Control of noise in gene expression by transcriptional reinitiation

Rajesh Karmakar

Ramakrishna Mission Vidya Mandira, Belur Math, Belur, Howrah, West Bengal, India
E-mail: rkarmakar@vidyamandira.ac.in

Received 8 July 2019
Accepted for publication 23 March 2020
Published 15 June 2020

Abstract. Gene expression is a random or noisy process. The process consists of several random events, among which the reinitiation of transcription by RNA polymerase (RNAP) is an important one. The RNAP molecules can bind the gene only after the promoter gets activated by transcription factors. Several transcription factors bind the promoter to put the gene in the active state. The gene turns into the inactive state as the bound transcription factors leave the promoter. During the active period of the gene, many RNAP molecules transcribe the gene to synthesize messenger RNA (mRNA). The binding of RNAP to the active state of the gene is a probabilistic process and therefore introduces noise or fluctuations in the mRNA and protein levels. In this paper, we analytically calculate the Fano factor in mRNA and protein levels, and also the probability distribution of mRNA numbers exactly with the binding event of RNAPs in the gene transcription process. The analytically calculated expression of the Fano factor of proteins shows excellent agreement with an experimental result. Then we show that the Fano factor in mRNA levels can be sub-Poissonian due to the reinitiation of transcription by RNAP, and the mean mRNA level can be increased without increasing the Fano factor. Our study show that the Fano factor can also be reduced keeping mRNA levels fixed. We find that the reinitiation of transcription can behave as a fine-tuned control process to regulate the mRNA/protein level in the cell.

Keywords: stochastic gene expression, reinitiation of transcription, Fano factor
Gene expression and its regulation is an inherently stochastic or noisy process, and this stochasticity produces cell-to-cell variation in messenger RNA (mRNA) and protein levels. The variation or fluctuations in mRNA and protein levels is qualitatively known as noise. Genetically identical individuals can be very different even if they grow in identical environmental conditions, because of noisy gene expression. There are a large number of theoretical and experimental studies at the single-cell level that establish the stochastic nature of gene expression [1–16]. The effect of noise or fluctuations in gene expression can be beneficial or detrimental to the progression of cellular activities [5, 9, 17–20]. Variability in mRNA and protein levels can be beneficial to adapt to sudden changes in environmental conditions [17]. The noise in gene expression can also generate phenotypic diversity in genetically identical cells and may play an important role in the survival of cell colonies [5]. Fluctuations in mRNA and protein levels can also have a detrimental role in the functioning of fine-tuned cellular processes [19]. Different studies show that the cellular system has its own mechanism to reduce noise or fluctuations in gene expression. There is a study showing that noise levels in proteins from essential genes are lower than those from nonessential genes [21]. It has also been shown that organisms could evolve to minimize the variability in gene expression by adopting a special technique, e.g. high transcription rate and low translation rate. The noise strength does not remain fixed throughout the life of a cell but rather decreases during the normal aging of a cell [22]. Many other processes exist by which the noise in gene expression can be reduced [23, 24].

Gene expression and regulation consists of several important biochemical steps, e.g. transcription, translation, degradation etc. The regulation can be achieved at the transcription or translation or degradation level, and the control of transcription is the dominant form of regulation of gene expression. Transcriptional regulation is mostly done by DNA binding proteins known as transcription factors. A group of transcription factors (TFIIA, TBP and associated factors) bind the promoter sequences on the
DNA and turn the gene into the active state or intermediate promoter complex \[25, 26\]. The transcription factors that modulate the transitions to the active (inactive) state are known as activators (repressors). A gene can either be active or inactive in the cell, and random transitions take place between the two states \[6–11\]. In the active state, the RNA polymerase (RNAP) proteins bind the gene and form a preinitiation complex (including RNAP). The bound RNAP then initiates the transcription of the protein-coding gene to synthesize mRNAs \[4, 5\]. The RNAP molecules transcribe the gene several times and synthesize a burst of mRNA. That is known as reinitiation of transcription by RNAP. The initiation of transcription by RNAP molecules is an important part of transcriptional regulation \[27\]. The control of transcription at the initiation stage is a key mechanism for the regulation of gene expression.

There are some gene regulatory systems where activators and repressors are both involved in transcriptional regulation \[4, 5, 12, 13\]. It has been shown that the gene can be in more than two states under the regulation of activators and repressors \[4, 5, 13, 29\]. It has also been shown that the stochastic effects due to random transitions between active and inactive states of a gene are much stronger than the stochastic effects caused by random production and degradation of single mRNA and protein molecules \[4, 5, 7, 9, 10, 29, 30\]. In the course of time, different model systems have been developed to explain the experimentally observed variability in gene expression. The simplest model is constitutive gene expression, in which a gene is always in the active state and mRNAs are synthesized at a constant rate from that state \[2, 11, 30, 31\]. In such a case, the fluctuations in mRNA level arise from the random birth and death of mRNAs. The study of cell-to-cell variation in mRNA and protein levels at the single-cell level is mostly done by assuming two states of a gene under the regulation of activators \[6–11, 14, 15\]. In the two-state model, the randomness in mRNA levels comes from the random transitions between the gene states and from the random birth and death of mRNAs from the active gene \[15, 30, 31\]. But the stochasticity due to reinitiation of transcription by RNAP is ignored in both constitutive and two-state gene activation models. Blake et al experimentally studied the synthetic GAL1* yeast promoter and identified a regulatory mechanism using stochastic simulations that agrees well with their experimental data \[4, 5\]. They observed that the pulsatile mRNA production through reinitiation is crucial for the dependence of noise or fluctuations on transcriptional efficiency. It has been shown that the reinitiation also enhances the reliability of transcriptional response in eukaryotic systems \[28\]. Sanchez and Kondiev studied the GAL1* yeast promoter also and analytically calculated the various moments neglecting the reinitiation step in gene transcription \[29\]. It has been identified that the reinitiation of transcription by RNAP plays a crucial role in the phenotypic variability in cells, but its functional role in noisy mRNA or protein levels is very poorly understood.

The stochasticity in gene expression is generally quantified by calculating the variance or coefficient of variation (CV) or Fano factor \[2–4, 7, 15, 29–31\]. mRNA synthesis from a gene constitutively is a Poisson process with unity Fano factor (FF). The two-state model gives rise to an increased FF in mRNA level (FF\(_m\)) due to random transitions between gene states. FF\(_m\) now becomes super-Poissonian (greater than unity) \[15, 30\]. It is observed that only negative feedback can reduce FF\(_m\) and move it into the sub-Poissonian (less than unity) regime \[14\]. With the two-state gene activation model, we
cannot think of the regime of sub-Poissonian Fano factor at the mRNA level without negative feedback in gene expression. In this article, we show that the reinitiation of transcription by RNAP in two-state gene activation process can move FF\textsubscript{m} into the sub-Poissonian regime. We observe that the mean mRNA level can be increased without increasing FF\textsubscript{m} by controlling the reinitiation of transcription. We also observe that FF\textsubscript{m} can be decreased to sub-Poissonian regime keeping the average mRNA level fixed. Again, the knowledge of FF may not always give complete information about the proteins. The distribution of mRNA/protein level helps to identify the more important information about responses of gene expression. Obtaining an exact analytical expression for the distribution of mRNA/protein has been a challenging task, and that always gives some extra insights. There are many works that have already been done on the exact solution of the distribution of mRNA and protein levels with and without feedback but excluding the reinitiation process [8, 11, 13, 14]. We also find the exact steady-state distribution of mRNA levels with reinitiation of transcription by RNAP.

2. The model and analysis

2.1. The model

We consider the gene regulation model where only activators regulate gene transcription. In this model, the activators bind the promoter region of the gene and turn the gene into the active state (G\textsubscript{2}) from the inactive state (G\textsubscript{1}). Once the gene turns into the active state, the RNAP then binds it and forms an initiation complex (G\textsubscript{3}). Then bound RNAP can follow two paths: either it unbinds and leaves the G\textsubscript{3} state or it starts transcription. As it starts transcription, it moves forward along the gene and the G\textsubscript{3} state turns into G\textsubscript{2} again. Another RNAP can bind the gene again to form an initiation complex and then start transcription. So, RNAP can bind multiple times in the active state of the gene, and a burst of mRNAs are synthesized before the gene turns into the inactive state (G\textsubscript{1}) from the active state (G\textsubscript{2}). The biochemical reactions and the corresponding rate constants are shown in figure 1. The model considered here is a part of the network identified in the synthetic GAL\textsubscript{1}\textsuperscript{*} promoter by Blake et al [4]. The GAL\textsubscript{1}\textsuperscript{*} promoter is regulated by both activator and repressor, and their concentrations are further controlled by external inducer GAL and ATC respectively. In an experiment, the activation (repression) probability of the promoter increases (decreases) with increasing GAL (ATC) concentration [4]. The GAL\textsubscript{1}\textsuperscript{*} promoter may behave as an activator-only system with the full induction of ATC ([ATC] = 500 ng ml\textsuperscript{-1}). Thus, the promoter architecture proposed by them resembles our model given in figure 1 with full ATC induction.

2.2. Analysis

Let us consider that \(l\) copies of a particular gene exist in the cell. Let \(p(l_2, l_3, m, n, t)\) be the probability that at time \(t\) there are \(m\) mRNAs and \(n\) protein molecules with \(l_2\) genes in the active state (G\textsubscript{2}) and \(l_3\) genes in the initiation complex (G\textsubscript{3}). The number
of genes in the inactive state is $l_1 = l - l_2 - l_3$ as the number of gene copies is conserved. The time evaluation of the probability is given by the Master equation \[32\]

\[
\frac{\partial p(l_2, l_3, m, n, t)}{\partial t} = k_1[(l - l_2 - l_3 + 1)p(l_2 - 1, l_3, m, n, t) - (l - l_2 - l_3)p(l_2, l_3, m, n, t)] \\
+ k_2[(l_2 + 1)p(l_2 + 1, l_3, m, n, t) - l_2p(l_2, l_3, m, n, t)] \\
+ k_3[(l_2 + 1)p(l_2 + 1, l_3 - 1, m, n, t) - l_2p(l_2, l_3, m, n, t)] \\
+ k_4[(l_3 + 1)p(l_2 - 1, l_3 + 1, m, n, t) - l_3p(l_2, l_3, m, n, t)] \\
+ J_m[(l_3 + 1)p(l_2 - 1, l_3 + 1, m - 1, n, t) - l_3p(l_2, l_3, m, n, t)] \\
+ k_m[(m + 1)p(l_2, l_3, m + 1, n, t) - mp(l_2, l_3, m, n, t)] \\
+ J_p[mp(l_2, l_3, m, n - 1, t) - mp(l_2, l_3, m, n, t)] \\
+ k_p[(n + 1)p(l_2, l_3, m, n + 1, t) - np(l_2, l_3, m, n, t)]
\] (1)

The standard generating function technique is now applied for the expression of the means and variances of mRNA and protein in the steady state \[32\].

The expressions for the mean and FF of mRNAs ($\langle m \rangle$, $\text{FF}_m$) and proteins ($\langle p \rangle$, $\text{FF}_p$) in terms of the rate constants (figure 1) in the steady state are given by (for $l = 1$)

\[
\langle m \rangle = \frac{k_1k_3J_m}{a_2k_m}, \quad \text{FF}_m = 1 + \frac{J_mk_3(a_2 - k_1a_1)}{a_2(a_1k_m + a_2)}
\] (2)

\[
\langle p \rangle = \frac{\langle m \rangle J_p}{k_p}, \quad \text{FF}_p = 1 + \frac{J_p}{k_p + k_m} + \frac{J_pJ_mk_3(a_4 + a_1)}{k_m(k_p + k_m)a_4} + \frac{J_pJ_mk_3(a_6 - k_1k_pa_8)}{k_p(k_p + k_m)a_5} - \frac{J_pJ_mk_1k_3}{k_mk_pk_a2}
\] (3)

where $a_1 = k_m + J_m + k_1 + k_2 + k_3 + k_4$, $a_2 = k_1J_m + k_2J_m + k_3k_4 + k_1k_3 + k_1k_4$, $a_3 = k_m + J_m + k_1$, $a_4 = k_1J_m + J_m + k_1$, $a_5 = a_4\{a_2 + k_pa_1 - k_p - a_1\}$, $a_6 = (k_pa_7 + k_1)a_4 + k_pa_8(J_m + k_1)$, $a_7 = a_5(k_m + k_1)/(a_1k_m + a_2)$ and $a_8 = k_5(k_m + k_1)/(a_1k_m + a_2)$.
Control of noise in gene expression by transcriptional reinitiation

Figure 2. (a) Plot of $\text{FF}_p$ with the transcriptional efficiency. The rate constants are $k_1 = 0.02 + 0.2\text{[GAL]}$, $k_2 = 0.01 + 0.1\text{[GAL]} + 0.077/\text{[GAL]}$, $k_3 = 50.0$, $k_4 = 10.0$, $J_m = 1.0$, $k_m = 1.0$, $J_p = 5.0$, $k_p = 0.0125$ with reinitiation (red solid line) (figure 1) and without reinitiation (red dotted line) [15]. The hollow circles (blue) are the experimental data points with full ATC induction [4]. The inset shows the variation of mean protein number with GAL concentration with reinitiation (red solid line) and without reinitiation (red dotted line). (b) Plot of relative percentage error of $\text{FF}_p$ with the transcriptional efficiency. This shows the excellent agreement between the experimental data and analytically calculated curve for intermediate and higher values of transcriptional efficiency and a little disagreement at lower values.

Blake et al measured $\text{FF}_p$ as a function of transcriptional efficiency by varying the inducer [GAL] (0% to 2%) with full induction of [ATC] (500 ng ml$^{-1}$) and also by varying [ATC] with full induction of [GAL] (2%) [4]. They found maximum $\text{FF}_p$ at intermediate values of transcriptional efficiency. We compared our analytical result (equation (3)) with the experimental observations of Blake et al with full [ATC] induction. The plot of $\text{FF}_p$ against transcriptional efficiency along with the experimental data points is shown in figure 2(a). We also plot the mean protein level $\langle p \rangle$ against GAL concentration (inset) with the same rate constants as given in Blake et al [4]. We also plot $\text{FF}_p$ against transcriptional efficiency and $\langle p \rangle$ against GAL concentration (inset) without the reinitiation of transcription by RNAP (red dotted curves in figure 2(a)). We plot the percentage relative error between the experimental data points and analytically calculated values against transcriptional efficiency in figure 2(b). That shows the excellent agreement between the experimental data in Blake et al [4] and our analytically calculated curve for intermediate and higher values of transcriptional efficiency, and a little disagreement at the lower values of transcriptional efficiency. The initial disagreement may be due to the presence of repressors in the experimental system. We also observe the large deviation of $\text{FF}_p$ and $\langle p \rangle$ from the experimental data points without the reinitiation of transcription by RNAP.

The transcriptional reinitiation by RNAP affects $\text{FF}$ at the mRNA level first. Post-transcription processes then further amplify that effect. So, we study $\text{FF}_m$ to explore the role of reinitiation of transcription. In the expression for $\text{FF}_m$ in equation (2), the first term arises due to the random birth and death of mRNA molecules and the second term arises due to the random transitions between different gene states, $G_1$, $G_2$ and $G_3$. The diagram shows the relationship between these processes and the resulting fluctuations in mRNA levels.

https://doi.org/10.1088/1742-5468/ab8382
Equation (2) shows that the reinitiation can drive \( FF_m \) to the sub-Poissonian level depending on the condition, given by

\[
 a_2 < k_1 a_1 \text{ or } (J_m + k_1) < \frac{k_1}{k_2}(k_1 + k_2 + k_m).
\]  

There is a critical value \( J_m^c \) for the given value of \( k_1, k_2 \) and \( k_4 \), which is given by

\[
 J_m^c = \frac{k_1}{k_2}(k_1 + k_2 + 1) - k_4.
\]  

The transcriptional reinitiation process in figure 1 is mainly controlled by the rate constants \( k_3 \) and \( J_m \). We observe the variation of \( FF_m \) with \( k_3 \) for the rate constants \( k_1 = 6.02 \) and \( k_2 = 3.012 \) (i.e. high GAL concentration (30%)) with \( J_m \) as parameter (figure 3(a)). For the rate constant considered in figure 3, the critical value of \( J_m \) is \( J_m^c = 10.048 \), for which \( FF_m \) is Poissonian (green solid line in figure 3(a)). For \( J_m \) greater than \( J_m^c \), \( FF_m \) always lies in the super-Poissonian phase (blue dashed line for \( J_m = 13.048 \) in figure 3(a)) whereas for \( J_m \) less than \( J_m^c \) \( FF_m \) always lies in the sub-Poissonian phase (red dotted line for \( J_m = 1.048 \) in figure 3(a)). The green solid line (\( FF_m = 1 \)) in figure 3(a) clearly separates the sub-Poissonian and super-Poissonian phases. The degree of deviation of \( FF_m \) towards the sub-Poissonian phase increases with decreasing \( k_2 \) and \( k_4 \). The variation of \( FF_p \) is plotted against \( k_3 \) with the above-mentioned three different values of \( J_m \). Translation in gene expression produces a burst of proteins from each mRNA. That bursting process adds some extra fluctuations to the protein levels (equation (3)). The straight solid line (for \( J_m = 10.048 \)) and curved dot-dashed line (for \( J_m = 8.048 \)) in figure 3(a) become curved and straight respectively in figure 3(b) due to translation in protein synthesis. In the previous models and calculations, \( FF_m \) is always found to be greater than unity [9, 15, 29, 30]. But, with the inclusion of reinitiation processes in gene transcription, we observe that two distinct phases, namely super-Poissonian and sub-Poissonian Fano factors separated by a Poissonian one, are possible. Depending on the values of different rate constants, \( FF_m \) can be in any one of the phases.

The rate constants for gene expression for different genes are not unique. The different literature on gene expression indicated different values of rate constants for different steps. Figures 4(a) and (b) show the more intense variation of \( FF_m \) with \( k_3 \) for the rate constants given in Kaern et al [9]. We also calculate \( FF_m \) and \( FF_p \) from stochastic
Control of noise in gene expression by transcriptional reinitiation

Figure 4. Plot of Fano factors (a) $F_{F_m}$ and (b) $F_{F_p}$ against the rate constant $k_3$ and $J_m$ as parameter. Other rate constants are $k_1 = 10.0$, $k_3 = 10.0$, $k_4 = 1.0$, $k_m = 0.1$, $J_p = 0.2$ and $k_p = 0.05$. The rate constants are taken from Kaern et al [9], except $k_1$ which is chosen arbitrarily. The value of $J_m$ are: 30.0 (blue dashed), 19.1 (green solid) and 10.0 (red dotted). $J_m = 19.0$ for the (cyan dash-dot) straight line in the inset in (b). The hollow circles in (c) and (d) are generated from stochastic simulation using the Gillespie algorithm corresponding to the curves in (a) and (b).

Simulation using the Gillespie algorithm with the rate constants corresponding to the curves in figures 4(a) and (b) [33]. Figures 4(c) and (d) show that our analytical results agree well with the simulation results.

To observe the clearer variation of $F_{F_m}$ with other rate constants over a wide range, we use surface plots with different rate constants (figures 5–7). Figures 5–7 show that higher values of $k_1$ and lower values of $k_2$, $k_4$ and $J_m$ favour sub-Poissonian $F_{F_m}$. At a very high value of $k_1$, $F_{F_m}$ becomes independent of $k_2$ (figure 6(a)). We see from equation (5) that the critical value of $J_m$ increases with increasing rate constant $k_1$ and decreases with rate constants $k_2$ and $k_4$. That is reflected in figures 5(b) and 7(b). We observe from equation (2) that $k_1$, $k_3$ and $J_m$ play a crucial role in the determination of mean mRNA level. Figure 8 shows that the mean mRNA level does not increase much with increasing $k_1$ and $k_3$ but increases more with $J_m$. Therefore, the variation of $k_1$ or $k_3$ is sufficient to change the mean mRNA level by a small amount but the variation of $J_m$ is necessary for a greater change. This behaviour is similar to ‘coarse’ and ‘fine’ control knobs in electronic devices. Now, if $k_1$ and $J_m$ are increased (keeping $J_m < J_m^c$) at the same time as the mean mRNA number, $F_{F_m}$ moves to the sub-Poissonian regime (figures 5(b) and 8(b)).

In the experiment of Blake et al, the rate constants $k_1$ and $k_2$ are function of GAL concentration. $k_1$ increases monotonically with increasing GAL concentration but $k_2$ initially decreases (up to 3%) and then increases with increasing GAL concentration (beyond 3%). Figure 9(a) shows that the average mRNA level does not increase with increasing GAL concentration beyond 2%. In figure 9(b), we see that $F_{F_m}$ can be sub-Poissonian with different values of GAL concentration with the given reinitiation rate constant $J_m$. Higher GAL concentration is required for sub-Poissonian $F_{F_m}$ for higher $J_m$. That is, $F_{F_m}$ can be decreased without changing the mean mRNA level by varying GAL concentration.
Control of noise in gene expression by transcriptional reinitiation

Figure 5. The variation of $FF_m$ (a) with $J_m$ and $k_3$ for the rate constants $k_1 = 10.0, k_2 = 10.0, k_4 = 1.0$ and (b) with $J_m$ and $k_1$ for $k_2 = 10.0, k_3 = 50.0, k_4 = 1.0$.

Figure 6. The variation of $FF_m$ (a) with $k_1$ and $k_2$ for the rate constants $k_3 = 50, k_4 = 1, J_m = 19$ and (b) with $k_2$ and $k_3$ for the rate constants $k_1 = 10, k_4 = 1, J_m = 19$.

Figure 7. The variation of $FF_m$ (a) with $k_2$ and $k_4$ for the rate constants $k_3 = 50, k_1 = 10, J_m = 19$ and (b) with $J_m$ and $k_2$ for the rate constants $k_1 = 10, k_3 = 50, k_4 = 1$.

Concentration only (keeping other rate constants fixed). Sanchez and Kondev observed that FF can be decreased substantially with fixed mean by varying [GAL] only beyond 2% in GAL1* promoter [29].

https://doi.org/10.1088/1742-5468/ab8382
Figure 8. Mean mRNA number varies with (a) $k_1$ and $k_3$ (other rate constants are $J_m = 5$, $k_2 = 10$, $k_4 = 1$), (b) $k_1$ and $J_m$ (other rate constants are $k_2 = 10$, $k_3 = 50$, $k_4 = 1$) and (c) $k_3$ and $J_m$ (other rate constants are $k_1 = 10$, $k_2 = 10$, $k_4 = 1$).

Figure 9. (a) Variation of mean mRNA ($\langle m \rangle$) and (b) FF$_m$ with [GAL] and $J_m$. The other rate constants are $k_3 = 50.0$, $k_4 = 1.0$.

3. Steady-state probability density function for mRNA

For a single copy of the gene ($l = 1$), equation (1) can be decomposed into three different equations corresponding to the three different gene states:

\[
\frac{\partial p_1(m, t)}{\partial t} = k_2 p_2(m, t) - k_1 p_1(m, t) + k_m[(m + 1)p_1(m + 1, t) - mp_1(m, t)]
\]

(6)

\[
\frac{\partial p_2(m, t)}{\partial t} = k_1 p_1(m, t) + k_4 p_3(m, t) - (k_2 + k_3)p_2(m, t) + J_m p_3(m - 1, t) + k_m[(m + 1)p_2(m + 1, t) - mp_2(m, t)]
\]

(7)

\[
\frac{\partial p_3(m, t)}{\partial t} = k_3 p_2(m, t) - k_4 p_3(m, t) - J_m p_3(m, t) + k_m[(m + 1)p_3(m + 1, t) - mp_3(m, t)]
\]

(8)

where $p_i(m, t)$ ($i = 1, 2, 3$) is the probability that the gene is in the state $G_i$ ($i = 1, 2, 3$) at time $t$ and the number of mRNA molecules is $m$. 

https://doi.org/10.1088/1742-5468/ab8382
Now using the standard generating function technique [13, 32], we have the total probability density function from the Master equations (6) and (7):

\[
p(m) = \frac{(-h_4)^m \Gamma(h_1 + m) \Gamma(h_2 - h_3) \Gamma(h_2 + h_3)}{\Gamma(m + 1) \Gamma(h_1) \Gamma(h_2 - h_3 + m) \Gamma(h_2 + h_3 + m)} p \times F_q[\{h_1 + m\}, \{h_2 - h_3 + m, h_2 + h_3 + m\}, h_4]
\]  

(9)

where \( h_1 = \frac{n_1}{k_1} \), \( h_2 = -\frac{1}{2} + \frac{n_2}{k_2} \), \( h_3 = \frac{1}{2} \sqrt{1 - 2s_1 + s_1^2 - 4s_2} \), \( h_4 = s_3 \) and \( s_1 = 1 + J_m + k_1 + k_2 + k_3 + k_4 \), \( s_2 = k_3J_m + k_2J_m + k_3k_4 + k_1k_2k_4 + k_1k_3k_4 + k_2k_3k_4 \), \( s_3 = -J_m k_3k_4 \), \( s_4 = -J_m k_1k_3k_4 \).

The plot of \( p(m) \) versus \( m \) is shown in figure 10(a) (figure 10(b)) corresponding to the different FF\(_m\) regimes shown in figure 3(a) (figure 4(a)). In figure 10(b), the red dotted curve (\( J_m = 1.0 \)) and the cyan dot-dashed curve (\( J_m = 10.0 \)) are both in sub-Poissonian FF\(_m\) but with different mean values. The dot-dashed curve for \( J_m = 10.0 \) has higher mean but lower FF\(_m\). The mean mRNA level can thus be increased, and at the same time FF\(_m\) can be decreased, with reinitiation in gene transcription.

In figure 11(a), we draw two curves by changing the GAL concentrations only for same mean (\( \langle m \rangle = 15.5 \)) but with different FF\(_m\). The red dotted curve is for 10\% [GAL] with super-Poissonian FF\(_m\) and the blue dashed curve is for 60\% [GAL] with sub-Poissonian FF\(_m\). In figure 11(b), we see that mean mRNA level can be increased without increasing FF\(_m\). The red dotted curve, blue dashed curve and cyan dot-dashed curve have the same FF\(_m\) but with increasing mean. The green solid curve in figure 11(b) has higher mean but lower FF\(_m\) than the red dotted and blue dashed curves. That kind of behaviour of mean and FF can also be observed with lower values of GAL concentration but in a narrow range of \( k_3 \) and \( J_m \). With higher GAL concentrations, \( J_m \) becomes high and FF\(_m\) become sub-Poissonian over a wide range of \( k_3 \) and \( J_m \).
4. Conclusion

Noise or stochasticity in gene expression produces fluctuations in mRNA and protein levels. The fluctuations in protein levels can corrupt the quality of intracellular signals, thereby affecting cellular functions negatively. Different studies show that the cellular system has different mechanisms to control the inherent stochasticity. Negative feedback is one such mechanism by which noise (FF) reduction takes place [14]. The network architecture, consisting of four regulatory genes, helps to reduce FF and CV during aging [22]. A coherent feed-forward network with three genes can also show the lowest variance level in proteins [34]. In this study, we observed that the reinitiation of transcription by RNAP can also reduce FF during gene transcription.

To study the contribution of reinitiation in the variability of mRNA and protein levels, we considered a simple two-state model with reinitiation of RNAP (figure 1) and calculated the Fano factor in mRNA and protein levels using simple mathematics. Our model network is similar to the regulatory network of GAL1* yeast promoter with full ATC induction [4]. So, we compared our analytically calculated FF of protein levels with the experimental results of Blake et al at full ATC induction and found excellent agreement. Then we analyzed the Fano factor in mRNA levels and observed three different phases, namely, Poissonian, sub-Poissonian and super-Poissonian. We found that there exists a critical value \( J_m^c \) for which FF\(_m\) is always Poissonian. For \( J_m < J_m^c \) (\( J_m > J_m^c \)) FF\(_m\) lies in the sub-Poissonian (super-Poissonian) phase. That result is distinct from the well-studied constitutive and two-state gene activation model. In those two models, FF in mRNA levels is Poissonian and super-Poissonian respectively [15, 30]. The minimum value of FF\(_m\) in the two-state model can be very close to unity but never less than unity.

In our model network, we included the reinitiation step in the two-state gene activation model and found a regime of sub-Poissonian Fano factor in mRNA levels. The reinitiation process therefore removed the extra additive term in FF\(_m\) due to random
transitions between the gene states in the two-state model completely rather decreased it further below unity. That shows that the reinitiation step in gene expression has the strong ability to control the fluctuations in mRNA and protein levels. In the constitutive and two-state model, mRNA synthesis takes place at a constant rate from the active state of the gene. But the reinitiation process in gene transcription adds a refractory period after each mRNA synthesis, and that reduces the effective rate of mRNA synthesis and the variance although the mean mRNA level increases. This results in $FF_m$ going below unity in gene transcription with reinitiation.

In the two-state gene activation model, the activation ($k_1$) and deactivation ($k_2$) rate constants mostly regulate $FF$ in the mRNA level. A fast (slow) transition between the active and inactive states decreases (increases) $FF$ in the mRNA level [6]. Thus, sub-Poissonian $FF_m$ in our model network is favorable with higher values of $k_1$ and $k_2$ (figure 6(a)). It can be shown that the mean mRNA level decreases (increases) with increasing $k_2$ ($k_1$). But with high $k_1$, $k_2$ has the least effect on mean mRNA levels. In the experiment of Blake et al, $k_1$ and $k_2$ are both functions of GAL concentration and can be made high. High $k_1$ and $k_2$ lead to the mean mRNA level going up to almost saturation but FF$_m$ decreasing to the sub-Poissonian regime (figures 9(a) and (b)). We can visualize that result from the distributions of mRNA levels in figure 11(a). Therefore, one can adjust FF$_m$ to the sub-Poissonian regime without changing the mean mRNA by regulating the GAL concentrations in the system.

In our model network, the desired mean and FF$_m$ can also be obtained by choosing the rate constants $k_3$ and $J_m$ properly (figures 8(c), 5(a) and 11(b)). FF$_m$ shows a dip when plotted against $k_3$ and $J_m$ whereas the mean increases with those rate constants (figures 5(a) and 8(c)). Therefore, the mean can be increased keeping FF$_m$ fixed (the red dotted line, blue dashed line and cyan-dot-dashed line) or it can be increased while FF$_m$ decreases (red dotted line, green solid line) (figure 11(b)). Like $k_1$ and $k_2$ in the experiment of Blake et al, the rate constants $k_3$ and $J_m$ are not controlled externally. But a cell can change their values to fulfill its requirements. There are several instances where the cellular system has the ability to adapt to unpredictable environmental changes by regulating its internal states, mean protein levels, fluctuations about the mean levels etc [17, 18, 20–22, 35, 36]. This ability is crucial for survival and proper functioning of cells.

References

[1] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell Science 297 1183–6
[2] Ozbudak E M, Thattai M, Kurtser I, Grossman A D and Oudenaarden A van 2002 Regulation of noise in the expression of single gene Nat. Genet. 31 69–73
[3] Swain P S, Elowitz M B and Siggia E D 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression Proc. Natl Acad. Sci. 99 12795–800
[4] Blake W J, Kaern M, Cantor C R and Collins J J 2003 Noise in eukaryotic gene expression Nature 422 633–7
[5] Blake W J, Balázsi G, Kohanski M A, Isaas F J, Murphy K F, Kuang Y, Cantor C R, Walt D R and Collins J J 2006 Phenotypic consequences of promoter-mediated transcriptional noise Mol. Cell 24 853–65
[6] Karmakar R and Bose I 2004 Graded and binary responses in stochastic gene expression Phys. Biol. 1 197
[7] Raser J M and O’Shea E K 2005 Noise in gene expression: origins, consequences, and control Science 309 2010–3
[8] Horos J E M, Schultz D, Innocentini G C P, Wang J, Walczak A M, Onuchic J N and Wolynes P G 2005 Self regulating gene: an exact solution Phys. Rev. E 72 051907
Control of noise in gene expression by transcriptional reinitiation

[9] Kaern M, Elston T C, Blake W J and Collins J J 2005 Stochasticity in gene expression: from theories to phenotypes Nat. Rev. Genet. 6 451–64

[10] Raj A, Peskin C S, Tranchina D, Vargas D Y and Tyagi S 2006 Stochastic mRNA synthesis in mammalian cells PLoS Biol. 4 e309

[11] Shahrezaei V and Swain P S 2008 Analytical distributions for stochastic gene expression Proc. Natl Acad. Sci. 105 17256–61

[12] Rossi F M V, Kringlestein A M, Spicher A, Guicherit O M and Blau H M 2000 Transcriptional control: rheostat converted to on/off switch Mol. Cell 6 723–8

[13] Karmakar R 2010 Conversion of graded to binary responses in an activator–repressor system Phys. Rev. E 81 021905

[14] Kumar N, Platini T and Kulkarni R V 2014 Exact distribution for stochastic gene expression models with bursting and feedback Phys. Rev. Lett. 113 268105

[15] Paulson J 2005 Models of stochastic gene expression Phys. Life Rev. 2 157–75

[16] Sanchez A, Choubey S and Kondev J 2013 Regulation of noise in gene expression Annu. Rev. Biophys. 42 469–91

[17] Maamar H, Raj A and Dubnau D 2007 Noise in gene expression determines cell fate in bacillus subtilis Science 317 526–9

[18] Acar M, Mettetal J T and Oudenaarden A v 2008 Stochastic switching as a survival strategy in fluctuating environments Nat. Genet. 40 471–5

[19] Magee J A, Abdulkadir S A and Milbrandt J 2003 Haploinsufficiency at the Nkx 3.1 locus. a paradigm for stochastic, dose-sensitive gene regulation during tumour initiation Canc. Cell 3 273–83

[20] Weinberger L S, Burnett J C, Toettcher J E, Arkin A P and Schaffer D V 2005 Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 tat fluctuations drive phenotypic diversity Cell 122 169–82

[21] Fraser H B, Hirsh A E, Giaever G, Kumm J and Eisen M B 2004 Noise minimization in eukaryotic gene expression PLoS Biol. 2 1–5

[22] Liu P, Song R, Elison G L, Peng W and Acar M 2017 Noise reduction as an emergent property of single-cell aging Nat. Commun. 8 1–13

[23] Bundschuh R, Hayot F and Jayaprakash C 2003 The role of dimerization in noise reduction of simple genetic networks J. Theor. Biol. 220 261–9

[24] Struhl K 1999 Fundamentally different logic of gene regulation in eukaryotes and prokaryotes Cell 98 1–4

[25] Liu B, Yuan Z, Aihara K and Chen L 2014 Reinitiation enhances reliable transcriptional responses in eukaryotes J. R. Soc. Interface 11 0326

[26] van Kampen N G 1985 Stochastic Processes in Physics and Chemistry (Amsterdam: North-Holland)

[27] Gillespie D T 1977 Exact stochastic simulation of coupled chemical reactions J. Phys. Chem. 81 2340–61

[28] Ghosh B, Karmakar R and Bose I 2005 Noise characteristics of feed forward loops Phys. Biol. 2 36–45

[29] Briat C, Gupta A and Khammash M 2016 Antithetic integral feedback ensures robust perfect adaptation in noisy biomolecular networks Cell Syst. 2 17–28

https://doi.org/10.1088/1742-5468/ab8382