

RESEARCH

Open Access



Translation regulation by RNA stem-loops can reduce gene expression noise

Candan Çelik^{1,2*} , Pavol Bokes^{1,3} and Abhyudai Singh⁴

From The 19th conference on Computational Methods in Systems Biology
Bordeaux, France. 22-24 September 2021. <https://cmsb2021.labri.fr/>

*Correspondence:
candanc@aydin.edu.tr

¹ Department of Applied
Mathematics and Statistics,
Comenius University,
84248 Bratislava, Slovakia

² Department of Industrial
Engineering, Istanbul Aydin
University, 34295 Istanbul, Turkey

³ Mathematical Institute,
Slovak Academy of Sciences,
81473 Bratislava, Slovakia

⁴ Department of Electrical
and Computer Engineering,
University of Delaware,
Newark 19716, USA

Abstract

Background: Stochastic modelling plays a crucial role in comprehending the dynamics of intracellular events in various biochemical systems, including gene-expression models. Cell-to-cell variability arises from the stochasticity or noise in the levels of gene products such as messenger RNA (mRNA) and protein. The sources of noise can stem from different factors, including structural elements. Recent studies have revealed that the mRNA structure can be more intricate than previously assumed.

Results: Here, we focus on the formation of stem-loops and present a reinterpretation of previous data, offering new insights. Our analysis demonstrates that stem-loops that restrict translation have the potential to reduce noise.

Conclusions: In conclusion, we investigate a structured/generalised version of a stochastic gene-expression model, wherein mRNA molecules can be found in one of their finite number of different states and transition between them. By characterising and deriving non-trivial analytical expressions for the steady-state protein distribution, we provide two specific examples which can be readily obtained from the structured/generalised model, showcasing the model's practical applicability.

Keywords: Stochastic gene expression, Master equation, Stochastic simulation

Background

Biochemical processes such as stochastic gene expression are inherently subject to random fluctuations that lead to noise in the number of constituents [1]. Quantifying the dynamics and the noise in such stochastic processes is an intense study of various research areas. Under simplest assumptions, gene expression is described as a two-step stochastic process comprised of transcription and translation that play a significant role in determining the levels of gene products. While RNA polymerase enzymes produce mRNA molecules in the former, protein synthesis takes place by ribosomes in the

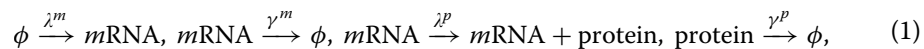


latter. Because of the similarity, it is often referred to as the (classical) two-stage gene-expression model. How gene-expression regulation affects the level of gene products such as mRNA and protein is a question of interest.

The contributions to gene expression noise give rise to cell-to-cell variability in the mRNA and protein levels [2–9]. The noise emerges from different sources, namely *intrinsic* and *extrinsic* noise [10, 11]; yet, structural elements such as stem-loops can also contribute to noise by binding to an untranslated region of mRNA [12, 13]. The untranslated regions of mRNAs often contain these stem-loops that can reversibly change configurations making individual mRNAs translationally active/inactive [14].

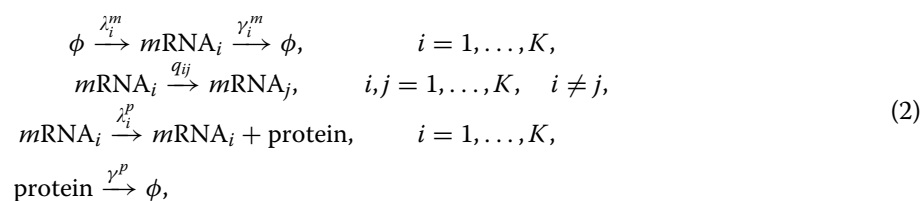
From a mathematical perspective, the dynamics of gene-expression mechanisms can be described in deterministic and stochastic settings by means of ordinary differential equations (ODEs) and Master equation formulation, respectively. On the other hand, hybrid models have also been proposed as a combination of the preceding two [15–17]. Only a few of those provide an explicit solution to the (classical) two-stage gene-expression model [18, 19]; most of the studies are based on Monte Carlo simulations, which are usually computationally expensive.

In recent decades, the (classical) two-stage model of gene expression has been extensively utilised to elucidate the underlying mechanisms of stochastic processes in living cells [19–22]. In particular, it has been extended by the regulation of transcription factors, which affect gene expression by modulating the binding rate of RNA polymerase [23]. Specifically, the stochastic dynamics of the classical two-stage model of gene expression is described by the reaction scheme [18, 24]



where λ^m is the mRNA production rate, λ^p is the protein translation rate, and γ^m and γ^p are the decay rate constants of mRNA and protein species, respectively. Here and henceforth, m and p in the superscript indicate the mRNA and protein species, respectively.

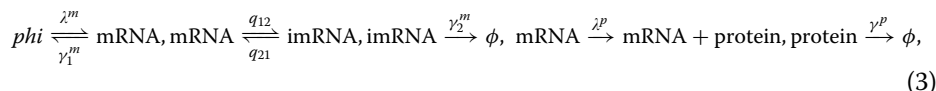
As a generalisation of the (classical) two-stage model, some studies in the literature consider a set of multiple gene states and investigate the dynamics of stochastic transitions among these states [25–28]. Here, we study a structuration/generalisation of the classical two-stage gene-expression model (1), which takes into account multiple mRNA states. More specifically, after being transcribed, mRNA molecules are considered to be transitioning among their different states at constant reaction rates. Subsequently, the nascent mRNA molecule is translated, and protein is degraded. The schematic of the reactions describing this system is given by the following set of chemical reactions:



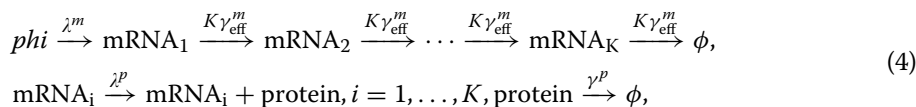
where λ_i^m and γ_i^m are the production and decay rates for an mRNA molecule in i -th state, respectively. The term q_{ij} , $i \neq j$, denotes the mRNA transition rate from state i to state j , λ_i^p and γ^p are the protein translation and decay rates, respectively. The subscript is reserved for multiple mRNA states. All model parameters and their biological meaning are listed in Table 1.

The chemical reactions in (2) correspond to mRNA transcription and decay, transitions among multiple mRNA states, protein translation, and protein decay, respectively. Throughout this paper, we refer to model (2) as the *generalised two-stage model*, by which we mean that the model is treated as an extension of the classical two-stage model concerning the structuration of mRNA. We note that the (classical) two-stage model has been extended in this manner by the inclusion of an mRNA activation/inactivation loop recently [29]; however, here we generalise the results of [29] for a more comprehensive model. Additionally, we reanalyse published data on the influence of RNA stem loops on gene expression noise and explore the influence of kinetic rate parameters on predicted noise reduction ratios. From a biologically relevant standpoint, a similar model involving multiple mRNA states has recently been studied to quantify protein variability arising from mRNA-microRNA interactions [30].

In what follows, we present two specific examples which can be obtained from the structured/generalised model (2): the mRNA inactivation loop model and the multiphasic mRNA model. These models are given by the reactions



where the abbreviation imRNA stands for an inactive mRNA molecule, and by



respectively. Here, the reaction system (3) accounts for the activation/inactivation of an mRNA molecule modelled by involving a pair of reversible chemical reactions. In (4), an mRNA molecule is considered to move through its finite lifetime stages, which

Table 1 Model parameters and their biological meaning used in all model variations

Parameter	Meaning
λ^m	mRNA production rate
γ^m	mRNA decay rate
λ^p	Protein translation rate
γ^p	Protein decay rate
λ_i^m	mRNA production rate in i -th state
γ_i^m	mRNA decay rate in i -th state
q_{ij}	mRNA transition rate from state i to j
λ_i^p	Protein translation rate in i -th state
γ_{eff}^m	Effective mRNA decay rate
K	The number of mRNA states

corresponds to the ageing of an mRNA. For a detailed discussion of these models, we refer the reader to Sections The mRNA inactivation loop model and Multiphasic mRNA lifetime and also the reference [29].

This paper is structured as follows. The core part of this study is given in Section Methods, where the generalised model is introduced in an in-depth analysis of a modelling framework. Specifically, in Section Model formulation, a brief review of the classical two-stage gene-expression model is given in stochastic settings; the underlying chemical master equation (CME) is transformed into a partial differential equation (PDE) for the generating function. Then, the main focus of this paper, which is the introduction of a generalization of the two-stage model, along with its corresponding CME and PDE, is presented. In Section Solution, a power series solution to the PDE is obtained. In Section Marginal distributions and moments, not only are the marginal mRNA and protein distributions obtained using the non-trivial analytical formula for the generating function, but the moments of the protein distributions are also determined by utilising factorial cumulants. The protein distribution is thereby recovered. Section Results pertains to data analysis, its interpretation, and summarises some of the key results of our mathematical analysis. The paper is concluded in Section Conclusions..

Results

The motivation for our mathematical analysis stems from a recent experimental study [12] on the influence of RNA stem loops on gene expression noise. Stem loops appear when two palindromic sequences on the chain of nucleic acids align and form hydrogen bonds. The aligned palindromic sequences then form the “stem” and the nucleic acids in between form the “loop” of a stem loop. Another term is “hairpin loop” because of resemblance.

The authors of [12] constructed several variants of a gene encoding for a fluorescent reporter protein. Although the constructs encode for the same reporter protein, they differ in palindromic sequences in the untranslated region at the 5’ end of the gene (5’UTR). The formation of a stem loop interferes with translation; the higher the stability of a stem loop, the greater the interference; the lower the mean. The authors also show that this is associated with an increase in the coefficient of variation (CV).

Previous theoretical studies indicate that different noise metrics can lead to different interpretations of the effects of a particular mechanism on gene expression noise. The most common are the squared coefficient of variation and the Fano factor defined by

$$CV^2 = \frac{\langle P^2 \rangle - \langle P \rangle^2}{\langle P \rangle^2}, \quad F = \frac{\langle P^2 \rangle - \langle P \rangle^2}{\langle P \rangle},$$

where P stands for the reporter protein and $\langle . \rangle$ are the averaging brackets. In Fig. 1, in addition to showing the dependence of the CV^2 on mean (thus reproducing Fig. 1 of [12]), we also show the dependence of $F = \langle P \rangle CV^2$ on the mean. Notably, decreasing the mean (which is associated with greater stem loop stability) decreases the Fano factor.

In order to explain the apparently contradictory interpretations, we fit the classical two-stage (transcription-translation) model (1) of gene expression [15, 24]. The model is described in full mathematical detail in Section Model formulation. For the purposes

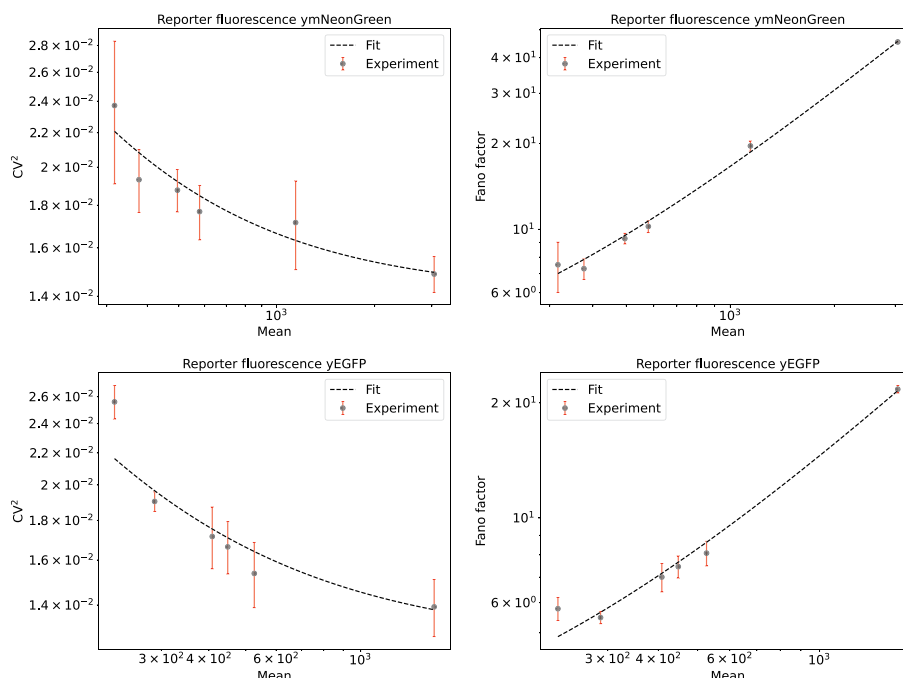


Fig. 1 Dependence of protein noise on protein mean for different 5'UTR constructs. The yEGFP reporter (bottom) and the ymNeonGreen reporter (top) constructs are treated separately. The use of a log-log scale is adopted from [12]. The dots give the experimental values taken from [12] (see Table 2). Each dot is a result of multiple experiments, and the error bars indicate the standard deviation (SD). These were obtained from the standard deviation of the (nonsquared) coefficient of variation by Taylor formula: $SD_{CV^2} = 2CVSD_{CV}$, $SD_F = \langle P \rangle SD_{CV^2}$. The dashed lines give the linear and hyperbolic dependence of the F and CV^2 , respectively, which are predicted by the two-stage gene expression model (cf. (5)). The protein translation rate λ^p and the mRNA decay rate γ^m are being varied to change the mean levels. Note that the use of the log-log scale results in a slight curvature of the line (with a nonzero intercept)

Table 2 Protein mean and noise (CV) values for the yEGFP and the ymNeonGreen reporters obtained from [12]. The first column denotes distinct constructs that are of different stabilities. For instance, the L0 (P_{TEF1}) construct is driven by a strong promoter P_{TEF1} , generating a large abundance of protein molecules per cell, whereas P_{PAB1} is a mid-range promoter. The hyphen symbol denotes undetermined values

Construct	yEGFP		ymNeonGreen	
	μ	CV (%)	μ	CV (%)
L0 (P_{TEF1})	1560	11.8 ± 0.5	3050	12.2 ± 0.3
U	526	12.4 ± 0.6	–	–
M1Ug	408	13.1 ± 0.6	–	–
M3g	226	16.0 ± 0.4	–	–
G ₁₀	448	12.9 ± 0.5	–	–
G ₁₄	–	–	317	15.4 ± 1.5
M3Wn	–	–	1143	13.1 ± 0.8
M3n	–	–	579	13.3 ± 0.5
M3Un	–	–	377	13.9 ± 0.6
L0 (P_{PAB1})	288	13.8 ± 0.2	495	13.7 ± 0.4

of the current section, we mention that it predicts that the stationary protein mean and Fano factor of the form

$$\langle P \rangle = \frac{\lambda^m \lambda^p}{\gamma^m \gamma^p}, \quad F = 1 + \frac{\lambda^p}{\gamma^m + \gamma^p},$$

where λ^m is the mRNA production rate, λ^p is the protein translation rate, γ^m and γ^p are the decay rate constants of mRNA and protein species, respectively. We note that here in the expressions for the classical two-stage model, we omit the subscript i on mRNA in the generalised model (2) because there is only one mRNA state. Provided that the protein is more stable than the mRNA ($\gamma^p \ll \gamma^m$), we can simplify to

$$F = 1 + \frac{\lambda^p}{\gamma^m} = 1 + \frac{\gamma^p \langle P \rangle}{\lambda^m}, \quad CV^2 = \frac{F}{\langle P \rangle} = \frac{1}{\langle P \rangle} + \frac{\gamma^p}{\lambda^m}. \tag{5}$$

Stem loops do not affect the transcription rate λ^m or the protein stability γ^p , but they can affect the protein mean through translation rate λ^p and mRNA decay rate γ^m . Thus, the two-stage model predicts an increasing linear dependence of the Fano factor, and a decreasing hyperbolic dependence of the CV^2 , on the mean. In Fig. 1, the Fano factor data are fit by a straight line using simple linear regression. The slope of the regression line corresponds to the fraction λ^p/γ^m in (5), which is calculated as 0.0141 and 0.0124 for the ymNeonGreen and yEGFP reporters, respectively. The regression coefficients are reused for the hyperbolic dependence of the CV^2 . The fits seem to be satisfactory, leading us to attribute the changes in the noise to the decrease of mean rather than an active control of noise by the stem–loop mechanism. In the same Fig. 1, a reporter of gene expression, yeast-enhanced green fluorescent protein (yEGFP), and a monomeric protein, yellow mNeonGreen (ymNeonGreen), are used to obtain the experimental results.

Let us address the question of noise control by stem–loop formation theoretically. For reasons of mathematical elegance, we will introduce a general model that extends the classical two-stage model (1) by multiple transcript states in Section Model formulation and provide a thorough analysis of the mRNA inactivation model (3) in Sections Solution-Marginal distributions and moments. Here we discuss the special case with two states, one of them translationally active (without a stem–loop), the other translationally inactive (with a stem–loop) (cf. Eq. (3)). This special case is analysed in Section The mRNA inactivation loop model. Importantly, we note that our results pertain to this special case; therefore, we drop the subscript i on mRNA species (cf. Eqs. (45) and (46)). Using standard methods, we derive that the mean is given by

$$\langle P \rangle = \frac{\lambda^p \lambda^m}{\gamma^p \gamma_{\text{eff}}^m},$$

where

$$\gamma_{\text{eff}}^m = \gamma_1^m + \frac{q_{12} \gamma_2^m}{\gamma_2^m + q_{21}} \tag{6}$$

gives an effective mRNA decay rate constant. The Fano factor satisfies

$$F = 1 + \frac{\lambda^P}{\gamma^P + \gamma_1^m + \frac{q_{12}(\gamma^P + \gamma_2^m)}{\gamma^P + \gamma_2^m + q_{21}}} \tag{7}$$

The above equations give the steady-state protein mean and Fano factor as function of the model parameters (degradation rate constants $\gamma_1^m, \gamma_2^m, \gamma^P$ of active/inactive mRNA and protein; inactivation/activation rate constants q_{12}, q_{21} ; translation rate constant λ^P). The formula for the mean implies, in particular, that making the stem–loop more stable (i.e. decreasing q_{21}) decreases the mean. The noise requires a more subtle analysis, which is given below.

In order to compare the protein noise in the current model to that exhibited by the classical two-stage model (without the inactivation–activation loop) we define the baseline Fano factor as

$$F_0 = 1 + \frac{\lambda^P}{\gamma^P + \gamma_{\text{eff}}^m} = 1 + \frac{\lambda^P}{\gamma^P + \gamma_1^m + \frac{q_{12}\gamma_2^m}{\gamma_2^m + q_{21}}} \tag{8}$$

which can be obtained from (7) by first setting $q_{12} = 0$ (no inactivation) and then replacing the mRNA decay rate γ_1^m by its effective value (6). Adjusting the mRNA decay rate maintains the same species means in the baseline model like in the full model extended by the inactivation loop.

The protein variability formulae (7) and (8) can equivalently be expressed in terms of the squared coefficient of variation [31, 32] $CV^2 = F/\langle P \rangle$ and $CV_0^2 = F_0/\langle P \rangle$. We find that

$$CV^2 = \frac{1}{\langle P \rangle} + \frac{1}{\langle M \rangle} \frac{\gamma^P}{\gamma^P + \gamma_1^m + \frac{q_{12}(\gamma^P + \gamma_2^m)}{\gamma^P + \gamma_2^m + q_{21}}} \tag{9}$$

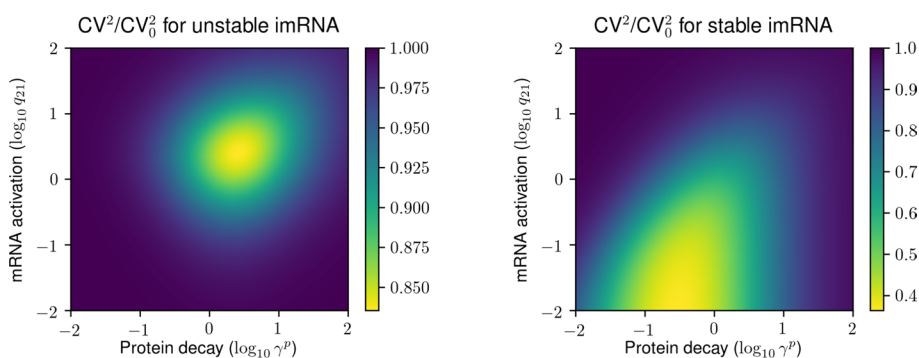


Fig. 2 Fractional protein noise reduction by the mRNA inactivation loop as function of protein decay and mRNA activation rate constants. The colour of the heat map gives the protein noise (the squared coefficient of variation) in the two-stage model extended by the mRNA inactivation loop relative to the protein noise in a baseline two-stage model without the mRNA inactivation loop (adjusting the mRNA decay rate to obtain the same species means). The mRNA mean is set to $\langle M \rangle = 10$, and protein mean is $\langle P \rangle = 500$. The mRNA decay rate is set to $\gamma_1^m = 1$ without loss of generality; the inactive mRNA decay rate is either the same as that of active mRNA ($\gamma_2^m = 1$; left panel) or set to zero ($\gamma_2^m = 0$; right panel). The inactivation rate constant is $q_{12} = 3$

$$CV_0^2 = \frac{1}{\langle P \rangle} + \frac{1}{\langle M \rangle} \frac{\gamma^p}{\gamma^p + \gamma_1^m + \frac{q_{12}\gamma_2^m}{\gamma_2^m + q_{21}}}, \tag{10}$$

where $\langle M \rangle = \lambda^m / \gamma_{eff}^m$ is the mean value of the activated mRNA.

Comparing (9) to (10), we see that $CV^2 < CV_0^2$, allowing us to conclude that the inclusion of the mRNA inactivation loop decreases protein noise. The key ingredient that distinguishes (9) from (10) is the Michaelis-Menten-type term in the denominator that involves the protein decay rate γ^p , the mRNA activation rate q_{21} and the inactive mRNA decay rate γ_2^m . Figure 2 explores the dependence of the fractional protein noise reduction CV^2 / CV_0^2 on these parameters. Without loss of generality, the active mRNA decay rate is set to one, and two plausible alternatives are considered for the decay of the inactive mRNA molecule: one, imRNA is unstable, decaying with the same rate as mRNA (left panel of Fig. 2); two, imRNA is stable, i.e. protected from degradation, and does not decay (right panel of Fig. 2). We observe qualitatively different noise reduction patterns for the two alternatives. For an unstable imRNA, there is an optimal combination of mRNA activation and protein decay rate constants that minimise the protein noise (Fig. 2, left panel). We observe that the optimal values are both greater than the unit value of the mRNA decay rate. This requirement runs counter to the biological evidence that proteins are typically more stable than mRNAs. The maximal reduction of noise is moderate (around 15% for the chosen parameter set). For a stable imRNA, the optimum is approached by making the mRNA activation as slow as possible (Fig. 2, right panel). The optimal value of protein decay rate constant is then less than that of the active mRNA, and the maximal reduction of noise is more pronounced (around 60% for the chosen parameter set).

Our analysis goes beyond the first and second moments (means and variances). In particular, for the mRNA inactivation loop model (3), we show that its steady state distribution is generated by Taylor-expanding the explicit function

$$G(x_1, x_2, z) = \exp \left(\frac{\lambda^m \lambda^p}{\gamma_{eff}^m \gamma^p} \int_1^z {}_2F_2 \left(\begin{matrix} 1, 1 + \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (s - 1) \right) ds \right. \\ \left. + \frac{\lambda^m (x_1 - 1)}{\gamma_{eff}^m} {}_2F_2 \left(\begin{matrix} 1, 1 + \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (z - 1) \right) \right. \\ \left. + \frac{q_{12} \lambda^m (x_2 - 1)}{\gamma_{eff}^m (\gamma_2^m + q_{21})} {}_2F_2 \left(\begin{matrix} 1, \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (z - 1) \right) \right).$$

The conjugate variables x_1 , x_2 , and z correspond to mRNA, imRNA, and protein species, respectively. Parameters r_1 , r_2 , and τ are parameter groupings defined by (53) and (55), respectively. The symbol ${}_2F_2$ stands for

$${}_pF_q \left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; \tilde{z} \right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{\tilde{z}^n}{n!}$$

which is the generalised hypergeometric function [33].

Our mathematical analysis thus provides a complete characterisation of the steady state distribution in the mRNA inactivation model in particular (as well as the generalised model in general), and extends the generating function result previously given for the two stage model in [18].

Conclusions

In this paper, we formulated and analysed a structuration/generalisation of the two-stage gene expression model in terms of having multiple mRNA states. Unlike the classical two-stage model, the generalised model considers multiple mRNA states, among which mRNA molecules are assumed to be transitioning at constant rates. Additionally, we demonstrated that the generalised model can be used to capture the dynamics of simpler models such as the inactivation loop model and the multiphasic mRNA model, which were analysed in detail as a particular interest of this paper.

We first introduced the corresponding chemical reaction system describing the generalised model and its mathematical description given by the CME. Then, we focused on seeking a solution to the corresponding PDE, which is obtained by transforming the CME using the generating function approach. A suitable ansatz was employed for converting the PDE to a system of ODEs. Subsequently, using the power series method, we sought a solution to the ODE system, which is then expressed in matrix form as a system of recurrence equations. We recovered the generating function of the stationary distribution of mRNA and protein amounts by means of the coefficients of power series, which are obtained by solving the recurrence relations under the initial conditions.

Furthermore, the sought-after solution was then used to characterise the marginal protein and mRNA distributions. To determine the protein distribution, we used the factorial moments, which are calculated from the factorial cumulants. Additionally, we demonstrated that the mRNA distributions are Poissonian. Obtaining a Poisson distribution is evident for any monomolecular chemical reaction system [34]; therefore, we derived the protein mean and Fano factor and thus expressed it in terms of the first two factorial moments. We then provided two different examples to which the generalised model and its results can be applied.

The first example concerns the inactivation loop model. We demonstrated that integrating the mRNA inactivation loop into the classical two-stage framework for gene expression results in reduced values of protein noise. Nevertheless, we note that certain conditions on the parameter rates must be met to obtain a significant protein noise reduction. These constraints take different forms depending on the interaction between the mRNA form and the mRNA degradation pathway. The first option is that the formation of the inactivation loop does not interfere with degradation, meaning that the inactive form degrades with the same rate constant as the active form. The second option is that the formation of the loop interferes with degradation so that its instantaneous degradation rate becomes zero. In both cases, protein stability must be

optimally chosen to maximise noise reduction; the protein can be neither too stable nor too unstable. However, if inactive mRNAs are subject to degradation, noise reduction is optimised for relatively low protein stabilities, whereas if inactive mRNAs are protected from degradation, noise reduction is optimised for more realistic, larger values of protein stabilities. For inactive mRNAs that do not degrade, optimal noise reduction requires low mRNA activation rates, whereas relatively fast rates of activation optimise noise reduction if inactive mRNAs degrade. Generally, the stability of the inactive mRNA form sustains greater reductions of protein noise for wider and more realistic parameter values. Overall, the noise analysis suggests that the mRNA inactivation loop may play a role in controlling gene expression noise, while also highlighting the limitations of its effect. It is worth noting that one can also compare the protein variance between the extended and canonical two-stage models using the mRNA autocovariance function [35]. The approach taken in this work has an additional advantage that we present a notably non-trivial distribution for protein, which is expressed in terms of the generalised hypergeometric series and is employed to obtain a recursive expression for the protein probability mass function.

As a second example, by making suitable parameter choices in the generalised model, we presented the multiphasic model in which an mRNA molecule is assumed to be transitioning through its lifetime stages. The solution obtained for the generalised model and the associated matrices (e.g., the transition matrix) were used to determine the first two moments of mRNA distributions, which allowed us to calculate the Fano factor for the multiphasic model.

We provided a biological example of the formation of RNA stem loops and performed data analysis to explain the influence of stem-loop structure on gene expression noise. Specifically, we based our extensive mathematical analysis on the two standard noise metrics: the CV² and the Fano factor. By doing so, our calculations allowed us to conclude that noise in gene expression can be reduced if stem loops restrict translation.

In summary, the paper provides a systematic mathematical analysis for protein–mRNA interactions in a structured gene expression model. We believe that the model and its results can be used in understanding the dynamics of underlying biochemical processes.

Methods

Model formulation

For the two-stage gene expression model (1), the probability $p_{m,n}(t)$ of observing m mRNA and n protein molecules at time t satisfies the CME

$$\begin{aligned} \frac{d}{dt} p_{m,n} = & \lambda^m (p_{m-1,n} - p_{m,n}) + \gamma^m ((m+1)p_{m+1,n} - mp_{m,n}) \\ & + \lambda^p m (p_{m,n-1} - p_{m,n}) + \gamma^p ((n+1)p_{m,n+1} - np_{m,n}), \end{aligned} \tag{11}$$

subject to initial condition

$$p_{m,n}(0) = \delta_{m,m_0} \delta_{n,n_0},$$

where $\delta_{i,j}$ represents the Kronecker delta symbol, which is one if $i = j$ and zero otherwise; m_0 and n_0 are the initial mRNA and protein amounts, respectively.

Our aim is to obtain a PDE rather than working with the CME (11). To this end, we introduce the probability generating function defined by

$$G(x, y, t) = \sum_m \sum_n x^m y^n p_{m,n}(t). \tag{12}$$

Multiplying the CME (11) by the factor $x^m y^n$ and summing over all m and n , and using (12), we arrive at the generating function which satisfies the linear first-order PDE

$$\frac{\partial G}{\partial t} = (\gamma^m(1-x) + \lambda^p x(y-1)) \frac{\partial G}{\partial x} + \gamma^p(1-y) \frac{\partial G}{\partial y} + \lambda^m(x-1)G. \tag{13}$$

Equation (13) has been used in [24] to derive mRNA and protein moments; it has been solved at steady state in [18]. Here we shall derive and study a generalisation of (13).

Without loss of generality, for the generalised model (2), the probability $P(\mathbf{m}, n, t)$ of observing m_1 mRNA copies in state 1, m_2 mRNA copies in state 2, and so on, at given time t satisfies the following CME,

$$\begin{aligned} \frac{dP(\mathbf{m}, n, t)}{dt} = & \sum_{i=1}^K \left(\lambda_i^m (\mathbb{E}_i^{-1} - 1)P + \gamma_i^m (\mathbb{E}_i - 1)m_i P + \sum_{j=1}^K q_{ij} (\mathbb{E}_i \mathbb{E}_j^{-1} - 1) \right. \\ & \left. \times m_i P + \lambda_i^p (\mathbb{E}_{K+1}^{-1} - 1)m_i P \right) + \gamma^p (\mathbb{E}_{K+1} - 1)nP, \end{aligned} \tag{14}$$

where $\mathbf{m} = [m_1 \ m_2 \ m_3 \ \dots \ m_K]$ is a vector of species copy numbers. Note that the step operator [36] \mathbb{E}_i in (14) is in the variable m_i , whereas \mathbb{E}_{K+1} is in the variable n ; $\mathbb{E}_i \mathbb{E}_j^{-1} - 1 = 0$ for $i = j$.

The multivariate probability generating function is given by

$$G(\mathbf{x}, y, t) = \sum_{m_1} \dots \sum_{m_K} \sum_n P(\mathbf{m}, n, t) x_1^{m_1} x_2^{m_2} \dots x_K^{m_K} y^n, \tag{15}$$

where $\mathbf{x} = [x_1 \ x_2 \ x_3 \ \dots \ x_K]$. Multiplying (14) by $x_1^{m_1} x_2^{m_2} \dots x_K^{m_K} y^n$ and summing over all m_1, m_2, \dots, m_K, n , and employing (15), we arrive at the PDE

$$\begin{aligned} \frac{\partial G(\mathbf{x}, y, t)}{\partial t} = & \sum_{i=1}^K \left(\lambda_i^m (x_i - 1)G + \gamma_i^m (1 - x_i) \frac{\partial G}{\partial x_i} + \sum_{j=1}^K q_{ij} (x_j - x_i) \frac{\partial G}{\partial x_i} \right. \\ & \left. + \lambda_i^p (y - 1)x_i \frac{\partial G}{\partial x_i} \right) + \gamma^p (1 - y) \frac{\partial G}{\partial y}. \end{aligned} \tag{16}$$

Note that the step operators $\mathbb{E}_i^{\pm 1}$ in (14) coincide with the variables $x_i^{\mp 1}$ while the copy number of species m_i correspond to the terms $x_i \partial_{x_i}$ in (16) for the generating function. In the next section, we will seek a solution to the PDE (16).

Solution

In this section, we shall provide a step-by-step breakdown of our solution method for solving the PDE (16). We are interested in the steady state; therefore, we set the time derivative in (16) to zero and rearrange the resulting equation to obtain

$$\sum_{i=1}^K \left(\lambda_i^m (x_i - 1)G + \left(\gamma_i^m (1 - x_i) + \sum_{j=1}^K q_{ij}(x_j - x_i) + \lambda_i^p (y - 1)x_i \right) \frac{\partial G}{\partial x_i} \right) + \gamma^p (1 - y) \frac{\partial G}{\partial y} = 0 \tag{17}$$

for the time-independent generating function $G(\mathbf{x}, y)$ of the stationary distribution. The probability normalisation condition translates to $G(1, \dots, 1) = 1$. Changing the variables according to

$$x_i = 1 + u_i, \quad y = 1 + v, \quad G = \exp(\varphi) \tag{18}$$

allows us to transform (17) into

$$\sum_{i=1}^K \left(\lambda_i^m u_i + \left(\lambda_i^p v(1 + u_i) - \gamma_i^m u_i + \sum_{j=1}^K q_{ij}(u_j - u_i) \right) \frac{\partial \varphi}{\partial u_i} \right) = \gamma^p v \frac{\partial \varphi}{\partial v}, \tag{19}$$

which is subject to the normalisation condition

$$\varphi(\mathbf{0}) = 0. \tag{20}$$

Below, we focus on seeking a solution to (19)–(20) using a suitable ansatz.

Let us first consider that the solution is of the form

$$\varphi(u_1, u_2, u_3, \dots, u_K, v) = \varphi_0(v) + u_1 \varphi_1(v) + \dots + u_K \varphi_K(v). \tag{21}$$

With this in mind, we obtain from (21) that

$$\frac{\partial \varphi}{\partial u_i} = \varphi_i(v), \quad \frac{\partial \varphi}{\partial v} = \varphi'_0(v) + u_1 \varphi'_1(v) + \dots + u_K \varphi'_K(v). \tag{22}$$

Inserting the partial derivatives (22) into (19), we get

$$\sum_{i=1}^K \left(\lambda_i^m u_i + \left(\lambda_i^p v(1 + u_i) - \gamma_i^m u_i + \sum_{j=1}^K q_{ij}(u_j - u_i) \right) \varphi_i - \gamma^p v u_i \varphi'_i \right) = \gamma^p v \varphi'_0. \tag{23}$$

Equation (23) can be rewritten as

$$\left(\gamma^p \varphi'_0 - \sum_{i=1}^K \lambda_i^p \varphi_i \right) v + \sum_{i=1}^K \left(\gamma^p v \varphi'_i + \left(\gamma_i^m - \lambda_i^p v + \sum_{j=1}^K q_{ij} \right) \varphi_i - \sum_{j=1}^K q_{ji} \varphi_j - \lambda_i^m \right) u_i = 0. \tag{24}$$

In order that (24) hold, we must necessarily have

$$\sum_{i=1}^K \lambda_i^p \varphi_i - \gamma^p \varphi'_0 = 0, \tag{25}$$

$$\gamma^p v \varphi'_i + \left(\gamma_i^m - \lambda_i^p v + \sum_{j=1}^K q_{ij} \right) \varphi_i - \sum_{j=1}^K q_{ji} \varphi_j = \lambda_i^m. \tag{26}$$

Thus far, we have converted the system of PDEs (17) into the system of ODEs (25)–(26). Next, we provide a detailed explanation of solving this system using the power series method.

Let us assume that the functions φ_0 and φ_i are of the power series form, i.e.,

$$\varphi_0(v) = \sum_{n=0}^{\infty} a_n v^n, \quad \varphi_i(v) = \sum_{n=0}^{\infty} b_n^{(i)} v^n \tag{27}$$

for $i \in \{1, \dots, K\}$. Differentiating (27) term by term we get

$$\varphi_0'(v) = \sum_{n=1}^{\infty} n a_n v^{n-1}, \quad \varphi_i'(v) = \sum_{n=1}^{\infty} n b_n^{(i)} v^{n-1}. \tag{28}$$

Inserting (27) and (28) into (26), and collecting same powers of v , we obtain the following system of recurrence relations

$$\left(\gamma_i^m + \sum_{j=1}^K q_{ij} + n \gamma^p \right) b_n^{(i)} - \sum_{j=1}^K q_{ji} b_n^{(j)} = \lambda_i^p b_{n-1}^{(i)} \tag{29}$$

for the coefficients $b_n^{(i)}$, where $i = 1, \dots, K$. For the sake of simplicity, equations (29) can be rewritten in matrix form as

$$(\mathbf{A} - \mathbf{Q}^\top + n \gamma^p \mathbf{I}) X_n = \mathbf{B} X_{n-1}, \quad n \geq 1, \tag{30}$$

where \mathbf{I} is the identity matrix and the vector X_n is defined as

$$X_n = [b_n^{(1)}, b_n^{(2)}, b_n^{(3)}, \dots, b_n^{(K)}]^\top.$$

In (30), \mathbf{A} is a $K \times K$ matrix defined by

$$\mathbf{A}_{ij} := \begin{cases} \gamma_i^m & \text{for } i = j, \\ 0 & \text{for } i \neq j, \end{cases} \tag{31}$$

\mathbf{Q} is a $K \times K$ matrix defined by

$$\mathbf{Q}_{ij} := \begin{cases} -\sum_{k \neq i} q_{ik} & \text{for } i = j, \\ q_{ij} & \text{for } i \neq j, \end{cases} \tag{32}$$

and \mathbf{B} is a $K \times K$ matrix defined by

$$\mathbf{B}_{ij} := \begin{cases} \lambda_i^p & \text{for } i = j, \\ 0 & \text{for } i \neq j. \end{cases} \tag{33}$$

In order to solve the recurrence relations (30) initial conditions are needed. These can be obtained from (26) by setting $v = 0$ for each $i \in \{1, 2, \dots, K\}$. The resulting system of linear equations is given in matrix form as

$$(\mathbf{A} - \mathbf{Q}^\top) X_0 = C, \tag{34}$$

where C is a column vector defined as $C = [\lambda_1^m \ \lambda_2^m \ \dots \ \lambda_K^m]^\top$.

Solving the system of algebraic equations (30) under the initial conditions (34) yields the terms of $b_n^{(i)}$; the sequence a_n can be obtained by substituting (27) and (28) into (25) and collecting same powers of v . By doing so, we get

$$a_n = \frac{1}{n\gamma^p} \sum_{i=1}^K \lambda_i^p b_{n-1}^{(i)}, \quad n \geq 1. \tag{35}$$

Note that the normalisation condition (20) implies that $a_0 = \varphi_0(0) = \varphi(\mathbf{0}) = 0$. Having found the sequences a_n and $b_n^{(i)}$, we combine (21) and (27) to obtain

$$\varphi(u, v) = \sum_{n=1}^{\infty} a_n u^n + \sum_{i=1}^K v_i \sum_{n=0}^{\infty} b_n^{(i)} u^n. \tag{36}$$

We return to the original variables in (36) via (18) to obtain the generating function of the stationary distribution of mRNA and protein amounts, which is given by

$$G(\mathbf{x}, y) = \exp \left(\sum_{n=1}^{\infty} a_n (y-1)^n + \sum_{i=1}^K (x_i - 1) \sum_{n=0}^{\infty} b_n^{(i)} (y-1)^n \right). \tag{37}$$

Equation (37) provides the sought-after steady-state solution to the PDE (16) and will be used in the following section.

Marginal distributions and moments

In this section, we use the analytical formula for the generating function (37) to obtain marginal mRNA distributions. We determine the moments of the protein distribution by way of the factorial cumulants, which allow us to recover the protein distribution. Additionally, we derive the protein Fano factor (variance-to-mean ratio) and express it in terms of the first two factorial moments.

Marginal mRNA distributions In the generating function (37), if we take $y = 1$, then we obtain the marginal mRNA distributions as

$$G^m(\mathbf{x}) = G(\mathbf{x}, 1) = \exp \left(\sum_{i=1}^K b_0^{(i)} (x_i - 1) \right) = \prod_{i=1}^K \exp \left(b_0^{(i)} (x_i - 1) \right), \tag{38}$$

from which we conclude that the steady state mRNA distributions are independent Poissons with means

$$\langle m_i \rangle = b_0^{(i)}.$$

Marginal protein distribution Likewise, by inserting $x_i = 1$ ($i = 1, \dots, K$) into (37), we can recover the generating function of the marginal protein distribution

$$G(y) = G(\mathbf{1}, y) = \exp \left(\sum_{n=1}^{\infty} a_n (y-1)^n \right), \tag{39}$$

where $\mathbf{1}$ is a K -dimensional row vector of ones.

Next, we determine the moments of the protein distributions. The factorial (combinatorial) moments h_n are obtained by expanding the generating function into a power series around $y = 1$:

$$G(y) = \sum_{n=0}^{\infty} h_n (y - 1)^n.$$

We aim to calculate the factorial moments h_n by way of the factorial cumulants a_n . To that end, we first differentiate (39) to obtain

$$DG(y) = G(y)D \ln G(y), \tag{40}$$

where D denotes the differential operator d/dy . Then, taking the $(n - 1)$ th derivative of (40), we get

$$D^n G(y) = \sum_{i=0}^{n-1} \binom{n-1}{i} D^i G(y) D^{n-i} \ln G(y),$$

which can be recast as

$$\frac{D^n G(y)}{n!} = \sum_{i=0}^{n-1} \left(1 - \frac{i}{n}\right) \frac{D^i G(y)}{i!} \frac{D^{n-i} (\ln G(y))}{(n-i)!}. \tag{41}$$

Evaluating (41) at $y = 1$ gives the factorial moments of the protein distribution

$$h_n = \sum_{i=0}^{n-1} \left(1 - \frac{i}{n}\right) a_{n-i} h_i, \quad \text{for } n \geq 1, \tag{42}$$

where $h_0 = 1$. The terms of h_n can be recursively obtained by inserting (35) into (42). Subsequently, by employing the recurrence method proposed in [37], we recover the protein distribution

$$p(n) = \sum_{j=1}^{\infty} \frac{(j+1)_n}{n!} h_{n+j} (-1)^j,$$

where $(x)_n$, n being a nonnegative integer, denotes the rising factorial or namely Pochhammer symbol.

Moments Clearly, the mRNA distributions in (38) are Poissonian. Therefore, mRNA Fano factor is equal to 1. The protein mean and Fano factor can be derived from the factorial moments (42). The first two factorial moments are given by

$$\langle n \rangle = h_1 = a_1 \quad \text{and} \quad \langle n(n-1) \rangle = 2h_2 = 2a_2 + a_1^2, \tag{43}$$

respectively. The Fano factor,

$$F = \frac{\langle n^2 \rangle}{\langle n \rangle} - \langle n \rangle = \frac{\langle n(n-1) \rangle}{\langle n \rangle} + 1 - \langle n \rangle = \frac{2a_2}{a_1} + 1, \tag{44}$$

is thus expressed in terms of the first two factorial cumulants a_1 and a_2 .

The mRNA inactivation loop model

In this section, we present a particular example of the generalised model (2), which we refer to as the *inactivation loop model*, whose reaction scheme is given by (3). Specifically, we provide an explicit representation of the stationary solution using the cumulants. Furthermore, we calculate the steady-state protein Fano factor and express it as a function of the model parameters. Let us note that a possible biological scenario that can implement this model is by a regulatory RNA that temporarily blocks mRNA function [38].

The inactivation loop model (3) can be readily obtained from the generalised model (2) by taking $K = 2$, which accounts for only two mRNA states denoting the active mRNA state m_1 and the inactive mRNA state m_2 . In what follows, we assume that a newly produced mRNA is active, i.e. that the transcription rate satisfies

$$\lambda_i^m = \lambda^m \delta_{i,1}, \quad \text{for } i = 1, 2. \tag{45}$$

Additionally, we assume that proteins are translated only from an active mRNA, so that we have

$$\lambda_i^p = \lambda^p \delta_{i,1}, \quad \text{for } i = 1, 2, \tag{46}$$

for the translation rate. Here, $\delta_{i,j}$ denotes the Kronecker delta symbol. *Cumulants* We aim to recover expressions for the inactivation loop model from the generalised model. The system of algebraic equations for this model follows from (34), taking the form of

$$\begin{aligned} (\gamma_1^m + q_{12})b_0^{(1)} - q_{21}b_0^{(2)} &= \lambda^m, \\ (\gamma_2^m + q_{21})b_0^{(2)} - q_{12}b_0^{(1)} &= 0, \end{aligned}$$

from which we recover

$$b_0^{(1)} = \frac{\lambda^m(\gamma_2^m + q_{21})}{(\gamma_1^m + q_{12})(\gamma_2^m + q_{21}) - q_{12}q_{21}}. \tag{47}$$

Combining (47) with (39) we find

$$\langle m_1 \rangle = \frac{\lambda^m}{\gamma_{\text{eff}}^m},$$

where

$$\gamma_{\text{eff}}^m = \gamma_1^m + \frac{q_{12}\gamma_2^m}{\gamma_2^m + q_{21}} \tag{48}$$

is the effective rate of mRNA decay. The recurrence relations (30) read

$$(\gamma_1^m + q_{12} + n\gamma^p)b_n^{(1)} - \lambda^p b_{n-1}^{(1)} - q_{21}b_n^{(2)} = 0, \tag{49}$$

$$(\gamma_2^m + q_{21} + n\gamma^p)b_n^{(2)} - q_{12}b_n^{(1)} = 0, \tag{50}$$

for $n \geq 1$. Solving the algebraic system (49)–(50) in $b_n^{(1)}$ yields

$$b_n^{(1)} = \frac{\lambda^p(\gamma_2^m + q_{21} + n\gamma^p)}{\gamma^{p^2}n^2 + \gamma^p(\gamma_2^m + \gamma_1^m + q_{21} + q_{12})n + \gamma_2^m\gamma_1^m + \gamma_1^mq_{21} + \gamma_2^mq_{12}}b_{n-1}^{(1)}, \tag{51}$$

which is a recursive expression whose first term (i.e. zeroth) is given by (47).

Explicit representation The recursive formula (51) can further be simplified by factorising its denominator as

$$b_n^{(1)} = \lambda^p \frac{\gamma_2^m + q_{21} + n\gamma^p}{\gamma^{p^2}(n + r_1)(n + r_2)}b_{n-1}^{(1)} \quad \text{for } n \geq 1, \tag{52}$$

where

$$r_{1,2} = \frac{\gamma_1^m + q_{12} + \gamma_2^m + q_{21} \pm \sqrt{(\gamma_2^m + q_{21} - \gamma_1^m - q_{12})^2 + 4q_{21}q_{12}}}{2\gamma^p} \tag{53}$$

are the opposite numbers to the roots of the quadratic in the denominator of (51). The sequence (52) can be rewritten as

$$b_n^{(1)} = \frac{\lambda^m(1 + \tau)_n}{\gamma_{\text{eff}}^m(1 + r_1)_n(1 + r_2)_n} \left(\frac{\lambda^p}{\gamma^p}\right)^n, \quad n \geq 1, \tag{54}$$

where

$$\tau = \frac{\gamma_2^m + q_{21}}{\gamma^p}, \tag{55}$$

and $(x)_n$ represents the rising factorial. Thus, a_n can be obtained from (35) as

$$a_n = \frac{\lambda^p}{n\gamma^p}b_{n-1}^{(1)}, \quad n \geq 1. \tag{56}$$

Inserting (54) into (56) gives

$$a_n = \frac{\lambda^m r_1 r_2}{\gamma_{\text{eff}}^m \tau} \frac{(\tau)_n}{n(r_1)_n(r_2)_n} \left(\frac{\lambda^p}{\gamma^p}\right)^n, \quad n \geq 1, \tag{57}$$

and, similarly, inserting (54) into (50) gives

$$b_n^{(2)} = \frac{q_{12}\lambda^m(\tau)_n}{\gamma_{\text{eff}}^m(\gamma_2^m + q_{21})(1 + r_1)_n(1 + r_2)_n} \left(\frac{\lambda^p}{\gamma^p}\right)^n, \quad n \geq 1. \tag{58}$$

Substituting (54), (57), and (58) into (37), we obtain an explicit representation of the stationary solution

$$G(x_1, x_2, z) = \exp \left(\frac{\lambda^m \lambda^p}{\gamma_{\text{eff}}^m \gamma^p} \int_1^z {}_2F_2 \left(\begin{matrix} 1, 1 + \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (s - 1) \right) ds \right. \\ \left. + \frac{\lambda^m (x_1 - 1)}{\gamma_{\text{eff}}^m} {}_2F_2 \left(\begin{matrix} 1, 1 + \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (z - 1) \right) \right. \\ \left. + \frac{q_{12} \lambda^m (x_2 - 1)}{\gamma_{\text{eff}}^m (\gamma_2^m + q_{21})} {}_2F_2 \left(\begin{matrix} 1, \tau \\ 1 + r_1, 1 + r_2 \end{matrix}; \frac{\lambda^p}{\gamma^p} (z - 1) \right) \right),$$

where

$${}_pF_q \left(\begin{matrix} a_1, \dots, a_p \\ b_1, \dots, b_q \end{matrix}; \tilde{z} \right) = \sum_{n=0}^{\infty} \frac{(a_1)_n \dots (a_p)_n}{(b_1)_n \dots (b_q)_n} \frac{\tilde{z}^n}{n!}$$

is the generalised hypergeometric function [33]. Furthermore, combining (47) and (56) yields an equivalent expression

$$\langle n \rangle = \frac{\lambda^p \lambda^m (\gamma_2^m + q_{21})}{\gamma^p ((\gamma_1^m + q_{12})(\gamma_2^m + q_{21}) - q_{12}q_{21})} = \frac{\lambda^p \lambda^m}{\gamma^p \gamma_{\text{eff}}^m}$$

for the protein mean given by (43) in terms of the model parameters. Likewise, substituting (56) and (51) into (44) and simplifying gives

$$F = 1 + \frac{b_n^{(1)}}{b_n^{(0)}} = 1 + \frac{\lambda^p}{\gamma^p + \gamma_1^m + \frac{q_{12}(\gamma^p + \gamma_2^m)}{\gamma^p + \gamma_2^m + q_{21}}} \tag{59}$$

for the steady-state protein Fano factor as function of the model parameters.

Multiphasic mRNA lifetime

In this section, we consider that mRNA molecules possess $K > 2$ stages of their lifetime, where the transition rates correspond to the ageing of an mRNA molecule. The chemical reaction system for this multiphasic model was given in (4). We note that kinetic proof reading cascades can be an interesting application of our multiphasic model [39].

By (4), there are K stages of an mRNA's molecule lifetime, each of which lasts $1/K \gamma_{\text{eff}}^m$ on average. The total mRNA lifetime is then $1/\gamma_{\text{eff}}^m$; γ_{eff}^m is thereby interpreted as the effective mRNA decay rate. The multiphasic mRNA decay in K steps leads to an Erlang-distributed lifetime with mean $1/\gamma_{\text{eff}}^m$ and variance $1/(\gamma_{\text{eff}}^m)^2$, whereas the lifetime distribution is exponential in the standard model (1).

The multiphasic model (4) can be obtained by making the following choices in the general model statement (2):

$$\lambda_i^m = \begin{cases} \lambda^m & \text{for } i = 1, \\ 0 & \text{for } i \neq 1, \end{cases}$$

and

$$\gamma_i^m = \begin{cases} K\gamma_{\text{eff}}^m & \text{if } i = K, \\ 0 & \text{otherwise.} \end{cases}$$

The transition matrix **Q** (32) for the multiphasic model takes the form of

$$\mathbf{Q} = K\gamma_{\text{eff}}^m \begin{pmatrix} -1 & 1 & & & \\ & -1 & 1 & & \\ & & \ddots & \ddots & \\ & & & -1 & 1 \\ & & & & 0 \end{pmatrix}, \tag{60}$$

and the matrix **A** (31) is given by

$$\mathbf{A} = K\gamma_{\text{eff}}^m \begin{pmatrix} 0 & & & & \\ & 0 & & & \\ & & \ddots & & \\ & & & 0 & \\ & & & & 1 \end{pmatrix}. \tag{61}$$

Inserting (61) and (60) into (34), we obtain the system of recurrence equations

$$K\gamma_{\text{eff}}^m \begin{pmatrix} 1 & & & & \\ -1 & 1 & & & \\ & -1 & \ddots & & \\ & & \ddots & 1 & \\ & & & -1 & 1 \end{pmatrix} \begin{pmatrix} b_0^{(1)} \\ b_0^{(2)} \\ \vdots \\ b_0^{(i)} \\ \vdots \\ b_0^{(K)} \end{pmatrix} = \begin{pmatrix} \lambda^m \\ 0 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{62}$$

from which, upon taking the *i*-th row of (62) and solving the recursive equations

$$-K\gamma_{\text{eff}}^m b_0^{(i-1)} + K\gamma_{\text{eff}}^m b_0^{(i)} = 0, \quad \text{for } 2 \leq i \leq K,$$

where $b_0^{(1)} = \lambda^m / K\gamma_{\text{eff}}^m$, we recover

$$b_0^{(i)} = \frac{\lambda^m}{K\gamma_{\text{eff}}^m}. \tag{63}$$

Formula (63) gives the mean of mRNA molecule in the *i*-th state of its lifetime. Note that the matrix **B** (33) takes the form of $\mathbf{B} = \lambda^p \mathbf{I}$, where **I** is the identity matrix.

Having found the first moments (i.e. means) (63), we then determine the second moments. Taking $n = 1$ in (30), we have

$$\begin{pmatrix} K\gamma_{\text{eff}}^m + \gamma^p & & & & \\ -K\gamma_{\text{eff}}^m & K\gamma_{\text{eff}}^m + \gamma^p & & & \\ & -K\gamma_{\text{eff}}^m & \ddots & & \\ & & \ddots & K\gamma_{\text{eff}}^m + \gamma^p & \\ & & & -K\gamma_{\text{eff}}^m & K\gamma_{\text{eff}}^m + \gamma^p \end{pmatrix} \begin{pmatrix} b_1^{(1)} \\ b_1^{(2)} \\ \vdots \\ b_1^{(i)} \\ \vdots \\ b_1^{(K)} \end{pmatrix} = \frac{\lambda^p \lambda^m}{K\gamma_{\text{eff}}^m} \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{pmatrix}, \tag{64}$$

from which we obtain the first term of the sequence $b_1^{(i)}$ as

$$b_1^{(1)} := u = \frac{\lambda^p \lambda^m}{K \gamma_{\text{eff}}^m (K \gamma_{\text{eff}}^m + \gamma^p)}. \tag{65}$$

Equation (64) implies that

$$-K \gamma_{\text{eff}}^m b_1^{(i-1)} + (K \gamma_{\text{eff}}^m + \gamma^p) b_1^{(i)} = \frac{\lambda^p \lambda^m}{K \gamma_{\text{eff}}^m}, \quad \text{for } 2 \leq i \leq K,$$

which can equivalently be rewritten as

$$b_1^{(i)} = u + v b_1^{(i-1)}, \quad 2 \leq i \leq K, \tag{66}$$

where we set

$$v = \frac{K \gamma_{\text{eff}}^m}{K \gamma_{\text{eff}}^m + \gamma^p} \tag{67}$$

for simplicity. Combining (66) and (65), we obtain

$$b_1^{(i)} = \frac{u}{1-v} + v^{i-1} \left(u - \frac{u}{1-v} \right), \quad 1 \leq i \leq K, \tag{68}$$

from which all the elements of $b_1^{(i)}$ (thereby the second moments) can be iteratively obtained. It is worth noting that one can derive higher moments using formula (30), but we limit our study to the first two moments.

Next, we focus on calculating the first two terms of the sequence a_n (35). Setting $n = 1, 2$ in (35) and inserting (63) and (68) into the resulting equations, respectively, we get

$$a_1 = \frac{\lambda^p \lambda^m}{\gamma^p \gamma_{\text{eff}}^m} \quad \text{and} \quad a_2 = \frac{\lambda^p u (K + v(-1 - K + v^K))}{2 \gamma^p (1-v)^2}. \tag{69}$$

Having found the first two terms of a_n , we are now ready to calculate the Fano factor. Inserting (69) into (44), and substituting (65) and (67) into the resulting expression yields

$$F_m = 1 + \frac{\lambda^p}{\gamma^p} \left(1 + \frac{\gamma_{\text{eff}}^m}{\gamma^p} \left(-1 + \left(\frac{K \gamma_{\text{eff}}^m}{K \gamma_{\text{eff}}^m + \gamma^p} \right)^K \right) \right),$$

where F_m stands for the multiphasic Fano factor.

Abbreviations

mRNA	Messenger RNA
CME	Chemical master equation
PDE	Partial differential equation
ODE	Ordinary differential equation
imRNA	Inactive mRNA
UTR	Untranslated region
CV	Coefficient of variation
SD	Standard deviation
yEGFP	Yeast-enhanced green fluorescent protein
ymNeonGreen	Yellow monomeric fluorescent protein

Acknowledgements

AS acknowledges support by ARO W911NF-19-1-0243 and NIH grants R01GM124446 and R01GM126557.

About this supplement

This article has been published as part of BMC Bioinformatics Volume 24 Supplement 1, 2023: Special Issue of the 19th International Conference on Computational Methods in Systems Biology. The full contents of the supplement are available online at <https://bmcbioinformatics.biomedcentral.com/articles/supplements/volume-24-supplement-1>.

Author contributions

CÇ performed simulations. CÇ and PB analysed and interpreted the results. AS conceived the research. CÇ, PB, and AS wrote the manuscript. All authors have read and approved the final manuscript.

Funding

This work has been supported by the Slovak Research and Development Agency under the contract No. APVV-18-0308 and the VEGA grants 1/0339/21 and 1/0755/22. The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

Not applicable.

Declarations**Ethical approval and consent to participate**

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no Conflict of interest.

Received: 9 March 2022 Accepted: 18 September 2024

Published online: 22 October 2024

References

1. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002;297(5584):1183–6.
2. Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Science*. 2012;336(6078):183–7.
3. Raser JM, O’Shea EK. Noise in gene expression: origins, consequences, and control. *Science*. 2005;309(5743):2010–3.
4. Sanchez A, Choubey S, Kondev J. Regulation of noise in gene expression. *Annu Rev Biophys*. 2013;42:469–91.
5. Dar RD, Shaffer SM, Singh A, Razoooky BS, Simpson ML, Raj A. Explains transcriptional bursting, the noise-versus-mean relationship in mRNA and protein levels. *PLOS ONE*. 2016;11(7):e158298. <https://doi.org/10.1371/journal.pone.0158298>.
6. Kim S, Jacobs-Wagner C. Effects of mRNA degradation and site-specific transcriptional pausing on protein expression noise. *Biophys J*. 2018;114(7):1718–29.
7. Fraser LCR, Dikdan RJ, Dey S, Singh A, Tyagi S. Reduction in gene expression noise by targeted increase in accessibility at gene loci. *Proceed Nat Acad Sci*. 2021;118(42):e2018640118. <https://doi.org/10.1073/pnas.2018640118>.
8. Modi S, Dey S, Singh A. Noise suppression in stochastic genetic circuits using PID controllers. *PLOS Comput Biol*. 2021;17(7):1–25. <https://doi.org/10.1371/journal.pcbi.1009249>.
9. Smith M, Soltani M, Kulkarni R, Singh A. Modulation of stochastic gene expression by nuclear export processes. In: 2021 60th IEEE Conference on Decision and Control (CDC). 2021;655–60.
10. Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceed Nat Acad Sci*. 2002;99(20):12795–800.
11. Thomas P. Intrinsic and extrinsic noise of gene expression in lineage trees. *Sci Rep*. 2019;9(1):474.
12. Dacheux E, Malys N, Meng X, Ramachandran V, Mendes P, McCarthy JEG. Translation initiation events on structured eukaryotic mRNAs generate gene expression noise. *Nucleic Acids Res*. 2017;45(11):6981–92.
13. Chiaruttini C, Guillier M. On the role of mRNA secondary structure in bacterial translation. *Wiley Interdiscip Rev: RNA*. 2020;11(3):e1579.
14. Roy B, Jacobson A. The intimate relationships of mRNA decay and translation. *Trends Genet*. 2013;29(12):691–9.
15. Bokes P, King JR, Wood ATA, Loose M. Transcriptional bursting diversifies the behaviour of a toggle switch: hybrid simulation of stochastic gene expression. *Bull Math Biol*. 2013;75(2):351–71.
16. Kurasov P, Mugnolo D, Wolf V. Analytic solutions for stochastic hybrid models of gene regulatory networks. *J Math Biol*. 2021;82(1):1–29.
17. Singh A, Hespanha JP. Stochastic hybrid systems for studying biochemical processes. *Philosoph Trans Royal Soc A: Math, Phys Eng Sci*. 1930;2010(368):4995–5011.
18. Bokes P, King JR, Wood ATA, Loose M. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J Math Biol*. 2012;64(5):829–54. <https://doi.org/10.1007/s00285-011-0433-5>.

19. Shahrezaei V, Swain PS. Analytical distributions for stochastic gene expression. *Proceedings of the National Academy of Sciences*. 2008
20. Peccoud J, Ycart B. Markovian modeling of gene-product synthesis. *Theor Popul Biol*. 1995;48(2):222–34.
21. Pendar H, Platini T, Kulkarni RV. Exact protein distributions for stochastic models of gene expression using partitioning of Poisson processes. *Phys Rev E*. 2013;87(4):042720.
22. Schnoerr D, Sanguinetti G, Grima R. Approximation and inference methods for stochastic biochemical kinetics—a tutorial review. *J Phys A: Math Theor*. 2017;50(9):093001.
23. Bartman CR, Hamagami N, Keller CA, Giardine B, Hardison RC, Blobel GA, et al. Transcriptional burst initiation and polymerase pause release are key control points of transcriptional regulation. *Mol cell*. 2019;73(3):519–32.
24. Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proceed Nat Acad Sci*. 2001;98(15):8614–9.
25. Li J, Ge H, Zhang Y. Fluctuating-rate model with multiple gene states. *J Math Biol*. 2020;81(4):1099–141.
26. Zhou T, Liu T. Quantitative analysis of gene expression systems. *Quant Biol*. 2015;3(4):168–81.
27. Szavits-Nossan J, Grima R. Mean-field theory accurately captures the variation of copy number distributions across the mRNA life cycle. *Phys Rev E*. 2022;105:014410. <https://doi.org/10.1103/PhysRevE.105.014410>.
28. Filatova T, Popović N, Grima R. Modulation of nuclear and cytoplasmic mRNA fluctuations by time-dependent stimuli: Analytical distributions. *Math Biosci*. 2022;347:108828.
29. Çelik C, Bokes P, Singh A. Protein noise and distribution in a two-stage gene-expression model extended by an mRNA inactivation loop. In: Cinquemani E, Paulevé L, editors. *Computational Methods in Systems Biology*. Cham: Springer International Publishing; 2021. p. 215–29.
30. Fan R, Hilfinger A. The effect of microRNA on protein variability and gene expression fidelity. *Biophys J*. 2023;122(5):905–23.
31. Paulsson J. Summing up the noise in gene networks. *Nature*. 2004;427(6973):415–8.
32. Singh A, Bokes P. Consequences of mRNA transport on stochastic variability in protein levels. *Biophys J*. 2012;103(5):1087–96.
33. Abramowitz M, Stegun IA, Romer RH. Handbook of mathematical functions with formulas, graphs, and mathematical tables. *Am J Phys*. 1988;56(10):958–8.
34. Jahnke T, Huisinga W. Solving the chemical master equation for monomolecular reaction systems analytically. *J Math Biol*. 2007;54(1):1–26.
35. Warren PB, Tănase-Nicola S, ten Wolde PR. Exact results for noise power spectra in linear biochemical reaction networks. *J Chem Phys*. 2006;25(14):144904.
36. Kampen NGV. *Stochastic Processes in Physics and Chemistry*. 3rd ed. North Holland: Elsevier; 2007.
37. Ham L, Schnoerr D, Brackston RD, Stumpf MPH. Exactly solvable models of stochastic gene expression. *J Chem Phys*. 2020;152(14):144106. <https://doi.org/10.1063/1.5143540>.
38. Rodríguez Martínez M, Soriano J, Tlusty T, Pilpel Y, Furman I. Messenger RNA fluctuations and regulatory RNAs shape the dynamics of a negative feedback loop. *Phys Rev—E Stat, Nonlinear, Soft Matter Phys*. 2010;81(3):031924.
39. Hopfield JJ. Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proceed Nat Acad Sci*. 1974;71(10):4135–9.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.