangle = -\langle n_\tau^{(1)}\rangle + N_1\pi_1(\tau) + N_2\pi_2(\tau). \tag{37}$$

Note that we can solve this equation without having to solve the full system (35). Namely, introducing the constants

$$\bar{N} = N_1 p_1 + N_2 p_2, \quad \Delta N = N_1 - N_2, \tag{38}$$

we get from Eq. (32)

$$N_1\pi_1(\tau) + N_2\pi_2(\tau) = \bar{N} + \Delta N\,(\pi_1(0) - p_1)\,\mathrm{e}^{-\epsilon\tau},$$

and this can be used to integrate Eq. (37), after putting it in the form

$$\mathrm{e}^{-\tau}\frac{\mathrm{d}}{\mathrm{d}\tau}\left(\mathrm{e}^{\tau}\langle n_\tau^{(1)}\rangle\right) = N_1\pi_1(\tau) + N_2\pi_2(\tau).$$

The solution is

$$\langle n_\tau^{(1)}\rangle = \bar{N} + \left(\langle n_0^{(1)}\rangle - \bar{N}\right)\mathrm{e}^{-\tau} + \frac{\Delta N}{1-\epsilon}\left(\pi_1(0) - p_1\right)\left(\mathrm{e}^{-\epsilon\tau} - \mathrm{e}^{-\tau}\right), \tag{39}$$

with the asymptotic value

$$\langle n_\infty^{(1)}\rangle = \lim_{\tau\to\infty}\langle n_\tau^{(1)}\rangle = \bar{N}. \tag{40}$$

For later use, we record here the complete solution of the system (35) because it will be needed at the next stage; it reads

$$\begin{aligned}\langle n_\tau^{(1)}\rangle_1 = \langle n_\infty^{(1)}\rangle_1 &+ \frac{\epsilon\,\Delta N\,p_1(\pi_1(0) - p_1)}{1-\epsilon}\,\mathrm{e}^{-\tau}\\ &+ \frac{[\epsilon(\Delta N\,p_1 - N_1) + N_1](\pi_1(0) - p_1)}{1-\epsilon}\,\mathrm{e}^{-\epsilon\tau}\\ &- \frac{\epsilon\,\Delta N(\pi_1(0) - p_1)(\pi_2(0) - p_1)}{1+\epsilon}\,\mathrm{e}^{-(1+\epsilon)\tau}\end{aligned} \tag{41}$$

for the partial mean value when the gene is in the state 1, and

$$\begin{aligned}\langle n_\tau^{(1)}\rangle_2 = \langle n_\infty^{(1)}\rangle_2 &+ \frac{\epsilon\, \Delta N\, p_2(\pi_1(0)-p_1)}{1-\epsilon}\, \mathrm{e}^{-\tau} \\ &- \frac{[\epsilon(\Delta N\, p_1 - N_1) + N_2](\pi_1(0)-p_1)}{1-\epsilon}\, \mathrm{e}^{-\epsilon\tau} \\ &+ \frac{\epsilon\, \Delta N(\pi_1(0)-p_1)(\pi_2(0)-p_1)}{1+\epsilon}\, \mathrm{e}^{-(1+\epsilon)\tau}\end{aligned} \tag{42}$$

for the partial mean value when the gene is in the state 2, with the asymptotic values

$$\langle n_\infty^{(1)}\rangle_1 = N_1 p_1 + \frac{\epsilon N_2 p_2}{1+\epsilon}, \quad \langle n_\infty^{(1)}\rangle_2 = N_2 p_2 + \frac{\epsilon N_1 p_1}{1+\epsilon}. \tag{43}$$

The ordinary differential equation governing the mean protein number density is obtained by averaging Eq. (2) which, in terms of the rescaled variables, gives

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\langle m_\tau\rangle = -\alpha\,\langle m_\tau\rangle + \beta\,\langle n_\tau^{(1)}\rangle. \tag{44}$$

The solution of this equation is as in Eq. (9):

$$\langle m_\tau\rangle = \langle m_0\rangle\, \mathrm{e}^{-\alpha\tau} + \beta\, \mathrm{e}^{-\alpha\tau} \int_0^\tau \langle n_{\tau'}^{(1)}\rangle\, \mathrm{e}^{\alpha\tau'}\, \mathrm{d}\tau'. \tag{45}$$

Using Eqs. (39) and (45), we integrate this to find

$$\begin{aligned}\langle m_\tau\rangle = \langle m_0\rangle\, \mathrm{e}^{-\alpha\tau} &+ \bar{N}\frac{\beta}{\alpha}(1-\mathrm{e}^{-\alpha\tau}) + \beta\left(\langle n_0^{(1)}\rangle - \bar{N}\right)\left(\frac{\mathrm{e}^{-\tau}-\mathrm{e}^{-\alpha\tau}}{\alpha-1}\right) \\ &+ \Delta N\frac{\beta}{1-\epsilon}(\pi_1(0)-p_1)\left(\frac{\mathrm{e}^{-\epsilon\tau}-\mathrm{e}^{-\alpha\tau}}{\alpha-\epsilon} - \frac{\mathrm{e}^{-\tau}-\mathrm{e}^{-\alpha\tau}}{\alpha-1}\right).\end{aligned} \tag{46}$$

with the asymptotic value

$$\langle m_\infty\rangle = \lim_{\tau\to\infty}\langle m_\tau\rangle = \bar{N}\frac{\beta}{\alpha}. \tag{47}$$

4.2 Variance

Passing to the case $r = 2$, we apply the operator $(z\,\partial/\partial z)$ to Eq. (3) twice and evaluate at $z = 1$ to obtain, for the partial second moments

$$\langle n_\tau^{(2)}\rangle_j = \sum_{n\geq 0} n^2\,\phi_n^j(\tau) = \left(z\frac{\partial}{\partial z}\right)^2 \phi^j(z,\tau)\Big|_{z=1} \quad (j=1,2), \tag{48}$$

the following system of ordinary differential equations,

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(2)}\rangle_1 &= -2\langle n_\tau^{(2)}\rangle_1 + (2N_1 + 1 - \epsilon p_2)\langle n_\tau^{(1)}\rangle_1 + \epsilon p_1 \langle n_\tau^{(1)}\rangle_2 + N_1\pi_1(\tau),\\ \frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(2)}\rangle_2 &= -2\langle n_\tau^{(2)}\rangle_2 + (2N_2 + 1 - \epsilon p_1)\langle n_\tau^{(1)}\rangle_2 + \epsilon p_2 \langle n_\tau^{(1)}\rangle_1 + N_2\pi_2(\tau).\end{aligned} \tag{49}$$

The corresponding differential equation for the total second moment

$$\langle n_\tau^{(2)}\rangle = \langle n_\tau^{(2)}\rangle_1 + \langle n_\tau^{(2)}\rangle_2 \tag{50}$$

is obtained by summing over j:

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(2)}\rangle = -2\langle n_\tau^{(2)}\rangle + (2N_1 + 1)\langle n_\tau^{(1)}\rangle_1 + (2N_2 + 1)\langle n_\tau^{(1)}\rangle_2 + N_1\pi_1(\tau) + N_2\pi_2(\tau). \tag{51}$$

Equivalently, we can derive a differential equation directly for the variance

$$V(n_\tau) = \langle n_\tau^{(2)}\rangle - \langle n_\tau^{(1)}\rangle^2 \tag{52}$$

by using Eq. (37) to deduce that

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(1)}\rangle^2 = 2\langle n_\tau^{(1)}\rangle \frac{\mathrm{d}}{\mathrm{d}\tau}\langle n_\tau^{(1)}\rangle = -2\langle n_\tau^{(1)}\rangle^2 + 2\langle n_\tau^{(1)}\rangle(N_1\pi_1(\tau) + N_2\pi_2(\tau))$$

and subtracting this result from Eq. (51) to arrive at

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\tau}V(n_\tau) = &-2V(n_\tau) + \langle n_\tau^{(1)}\rangle[1 - 2(N_1\pi_1(\tau) + N_2\pi_2(\tau))]\\ &+ 2N_1\langle n_\tau^{(1)}\rangle_1 + 2N_2\langle n_\tau^{(1)}\rangle_2 + N_1\pi_1(\tau) + N_2\pi_2(\tau).\end{aligned} \tag{53}$$

Again, we can solve Eqs. (51) and (53) without having to solve the full system (49), but here we now need the full solution of the system (35), Eqs. (41) and (42). For the variance, this solution has the following structure:

$$V(n_\tau) = A_1 + B_1\mathrm{e}^{-\tau} + C_1\mathrm{e}^{-2\tau} + D_1\mathrm{e}^{-\epsilon\tau} + E_1\mathrm{e}^{-(1+\epsilon)\tau} + F_1\mathrm{e}^{-2\epsilon\tau}, \tag{54}$$

with coefficients given by:

$$\begin{aligned}A_1 &= \bar{N} + \frac{(\Delta N)^2\, p_1(1-p_1)}{1+\epsilon},\\ B_1 &= -\frac{\epsilon\,\Delta N\,(\pi_1(0) - p_1)}{1-\epsilon},\\ C_1 &= -\frac{\epsilon\,(\Delta N)^2\,(\pi_1(0) - p_1)\,[2\pi_1(0) - \epsilon(\pi_1(0) - p_2) - 1]}{(1-\epsilon)^2(2-\epsilon)},\\ D_1 &= \Delta N\,(\pi_1(0) - p_1)\left[\frac{1}{1-\epsilon} + \frac{2\,\Delta N\,(1-2p_1)}{2-\epsilon}\right],\end{aligned}$$

$$E_1 = \frac{2\epsilon\,(\Delta N)^2\,(\pi_1(0) - p_1)\,[2\pi_1(0) - \epsilon(1-2p_1) - 1]}{(1+\epsilon)(1-\epsilon)^2},$$
$$F_1 = -\frac{(\Delta N)^2\,(\pi_1(0) - p_1)^2}{(1-\epsilon)^2}. \tag{55}$$

Our final goal will be to analyze the variance of the protein number density,

$$V(m_\tau) = \langle m_\tau^2\rangle - \langle m_\tau\rangle^2. \tag{56}$$

Using the solution for $\langle m_\tau\rangle$ in its integral representation, Eq. (45), the expression for $\langle m_\tau\rangle^2$ is

$$\begin{aligned}\langle m_\tau\rangle^2 = \langle m_0\rangle^2\,\mathrm{e}^{-2\alpha\tau} + 2\beta\,\mathrm{e}^{-2\alpha\tau}\int_0^\tau \langle m_0\rangle\langle n_{\tau'}^{(1)}\rangle\,\mathrm{e}^{\alpha\tau'}d\tau' \\ + \beta^2\mathrm{e}^{-2\alpha\tau}\int_0^\tau\int_0^\tau \langle n_{\tau'}^{(1)}\rangle\langle n_{\tau''}^{(1)}\rangle\,\mathrm{e}^{\alpha(\tau'+\tau'')}d\tau' d\tau.''\end{aligned} \tag{57}$$

The expression for $\langle m_\tau^2\rangle$ is obtained by first squaring Eq. (9) and then averaging, leading to:

$$\begin{aligned}\langle m_\tau^2\rangle = \langle m_0^2\rangle\,\mathrm{e}^{-2\alpha\tau} + 2\beta\,\mathrm{e}^{-\alpha\tau}\int_0^\tau \langle m_0\, n_{\tau'}\rangle\,\mathrm{e}^{\alpha\tau'}d\tau' \\ + \beta^2\mathrm{e}^{-2\alpha\tau}\int_0^\tau\int_0^\tau \langle n_{\tau'}n_{\tau''}\rangle\,\mathrm{e}^{\alpha(\tau'+\tau'')}d\tau' d\tau.''\end{aligned} \tag{58}$$

With these expressions at hand and in view of the fact that $\langle m_0 n_\tau\rangle = \langle m_0\rangle\langle n_\tau^{(1)}\rangle$, which means that the initial condition m_0 for protein number is independent of the mRNA process n_τ, we arrive at an explicit expression for the variance of protein number:

$$V(m_\tau) = \mathrm{e}^{-2\alpha\tau}\left[V(m_0) + \beta^2\underbrace{\int_0^\tau\int_0^\tau \mathrm{e}^{\alpha(s+s')}\left(\langle n_s n_{s'}\rangle - \langle n_s^{(1)}\rangle\langle n_{s'}^{(1)}\rangle\right)ds\,ds'}_{I_\tau}\right], \tag{59}$$

where $\langle n_s n_{s'}\rangle$ is the mRNA correlation function. Using the tower property of the conditional expectation and the Markov property of the solution of the ME, we get, for $s > s'$

$$\begin{aligned}\langle n_s n_{s'}\rangle &= \sum_{n'}\sum_{n}\sum_{j} n\,n'\,\Phi^j(n', s'; n, s) \\ &= \sum_{n',j'} n'\left[\sum_{n,j} n\,\Phi(n, s, j|n', s', j')\right]\phi_{n'}^{j'}(s') \\ &= \sum_{n',j'} n'\,\langle n_{s-s'}^{(1)}\rangle_{n',j'}\,\phi_{n'}^{j'}(s'),\end{aligned} \tag{60}$$

where the $\phi_{n'}^{j'}(s')$ are the components of the solution of the master equations at time s', the $\Phi(n, s, j|n', s', j')$ are the conditional probabilities as in Eq. (16) with $p = 2$,

and $\langle n^{(1)}_{s-s'}\rangle_{n',j'} = \sum_{n\geq 0} n\,\Phi(n,s,j|n',s',j')$ is the mean mRNA number at time s starting out with n' mRNA molecules and in promoter state j' at time s'. Now the latter is obtained directly by adapting Eq. (39) to this shifted initial time and these initial conditions, resulting in

$$\langle n^{(1)}_{s-s'}\rangle_{n',j'} = \bar{N}+(n'-\bar{N})\,\mathrm{e}^{-(s-s')}+\frac{\Delta N}{1-\epsilon}(\delta_{j',1}-p_1)\left(\mathrm{e}^{-\epsilon(s-s')}-\mathrm{e}^{-(s-s')}\right), \quad (61)$$

where δ is the Kronecker symbol ($\delta_{j',j}$=1 when $j'=j$ and $\delta_{j',j}=0$ when $j'\neq j$). From Eqs. (60) and (61), it follows that, for $s>s'$,

$$\begin{aligned}&\langle n_s n_{s'}\rangle - \langle n^{(1)}_s\rangle\langle n^{(1)}_{s'}\rangle \\ &= V(n_{s'})\,\mathrm{e}^{-(s-s')} + \frac{\Delta N}{1-\epsilon}\left(\langle n^{(1)}_{s'}\rangle_1 - \pi_1(s')\langle n^{(1)}_{s'}\rangle\right)(\mathrm{e}^{-\epsilon(s-s')}-\mathrm{e}^{-(s-s')}).\end{aligned} \quad (62)$$

From Eqs. (32), (39) and (41), it follows that the quantity $\langle n^{(1)}_s\rangle_1 - \pi_1(s)\langle n^{(1)}_s\rangle$ has the structure:

$$\langle n^{(1)}_s\rangle_1 - \pi_1(s)\langle n^{(1)}_s\rangle = A_2 + B_2\mathrm{e}^{-\epsilon s} + C_2\mathrm{e}^{-(1+\epsilon)s} + D_2\mathrm{e}^{-2\epsilon s}, \quad (63)$$

with coefficients:

$$\begin{aligned}A_2 &= \frac{\Delta N\, p_1(1-p_1)}{1+\epsilon},\\ B_2 &= \Delta N\,(1-2p_1)(\pi_1(0)-p_1),\\ C_2 &= \frac{\epsilon\,\Delta N\,[2\pi_1(0)+\epsilon(1-2p_1)-1]\,(\pi_1(0)-p_1)}{(1+\epsilon)(1-\epsilon)},\\ D_2 &= -\frac{\Delta N\,(\pi_1(0)-p_1)^2}{1-\epsilon}.\end{aligned} \quad (64)$$

Using (62),(54),(63), we find, for $s>s'$

$$\langle n_s n_{s'}\rangle - \langle n^{(1)}_s\rangle\langle n^{(1)}_{s'}\rangle = \sum_i K_i\,\mathrm{e}^{c_i s+d_i s'}, \quad (65)$$

and similarly, for $s'>s$,

$$\langle n_s n_{s'}\rangle - \langle n^{(1)}_s\rangle\langle n^{(1)}_{s'}\rangle = \sum_i K_i\,\mathrm{e}^{c_i s'+d_i s}, \quad (66)$$

with coefficients K_i, c_i, d_i given in Table 1.
Putting everything together, we are now in a position to evaluate the integral I_τ in Eq. (59): it has the form

Table 1 Coefficients in Eqs. (65) and (66)

i	c_i	d_i	K_i
1	-1	1	$A_1 - A_2\Delta N/(1-\epsilon)$
2	$-\epsilon$	ϵ	$A_2\Delta N/(1-\epsilon)$
3	-1	0	B_1
4	-1	-1	C_1
5	-1	$1-\epsilon$	$D_1 - B_2\Delta N/(1-\epsilon)$
6	-1	$-\epsilon$	$E_1 - C_2\Delta N/(1-\epsilon)$
7	$-\epsilon$	0	$B_2\Delta N/(1-\epsilon)$
8	$-\epsilon$	-1	$C_2\Delta N/(1-\epsilon)$
9	$-\epsilon$	$-\epsilon$	$D_2\Delta N/(1-\epsilon)$
10	-1	$1-2\epsilon$	$F_1 - D_2\Delta N/(1-\epsilon)$

$$
\begin{aligned}
I_\tau = &\sum_i K_i \int_0^\tau \left(\int_{s'}^\tau \mathrm{e}^{c_i s' + d_i s}\, \mathrm{e}^{\alpha(s+s')} ds \right) ds' \\
&+ \sum_i K_i \int_0^\tau \left(\int_s^\tau \mathrm{e}^{c_i s + d_i s'}\, \mathrm{e}^{\alpha(s+s')} ds' \right) ds,
\end{aligned}
\tag{67}
$$

so evaluating these integrals we get

$$
I_\tau = \sum_i \left[\frac{2K_i\, \mathrm{e}^{(2\alpha + c_i + d_i)\tau}}{(\alpha + d_i)(2\alpha + c_i + d_i)} - \frac{2K_i\, \mathrm{e}^{(\alpha + c_i)\tau}}{(\alpha + c_i)(\alpha + d_i)} + \frac{2K_i}{(\alpha + c_i)(2\alpha + c_i + d_i)} \right].
\tag{68}
$$

This gives us our final result for the protein number density variance:

$$
\begin{aligned}
V(m_\tau) = V(m_0)\mathrm{e}^{-2\alpha\tau} &+ \sum_i \frac{2\beta^2 K_i\, \mathrm{e}^{(c_i + d_i)\tau}}{(\alpha + d_i)(2\alpha + c_i + d_i)} \\
&- \sum_i \frac{2\beta^2 K_i\, \mathrm{e}^{(c_i - \alpha)\tau}}{(\alpha + c_i)(\alpha + d_i)} + \sum_i \frac{2\beta^2 K_i\, \mathrm{e}^{-2\alpha\tau}}{(\alpha + c_i)(2\alpha + c_i + d_i)},
\end{aligned}
\tag{69}
$$

with asymptotic value

$$
\begin{aligned}
\lim_{\tau\to\infty} V(m_\tau) &= \frac{\beta^2}{\alpha} \sum_{c_i + d_i = 0} \frac{K_i}{\alpha + d_i} \\
&= \frac{\beta^2}{\alpha} \left[\frac{A_1}{\alpha + 1} + \frac{A_2\, \Delta N}{1 - \epsilon} \left(\frac{1}{\alpha + \epsilon} - \frac{1}{\alpha + 1} \right) \right] \\
&= \frac{\beta^2}{\alpha(\alpha + 1)} \left[\bar{N} + (\Delta N)^2 \frac{(\alpha + \epsilon + 1)\, p_1(1 - p_1)}{(\alpha + \epsilon)(\epsilon + 1)} \right].
\end{aligned}
\tag{70}
$$

The expression (70) can be compared to the steady-state protein number variance obtained from the completely discrete protein expression model in Innocentini and

Hornos (2007). In that model, the protein copy number is treated as a discrete variable, whereas in the present model it is a continuous variable (density). Consequently, we expect to lose the contribution to the total variance that stems from discreteness of the protein birth-and-death process. And indeed, the term $\langle m_\infty \rangle = \bar{N}\frac{\beta}{\alpha}$ present in the steady-state variance, as computed in Innocentini and Hornos (2007), is missing from our expression (70). This term corresponds to the Poissonian component added to the protein variance by the discrete stochastic protein production and degradation processes. The hybrid model studied here is a good approximation of the full discrete model studied in Innocentini and Hornos (2007) when the neglected term is much smaller than the remaining terms in the variance. Considering the worst case, $p_1 = 0$ or $p_1 = 1$, when the switching contribution to the variance vanishes, we obtain the condition

$$\bar{N}\frac{\beta}{\alpha} \ll \bar{N}\frac{\beta^2}{\alpha(\alpha+1)}$$

which can be simplified to:

$$\alpha + 1 \ll \beta. \tag{71}$$

If the estimate (71) is satisfied then we can consider that the protein density follows an RDE [Eq. (2)] without losing accuracy of the steady-state variance. Let us recall that α stands for the mRNA to protein lifetime ratio and β for the average number of protein molecules produced by an mRNA molecule during its lifetime. In fact, the estimate (71) means that protein production should be efficient: one mRNA must produce several proteins. This is natural because the Kramers-Moyal expansion of the ME, which will lead to an RDE for protein population, works if the protein numbers are large compared to the mRNA numbers (Crudu et al. 2009, 2012).

5 Results

Following the approach discussed above, we have calculated the time-dependent probability distributions for mRNA molecules and protein density. More precisely, the dynamics of the probability distribution for the mRNA population is obtained by applying Eq. (4) to the exact solution of the master equations (6), written in terms of the original variables t and z. From that, we can compute, at each instant of time τ, the push-forward measure under the mapping given by Eq. (11), as defined by Eq. (14).

The result of this calculation is an ensemble of protein density values with their corresponding probabilities, $\{(m_\tau^k, \mathbf{P}(m_\tau^k)) : k = 1, \ldots, (L+1)^p\}$. Graphically, such ensembles will be represented by histograms where the probabilities are summed up within each bin. More precisely, if we fix a bin size and group together all m_τ^k belonging to the same bin, the probability assigned to that bin is simply the sum of all the probabilities $\mathbf{P}(m_\tau^k)$ corresponding to the m_τ^k in that bin.

In order to estimate the accuracy of our method, we compare the distributions obtained by our formalism with those from a Monte Carlo (MC) simulation of the model. We have used the MC simulation to generate trajectories of the mRNA process n_τ. Namely, let the τ_q be the random times when the birth-and-death process for

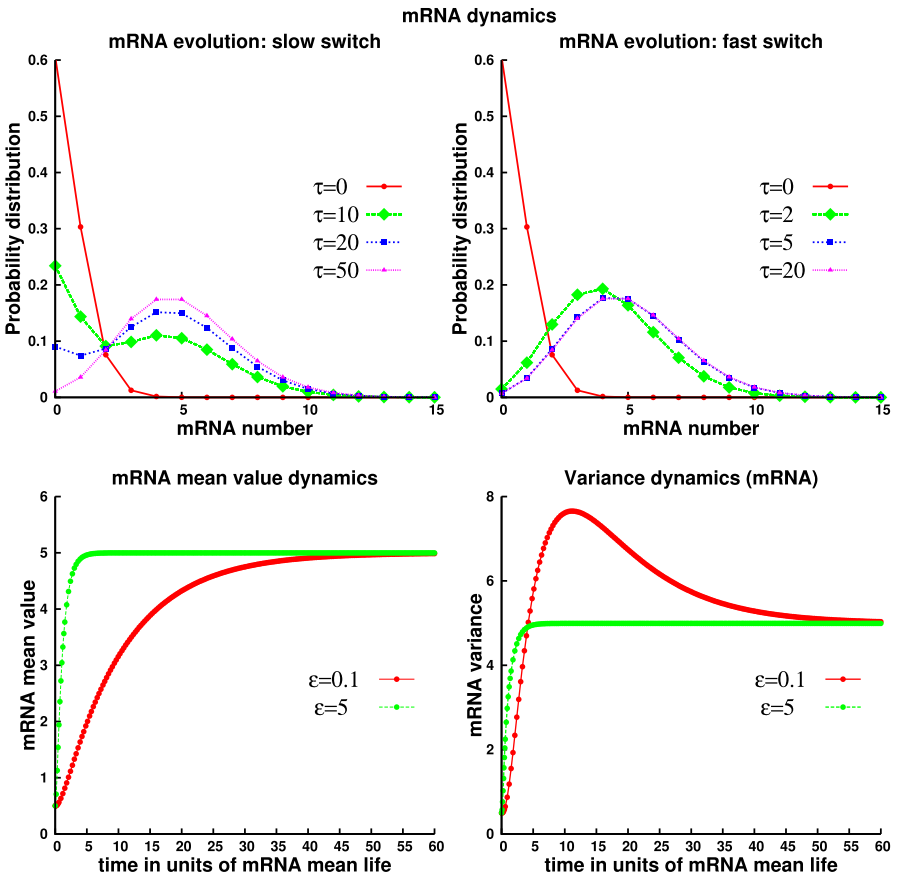


Fig. 1 (Colour figure online) mRNA dynamics in slow ($\epsilon = 0.1$) and fast ($\epsilon = 5$) switch regimes. Remaining parameters: $N_1 = 5$, $N_2 = 0.5$, $p_1 = 1$, $p_2 = 0$

mRNA molecules produces a change from n_τ to $n_\tau \pm 1$. Then, Eq. (11) can be used to directly compute samples of the protein process m_τ. Note that this makes our hybrid model much easier to simulate than the full discrete mRNA/protein model, since we avoid the separate simulation of the protein process, which is computationally costly.

As an example of our results, we exhibit in Fig. 1 the time evolution for the probability distribution of the mRNA population, its mean value and variance. In all cases, we have used as initial mRNA configuration the generating function $\phi(\mu) = \exp(N_2\mu)$, representing the gene with probability one in the off state, that is, the initial mRNA number follows a Poisson distribution with mean equal to N_2. On the other hand, the occupancy probabilities have been chosen as $p_1 = 1$, $p_2 = 0$, so as to produce a final equilibrium state which represents the gene in full activity and with mRNA number following a Poisson distribution with mean equal to N_1. Concerning the switching parameter ϵ, we have selected two values: $\epsilon = 5$ representing the "fast switch regime" and $\epsilon = 0.1$ representing the "slow switch regime." Specifically, in

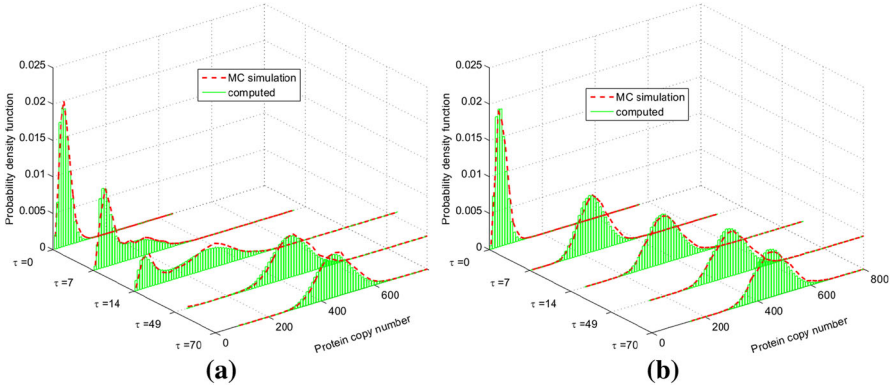


Fig. 2 (Colour figure online) Dynamical evolution of protein probability density in slow switch ($\epsilon = 0.1$) in **a** and fast switch ($\epsilon = 5$) in **b**, and comparison with MC simulation. Remaining parameters: $N_1 = 5$, $N_2 = 0.5$, $p_1 = 1$, $p_2 = 0$, $\alpha = 1/9$, $\beta = 10$

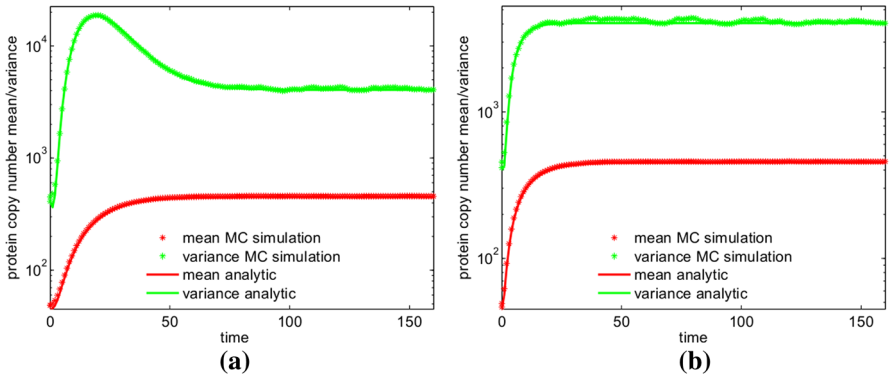


Fig. 3 (Colour figure online) Dynamical evolution of protein mean value and variance. Slow switch ($\epsilon = 0.1$) in **a** and fast switch ($\epsilon = 5$) in **b**. Remaining parameters: $N_1 = 5$, $N_2 = 0.5$, $p_1 = 1$, $p_2 = 0$, $\alpha = 1/9$, $\beta = 10$

Fig. 2 we exhibit, for the two switch regimes, a direct comparison between the distributions obtained by our method (green histograms) and those from MC simulation (red curves), and finally, in Fig. 3 we show the mean value and variance of the protein distribution, comparing the analytical formulas presented in Sect. 4 with the results of a direct simulation of the model.

The transient behavior of mRNA in the slow switch regime has a two-peak distribution, whereas otherwise it is unimodal (see Fig. 1, top). The same feature is present in the protein probability density (see Fig. 2). The bimodality is accompanied by an increase in the noise in the transient time, captured in the overshoots of Fig. 1, bottom right, and Fig. 3a. This phenomenon disappears in the fast switch regime, in accordance with the fact that increasing the gene switch parameter decreases the standard deviation in mRNA production, which is a well-known effect (Innocentini and Hornos 2007; Innocentini et al. 2013).

6 Discussion and Conclusion

The hybrid model presented here shows how to couple transcription and translation providing a complete picture of the entire dynamical process, without any restrictions on the parameter space. The randomness of protein synthesis due to the stochastic nature of transcription is exhibited in the dynamical behavior of the protein probability density. The main result is a full time-dependent solution for the probability distribution of mRNA as well as for the density probability for protein numbers—something that, to the best of our knowledge, has never been achieved before. Moreover, the distributions for protein number obtained by our method are in excellent agreement with those derived from MC simulations, at highly reduced computational cost. But there is a technical issue that must still be overcome. Namely, in order to improve the precision of our method, we must use joint probabilities with many events overlapping in a specific time interval (bigger values of p). This becomes difficult when the average number of mRNA is large, because it will imply in a bigger value for the cutoff L and thus increase the set of possible values for each n_q, making the size of our sample space, $(L+1)^p$, become unpractically large. Methods to bypass this difficulty are presently under investigation.

It is worth mentioning that pure random differential equation (RDE) models—where the processes of mRNA production and of protein production are treated on equal footing, using random differential equations for both—have been introduced in Lipniacki et al. (2006) for the continuous time case and in Ferreira et al. (2009, 2013) for the discrete time case. Similarly, pure master equations (ME) models—where the processes of mRNA production and of protein production are also treated on equal footing, but using master euations for both—have been discussed in the literature before; see, for instance (Innocentini and Hornos 2007; Shahrezaei and Swain 2008). Both of these approaches are highly interesting and logically perfectly consistent, but a closer look reveals some drawbacks. On the one hand, using pure RDE models means that mRNA is represented by a continuous random variable, which is problematic since the number of mRNA molecules is small, of the order of a few dozen per gene. On the other hand, pure ME models are hard to solve explicitly and one has to resort to simulations or appeal to some approximation scheme in order to simplify the equations and then find expressions for the protein distribution (a discrete probability distribution) that solve these simplified equations, rather than the original ones.

Recent experiments allowing real-time observation of the expression of stochastic protein synthesis in living *Escherichia coli* or *Bacillus subtilis* cells, with single molecule sensitivity (Cai et al. 2006; Ferguson et al. 2012), have shown that information about key parameters of protein expression can be extracted from the steady-state distribution. Furthermore, measurements of protein concentration can be integrated with mRNA tagging techniques, such as MS2, that monitor mRNA production. The model discussed here can be used to extract quantitative information on transcription and translation processes from measured mRNA and protein distributions. In addition, the ability to compute the shape of the protein distribution may be used to improve the understanding of stochasticity in biological decision-making processes.

Future research will also be dedicated to developing the model to include other phenomenological aspects of gene expression. One modification consists in allowing the protein synthesis/degradation rates to be random variables, thus taking into account the inherent noise due to the translational process. The model can also be extended to study eukaryotes, which requires introducing a time-delay accounting for the transport of mRNA from the nucleus to the cytoplasm. Another modification amounts to adding a nonlinear term to the RDE, reflecting a decrease in protein number due to other effects than just degradation, such as complex formation by dimerization: this will introduce a bifurcation parameter and ultimately implement the observed multi-stability in the steady state of protein population [the bifurcation theory for RDEs can be found in Arnold (1998)]. In contrast to multi-stability, the multi-modality originating in the controlling mechanism of protein synthesis, at the translational level, can be introduced by allowing the parameter B (or β) in Eq. (2) to be a matrix, turning the RDE for protein density into a vector equation. The entries of this matrix will encode the different levels of translational efficiency.

Finally, the model can be used as a building block for constructing mathematical models of gene regulatory networks. More concretely, the idea is to take several copies of our model and couple them by allowing the binding/unbinding rates controlling the on/off switch of any gene to become functions of the mean values of the proteins expressed by the other genes. Traditionally, this coupling is performed through Hill type functions which convert protein densities into binding/unbinding rates. This strategy is in accordance with the ubiquitous idea in physics that simple models serve as building blocks for more complicated ones.

Acknowledgments We would like to thank the referees for their insights. Work supported by FAPESP, SP, Brazil (G. I., contract 2012/04723-4) and CNPq, Brazil (G. I., contract 202238/2014-8; M. F., contract 307238/2011-3; F. A., contract 306362/2012-0). O. R. thanks CNRS and LABEX Epigenmed for support.

References

Abramowitz M, Stegun IA (1964) Handbook of mathematical functions with formulas, graphs and mathematical tables. Government Printing Office, U.S

Arnold L (1998) Random dynamical systems. Springer, Berlin

Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. Nature 422:633–637

Cai L, Friedman N, Xie X (2006) Stochastic protein expression in individual cells at the single molecule level. Nature 440(7082):358–62. doi:10.1038/nature04599

Cogburn R, Torrez WC (1981) Birth and death processes with random environments in continuous time. J Appl Probab 18(1):19–30

Crudu A, Debussche A, Muller A, Radulescu O (2012) Convergence of stochastic gene networks to hybrid piecewise deterministic processes. Ann Appl Probab 22(5):1822–1859

Crudu A, Debussche A, Radulescu O (2009) Hybrid stochastic simplifications for multiscale gene networks. BMC Syst Biol 3(1):89

Delbrück M (1940) Statistical fluctuations in autocatalytic reactions. J Chem Phys 8:120–124

Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) Stochastic gene expression in a single cell. Science 297(5584):1183–1186. doi:10.1126/science.1070919

Ferguson M, Le Coq D, Jules M, Aymerich S, Radulescu O, Declerck N, Royer C (2012) Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. Proc Natl Acad Sci USA 109(1):155–160

Ferreira RC, Bosco FAR, Briones MRS (2009) Scaling properties of transcription profiles in gene networks. Int J Bioinform Res Appl 5(2):178–186

Ferreira RC, Briones MRS, Antoneli F. A model of gene expression based on random dynamical systems reveals modularity properties of gene regulatory networks. arXiv:1309.0765 (2013)
Friedman N, Cai L, Xie XS (2006) Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys Rev Lett 97(16):168,302
Golding I, Paulsson J, Zawilski S, Cox E (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123(6):1025–36. doi:10.1016/j.cell.2005.09.031
Hornos JEM, Schultz D, Innocentini GCP, Wang J, Walczak AM, Onuchic JN, Wolynes PG (2005) Self-regulating gene: an exact solution. Phys Rev E 72(5):e051,907. doi:10.1103/PhysRevE.72.051907
Innocentini GCP, Forger M, Ramos A, Radulescu O, Hornos JEM (2013) Multimodality and flexibility in stochastic gene expression. Bull Math Biol 75:2600–2630
Innocentini GCP, Hornos JEM (2007) Modeling stochastic gene expression under repression. J Math Biol 55(3):413–431. doi:10.1007/s00285-007-0090-x
Iyer-Biswas S, Hayot F, Jayaprakash C (2009) Stochasticity of gene products from transcriptional pulsing. Phys Rev E 79:031,911
Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. Biophys J 81(6):3116–3136. doi:10.1016/S0006-3495(01)75949-8
Lipniacki T, Paszek P, Marciniak-Czochra A, Brasier AR, Kimmel M (2006) Transcriptional stochasticity in gene expression. J Theor Biol 238:348–367. doi:10.1016/j.jtbi.2005.05.032
Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) Regulation of noise in the expression of a single gene. Nat Genet 31:69–73
Paulsson J (2005) Models of stochastic gene expression. Phys Life Rev 2:157–175
Peccoud J, Ycart B (1995) Markovian modelling of gene product synthesis. Theor Popul Biol 48:222–234
Pirone J, Elston T (2004) Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. J Theor Biol 226:111–121
Ramos AF, Innocentini GCP, Hornos JEM (2011) Exact time-dependent solutions for a self-regulating gene. Phys Rev E 83(6):e062,902. doi:10.1103/PhysRevE.83.062902
Raser JM, O'Shea EK (2004) Control of stochasticity in eukaryotic gene expression. Science 304(5678):1811–1814. doi:10.1126/science.1098641
Shahrezaei V, Swain PS (2008) Analytical distributions for stochastic gene expression. Proc Natl Acad Sci USA 105(45):17256–17261. doi:10.1073/pnas.0803850105
van Kampen NG (2007) Stochastic processes in physics and chemistry, 3rd edn. Elsevier, Amsterdam
Yu J, Xiao J, Ren X, Lao K, Xie XS (2006) Probing gene expression in live cells, one protein molecule at a time. Science 311(5767):1600–1603. doi:10.1126/science.1119623