

Gene Expression in Self-repressing System with Multiple Gene Copies

Jacek Miękiś · Paulina Szymańska

Received: 1 August 2012 / Accepted: 7 January 2013 / Published online: 25 January 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract We analyze a simple model of a self-repressing system with multiple gene copies. Protein molecules may bound to DNA promoters and block their own transcription. We derive analytical expressions for the variance of the number of protein molecules in the stationary state in the self-consistent mean-field approximation. We show that the Fano factor (the variance divided by the mean value) is bigger for the one-gene case than for two gene copies and the difference decreases to zero as frequencies of binding and unbinding increase to infinity.

Keywords Self-repressing gene · Multiple gene copies

1 Introduction

One of the fundamental processes taking part in living cells is regulation of gene expression. It enables cells to differentiate and adapt to a changing environment. Gene expression is a complex process involving many biochemical reactions with proteins being final products. Produced proteins may in turn enhance or repress expression of other proteins. They may also regulate their own expression. Such regulatory networks in cells, from the smallest ones to those very complicated, have been arousing growing interest recently (Beckstein and Serrano 2000; Thattai and van Oudenaarden 2001; Kepler and Elston 2001; Simpson et al. 2003; Lipshtat et al. 2005;

J. Miękiś (✉)

Institute of Applied Mathematics and Mechanics, University of Warsaw, Banacha 2, 02-097 Warsaw, Poland

e-mail: miekisz@mimuw.edu.pl

P. Szymańska

College of Inter-Faculty Individual Studies in Mathematics and Natural Sciences, University of Warsaw, Warsaw, Poland

e-mail: p.m.szymanska@student.uw.edu.pl

Lipniacki et al. 2006; Hat et al. 2007; Komorowski et al. 2009; Loinger and Biham 2009). In many cases, biochemical processes take place in small volumes and may involve only few molecules. Deterministic approach dealing with macroscopic concentrations of molecules (such as ordinary differential equations of classical chemical kinetics) is then inappropriate. A small number of molecules taking part in gene expression results in significant random fluctuations and to take into account such fluctuations, many stochastic models were proposed (Thattai and van Oudenaarden 2001; Swain et al. 2002; Paulsson 2004, 2005).

In many cases, genes exist in several copies (Hat et al. 2007) (and references therein). Understanding the influence of the number of gene copies on the behavior of the system is crucial for designing experiments, which very often involve transfection—introducing an extra copy of the gene with a fluorescent marker in order to observe the evolution of the system. We must take into account that an additional copy of the gene might change the global behavior of the cell. It has been argued in Hat et al. (2007) that the knowledge of how the number of gene copies influences gene expression might lead to a better understanding of experimental data in cancer research. In fact, cancerous cells have, due to mutations, a larger number of gene copies, and thus predictions for tumor's invaded systems are not the same as for healthy ones.

The minimal model of gene expression, that is, of the production of protein molecules in living cells, consists of four fundamental biochemical processes: transcription (production of mRNA molecules), translation (production of protein molecules), and degradation of molecules of both types. One can compute in this model all moments of the number of protein molecules in the stationary states. In particular, a simple formula for the variance was derived in Thattai and van Oudenaarden (2001); see also Swain et al. (2002), Paulsson (2004, 2005) and Paszek (2007). Here, we lump transcription and translation into one process, that is, we use a standard approximation proposed in Kepler and Elston (2001), which is valid if transcription is much faster than translation.

We analyze a simple model of a self-repressing system with one or two gene copies. Protein molecules may bind to DNA promoters and repress their own transcription. We assume here that each gene copy can be in the unbound state or in the bound state with a lower transcription rate. Such an interaction of protein molecules with transcription factors makes the rigorous analysis of the cell dynamics very difficult. In the case of only one gene copy, exact results were obtained recently in Hornos et al. (2005) and Ramos et al. (2011). In particular, a stationary probability distribution of the number of protein molecules was presented as a series involving Kummer functions (Hornos et al. 2005). Time evolution of the probability distribution was considered in Ramos et al. (2011).

Here, we obtain explicit formulas for the variance of the number of protein molecules in the stationary state in the self-consistent mean-field approximation. Such approach, used commonly in statistical physics (Huang 1963; Ma 1985), was introduced recently in gene expression models in Ohkubo (2010); see Miękiś and Szymańska (2012) to compare a mean-field approximation in the Ising model of interacting spins and in a simple model of self-repressing gene. We also discuss two extreme cases: slow switching (binding/unbinding), where to get analytic results we

can use the conditional variance or simply perform an appropriate limit and fast gene switching, where we use an adiabatic approximation. We show analytically that in both extreme cases, the stationary variance of the number of protein molecules coincides with the mean-field approximation. We solved a truncated system of Master equations and showed that the solution agrees with the mean-field approximation for the whole range of the adiabaticity parameter.

The main goal of this paper is to establish how the number of gene copies influences the variance of produced proteins in a simple case of a self-repressing gene. We show that the two-gene system has a lower Fano factor (the variance divided by the mean value than the one-gene regulatory system). The difference disappears when the rate of switching becomes large as compared to production and degradation rates, that is in the adiabatic limit.

In Sect. 2, we analyze one-gene model. Two gene copies are discussed in Sect. 3. Section 4 is devoted to the fast switching gene case, and Sect. 5 to the slow switching one. Conclusions follow in Sect. 6.

2 Self-repressing Gene

Here, we analyze the simplest model of a self-regulating gene. We lump transcription and translation into one process, so we assume that proteins are produced directly out of DNA in one biochemical process (Kepler and Elston 2001). We will discuss here the repression—protein molecules may bind to a certain promoter region of their own DNA, and thus decrease or completely stop the transcription. In continuous models of chemical kinetic equations, the repression is modeled by the modification of a transcription rate, it might be given by a Hill function $h(n) = k/(1 + cn^h)$, where k is the maximal transcription rate, n the number of repressing protein molecules, c and h are constants (Komorowski et al. 2009).

Here, we will consider a stochastic model, where the gene (DNA) can be in two discrete states: unbound (on), denoted by 0 or bound (off), denoted by 1. In the generic case, the transcription rates for the on- and off-states are given by k_0 and k_1 respectively, but we set $k_1 = 0$, as it is often done. The protein degradation rate is denoted by γ . We consider a monomer binding and thus we assume that the binding rate is given by βn , where n is the number of proteins in the system, and the rate of switching the gene on (unbinding) is denoted by α ; see Fig. 1.

Let us introduce formally our model. We denote by $f_i(n, t)$, $i = 0, 1$ the joint probability that there are n protein molecules in the system at time t and the gene (DNA) is in the state i . The standard Master equation (Van Kampen 1997) can be written as:

$$\begin{aligned} \frac{d}{dt} f_0(n, t) &= k_0 [f_0(n-1) - f_0(n)] + \gamma [(n+1)f_0(n+1) - nf_0(n)] \\ &\quad - \beta n f_0(n) + \alpha f_1(n) \\ \frac{d}{dt} f_1(n, t) &= k_1 [f_1(n-1) - f_1(n)] + \gamma [nf_1(n+1) - (n-1)f_1(n)] \\ &\quad + \beta n f_0(n) - \alpha f_1(n) \end{aligned} \tag{1}$$

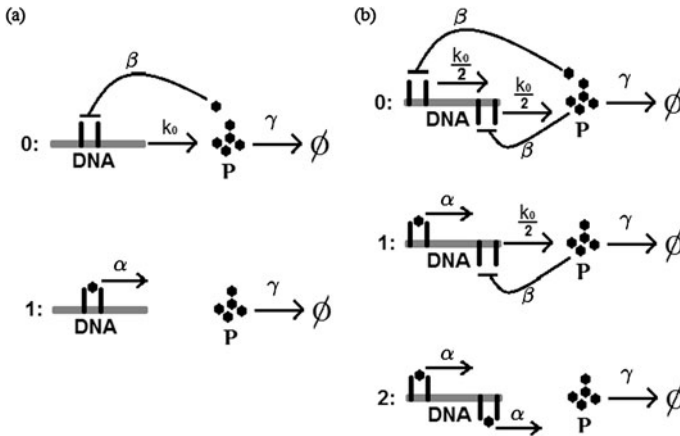


Fig. 1 (a) One-gene system. The protein is produced directly from the DNA in state 0 with the rate k_0 , a single molecule degrades with rate γ . The production is switched off when a protein binds to the promoter region. Binding and unbinding takes place with βn and α rates, respectively. (b) Two-gene system. Molecules bind to promoters of both gene copies independently

for $n \geq 1$.

For $n = 0$ we have $\frac{d}{dt} f_0(0, t) = -k_0 f_0(0) + \gamma f_0(1)$ and $f_1(0, t) = 0$.

Let us emphasize that n is the total number of molecules; one of them is bound to the promoter when the gene state is 1. It follows that $f_1(0, t) = 0$ all the time. We have also assumed that the bound protein cannot degrade. In this respect, our Master equation is different from the one discussed in Hornos et al. (2005); see also Qian et al. (2009).

We denote by (f_0, f_1) a stationary state of our system, that is a solution of (1) with time derivatives set to zero. Let A_0 and A_1 be probabilities (frequencies) that the gene is unbound or bound, respectively, in the stationary state, $A_i = \sum_{n=0}^{+\infty} f_i(n)$, $i = 0, 1$. The stationary expected number of protein molecules with respect to f_i is given by $\langle n \rangle_i = \sum_{n=0}^{+\infty} n f_i(n)$, obviously $\langle n \rangle = \langle n \rangle_0 + \langle n \rangle_1$ is the expected value with respect to $f = f_0 + f_1$. We introduce two generating functions:

$$F_0(z, t) = \sum_{n=0}^{+\infty} z^n f_0(n, t) \quad \text{and} \quad F_1(z, t) = \sum_{n=0}^{+\infty} z^n f_1(n, t).$$

We differentiate generating functions with respect to time, use (1), and after some simplifications, we get

$$\begin{aligned} \frac{\partial F_0(z, t)}{\partial t} &= (z - 1) \left[k_0 F_0(z, t) - \gamma \frac{\partial F_0(z, t)}{\partial z} \right] - \beta \cdot z \frac{\partial F_0(z, t)}{\partial z} + \alpha \cdot F_1(z, t) \\ \frac{\partial F_1(z, t)}{\partial t} &= (z - 1) \left[k_1 F_1(z, t) - \gamma \frac{\partial F_1(z, t)}{\partial z} \right] + F_1(z, t) \left(\gamma - \frac{\gamma}{z} \right) \\ &\quad + \beta \cdot z \frac{\partial F_0(z, t)}{\partial z} - \alpha F_1(z, t) \end{aligned} \tag{2}$$

Now we differentiate the above equations with respect to z once and twice, set $z = 1$, time derivatives to zero, and get the following algebraic equations for the moments of the stationary probability distribution of the number of protein molecules:

$$\begin{cases} A_0 + A_1 = 1 \\ \beta \langle n \rangle_0 - \alpha A_1 = 0 \\ k_0 A_0 - \gamma \langle n \rangle_0 - \beta \langle n^2 \rangle_0 + \alpha \langle n \rangle_1 = 0 \\ k_1 A_1 - \gamma \langle n \rangle_1 + \gamma A_1 + \beta \langle n^2 \rangle_0 - \alpha \langle n \rangle_1 = 0 \\ 2k_0 \langle n \rangle_0 - 2\gamma \langle n(n-1) \rangle_0 - \beta \langle n^2(n-1) \rangle_0 + \alpha \langle n(n-1) \rangle_1 = 0 \\ 2k_1 \langle n \rangle_1 - 2\gamma \langle n(n-1) \rangle_1 + 2\gamma (\langle n \rangle_1 - A_1) + \beta \langle n^2(n-1) \rangle_0 - \alpha \langle n(n-1) \rangle_1 = 0 \end{cases} \quad (3)$$

The above system is hierarchical, equations for lower moments involve higher moments (unlike equations in the classical model of unregulated gene expression analyzed in Thattai and van Oudenaarden 2001). It is not closed (there are more variables than equations) and, therefore, in principle cannot be solved. In order to get explicit formulas for moments, in particular the variance, one has to close somehow the infinite chain of equations. Several concepts and techniques were developed (Nasell 2003; Barzel and Biham 2011; Barzel et al. 2011). Here, we will use the so-called mean-field approximation well known in statistical physics of interacting particles (Huang 1963; Ma 1985; Miękisz and Szymańska 2012) and introduced recently in the context of regulatory genetic systems in Ohkubo (2010). Namely, we replace n in the switching term in (1) by its unknown expected value, that is instead of $\beta n f_0(n)$ we write $\beta \frac{\langle n \rangle_0}{A_0} f_0(n)$. It follows that (3) is replaced by

$$\begin{cases} A_0 + A_1 = 1 \\ \beta \langle n \rangle_0 - \alpha A_1 = 0 \\ k_0 A_0 - \gamma \langle n \rangle_0 - \beta \frac{\langle n \rangle_0}{A_0} \langle n \rangle_0 + \alpha \langle n \rangle_1 = 0 \\ k_1 A_1 - \gamma \langle n \rangle_1 + \gamma A_1 + \beta \frac{\langle n \rangle_0}{A_0} \langle n \rangle_0 - \alpha \langle n \rangle_1 = 0 \\ 2k_0 \langle n \rangle_0 - 2\gamma \langle n(n-1) \rangle_0 - \beta \frac{\langle n \rangle_0}{A_0} \langle n(n-1) \rangle_0 + \alpha \langle n(n-1) \rangle_1 = 0 \\ 2k_1 \langle n \rangle_1 - 2\gamma \langle n(n-1) \rangle_1 + 2\gamma (\langle n \rangle_1 - A_1) + \beta \frac{\langle n \rangle_0}{A_0} \langle n(n-1) \rangle_0 - \alpha \langle n(n-1) \rangle_1 = 0 \end{cases} \quad (4)$$

We obtained a closed system of equations. Let us observe that when one adds the third equation and the fourth one of either (3) or (4), results are the same (switching terms cancel out). The same applies to adding the fifth equation and the sixth one. Hence, independent of approximations, the following relations are always satisfied:

$$\begin{cases} \langle n \rangle = \frac{k_0}{\gamma} A_0 + \frac{k_1}{\gamma} A_1 + A_1 \\ \langle n(n-1) \rangle = \frac{k_0}{\gamma} \langle n \rangle_0 + \frac{k_1}{\gamma} \langle n \rangle_1 + \langle n \rangle_1 - A_1 \end{cases} \quad (5)$$

One can solve (4) (in fact we only need to solve first four equations), obtain the self-consistent value for $\langle n \rangle_i$, $i = 0, 1$, use (5) and $\text{var}(n) = \langle n(n-1) \rangle + \langle n \rangle - \langle n \rangle^2$

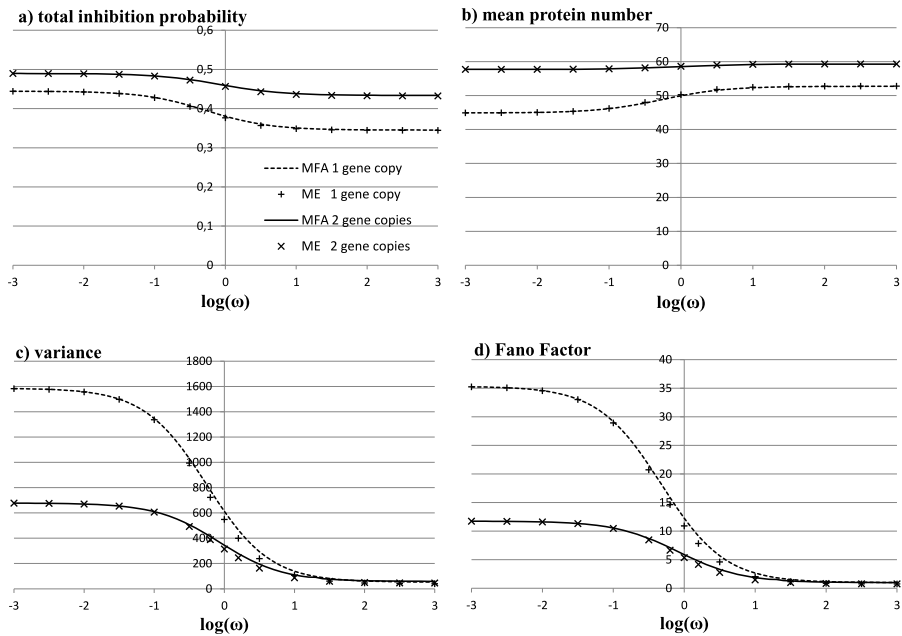


Fig. 2 Total inhibition probability, mean, variance, and Fano Factor of the number of protein molecules in the stationary state, plotted as a function of $\log(\omega)$ for $X^{eq} = 100$ and $X^{ad} = 40$. For the variance, the curves asymptotically approach values computed in the slow-switching case (1585 for the one-gene model and 678 for two gene copies) and in the adiabatic case (52 and 59 for the one and two-gene model, respectively). *Continuous and dotted lines* are obtained analytically within the mean-field approximation, crosses and pluses are points obtained by solving the system of Master equations (Eqs. (1) and (8)) restricted to the maximum of 200 particles

to get the expression for the variance of the number of protein molecules in the stationary state.

Here, we set $k_1 = 0$ and following Hornos et al. (2005) introduce new parameters: $X^{eq} = \frac{\alpha}{\beta}$ —equilibrium constant of the switching process, $X^{ad} = \frac{k_0+k_1}{2\gamma} = \frac{k_0}{2\gamma}$ —measure of protein concentration, and $\omega = \frac{\alpha}{\gamma}$ —adiabaticity parameter. It appears that all equations can be written in terms of these parameters.

From the first four equations of (4), we get the quadratic equation for A_1 ,

$$A_1^2[\omega - 2X^{ad}(\omega + 1) - X^{eq}] + A_1[(4X^{ad} + X^{eq} - 1)\omega + 4X^{ad} + X^{eq}] - 2X^{ad}(\omega + 1) = 0 \tag{6}$$

which has only one positive solution smaller than 1.

Equation (4) allows us to express $\text{var}(n)$ as a function of A_1 ,

$$\text{var}(n) = (2X^{ad} - 1)X^{eq}A_1 + 2[2X^{ad}(1 - A_1) + A_1] - [2X^{ad}(1 - A_1) + A_1]^2 \tag{7}$$

The variance as a function of $\log \omega$ is presented in Fig. 2. We see that the variance is a decreasing function of the switching rate.

We would like to check the validity of the mean-field approximation in two extreme cases: in the limits of the infinitely fast and infinitely slow switching. In the fast-switching case, we divide equations in (3) by α and assume that $\frac{k_i}{\alpha} = \frac{\gamma}{\alpha} = 0$. However, this does not help us in closing the system (3), the number of equations is still too small. It is usually assumed, for example, in Hornos et al. (2005) that in the fast switching case, in the so-called adiabatic limit, one may put $\langle n \rangle_i = A_i \langle n \rangle$, $i = 0, 1$. Such a procedure closes (3). We would like to point out however that this is another approximation and it is not true even in the limit $\alpha, \beta \rightarrow \infty$; see Sect. 4. In the slow-switching case, we assume that for a given gene state, the system attains its stationary state (if $k_1 = 0$, then of course all protein molecules are degraded in the stationary state). In such stationary states, we have from Thattai and van Oudenaarden (2001) formulas for the variance even in the model with transcription and translation; in our simplified model stationary states have the Poisson distribution and so the variance is equal to the expected value. Then we take into account switching between gene states—we simply use the conditional variance formula; see Sect. 5. Alternatively, we may close (3) by dividing equations by k_0 and assuming that $\frac{\alpha}{k_0} = \frac{\beta}{k_0} = 0$, details are shown in Sect. 5. We see in Fig. 2 that the mean field-approximation coincides with the fast-switching solution in the limit of the infinite ω and with the slow-switching one in the limit of zero ω .

To validate the mean-field approximation, we truncated the Master equation (1) by restricting the number of protein molecules to be at most 200. The rigorous solution of the truncated Master equation agrees with the mean-field solution for the whole range of the adiabaticity parameter ω as it can be seen in Fig. 2.

3 Repression with Two Gene Copies

Now we assume that the gene is present in two copies. It follows that the gene system can be in three states: 0, 1, and 2, where 0 means that both promoter sites are unbound, 1 means that exactly one promoter is bound, and 2 that both promoters are bound. Both copies of the gene produce proteins independently. To keep the mean expression approximately at the same level as in the one-gene case, we set $k_1 = \frac{k_0}{2}$ and $k_2 = 0$ so $X_{\text{ad}} = (k_0 + k_1 + k_2)/3\gamma = k_0/2\gamma$ as before. That is we assume that production rates of both genes are set to $\frac{k_0}{2}$. We also made calculations for the production rates of two genes equal to k_0 , they are literally copies of original genes. The mean and the variance are then approximately doubled, but the Fano factor (the variance divided by the mean) remains the same; see Fig. 3.

The Master equation now reads:

$$\begin{aligned} \frac{d}{dt} f_0(n) &= k_0[f_0(n-1) - f_0(n)] + \gamma[(n+1)f_0(n+1) - nf_0(n)] \\ &\quad - \beta n f_0(n) + \alpha f_1(n) \\ \frac{d}{dt} f_1(n) &= \frac{1}{2}k_0[f_1(n-1) - f_1(n)] + \gamma[nf_1(n+1) - (n-1)f_1(n)] \\ &\quad + \beta n f_0(n) - \alpha f_1(n) + 2\alpha f_2(n) - \beta(n-1)f_1(n) \end{aligned} \quad (8)$$

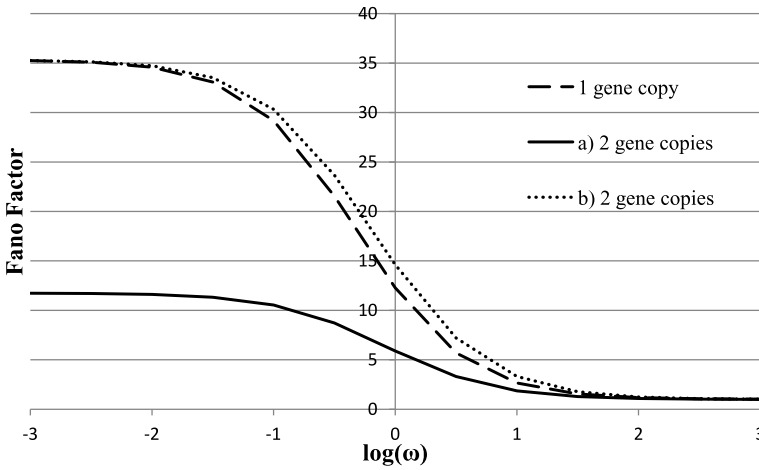


Fig. 3 Fano factor as a function of $\log(\omega)$, $X^{eq} = 100$ and $X^{ad} = 40$ for one-gene model, **(a)** for two-gene model with the production rate $k_0/2$ **(b)** for two-gene model with the production rate k_0

$$\frac{d}{dt} f_2(n) = \gamma[(n - 1)f_2(n + 1) - (n - 2)f_2(n)] - 2\alpha f_2(n) + \beta(n - 1)f_1(n)$$

for $n \geq 2$ and we may write similar equations for $n = 1$ and $n = 0$ with obvious terms not present.

We replace n in the switching term in (8) by its unknown expected value, that is instead of $\beta n f_0(n)$ and $\beta n f_1(n)$ we write $\beta \frac{\langle n \rangle_0}{A_0} f_0(n)$ and $\beta \frac{\langle n \rangle_1}{A_1} f_1(n)$ respectively. We introduce three generating functions, repeat the procedure of the previous section, and get a closed system of equations in the mean-field approximation,

$$\left\{ \begin{array}{l} A_0 + A_1 + A_2 = 1 \\ \beta \langle n \rangle_0 - \alpha A_1 = 0 \\ \beta \langle n \rangle_1 - 2\alpha A_2 - \beta A_1 = 0 \\ k_0 A_0 - \gamma \langle n \rangle_0 - \beta \frac{\langle n \rangle_0^2}{A_0} + \alpha \langle n \rangle_1 = 0 \\ (\frac{1}{2}k_0 + \gamma)A_1 - \gamma \langle n \rangle_1 + \beta \frac{\langle n \rangle_0^2}{A_0} - \alpha \langle n \rangle_1 - \beta \frac{\langle n \rangle_1^2}{A_1} + \beta \langle n \rangle_1 + 2\alpha \langle n \rangle_2 = 0 \\ 2\gamma A_2 - \gamma \langle n \rangle_2 + \beta \frac{\langle n \rangle_1^2}{A_1} - \beta \langle n \rangle_1 - 2\alpha \langle n \rangle_2 = 0 \\ 2k_0 \langle n \rangle_0 - 2\gamma \langle n(n - 1) \rangle_0 - \beta \frac{\langle n \rangle_0}{A_0} \langle n(n - 1) \rangle_0 + \alpha \langle n(n - 1) \rangle_1 = 0 \\ (k_0 + 2\gamma) \langle n \rangle_1 - 2\gamma A_1 - 2\gamma \langle n(n - 1) \rangle_1 + \beta \frac{\langle n \rangle_0}{A_0} \langle n(n - 1) \rangle_0 \\ - \alpha \langle n(n - 1) \rangle_1 - (\beta \frac{\langle n \rangle_1}{A_1} - \beta) \langle n(n - 1) \rangle_1 + 2\alpha \langle n(n - 1) \rangle_2 = 0 \\ 4\gamma \langle n \rangle_2 - 4\gamma A_2 - 2\gamma \langle n(n - 1) \rangle_2 \\ + (\beta \frac{\langle n \rangle_1}{A_1} - \beta) \langle n(n - 1) \rangle_1 - 2\alpha \langle n(n - 1) \rangle_2 = 0 \end{array} \right. \tag{9}$$

We add the fourth equation, the fifth, and the sixth one of (9) and then the last three equations of (9) and again as in the one-gene case we get relations which are

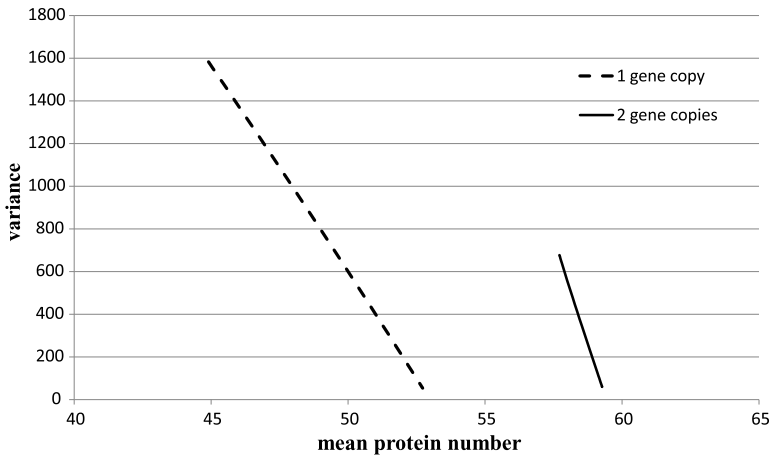


Fig. 4 Variance as a function of the mean, $X^{eq} = 100$ and $X^{ad} = 40$, so that ω is the only changing parameter, $-3 < \log(\omega) < 3$

satisfied independent of approximations:

$$\begin{cases} \langle n \rangle = \frac{k_0}{\gamma} A_0 + \frac{k_0}{2\gamma} A_1 + 2A_2 + A_1 \\ \langle n(n-1) \rangle = \frac{k_0(n)_0 + \frac{1}{2}k_0(n)_1}{\gamma} + 2\langle n \rangle_2 + \langle n \rangle_1 - 2A_2 - A_1 \end{cases} \quad (10)$$

As in the one-gene case, all equations can be expressed in terms of $X^{eq} = \frac{\alpha}{\beta}$, $X^{ad} = \frac{k_0+k_1+k_2}{3\gamma} = \frac{k_0}{2\gamma}$, and $\omega = \frac{\alpha}{\gamma}$. We proceed exactly in the same way as in the one-gene case. We solve the system (9) and get an expression for the probability of total inhibition, the expected value of the number of produced proteins, the variance, and the Fano factor as functions of $\log(\omega)$ in the stationary state; see Fig. 2. We see that the variance and the Fano factor are bigger for the one-gene case than for the two-gene case and that the difference decreases to zero as the rates of gene switching increase. In Fig. 4, we graph the variance as the function of the expected value of the number of proteins as we vary the adiabaticity parameter ω while keeping X^{eq} and X^{ad} fixed. We observe the linear dependence, the slope is bigger in the two-gene case than in the one-gene case.

4 Fast Switching Gene

Here, we consider the situation when gene states are switched infinitely fast. For simplicity, we discuss one-gene case. Let us assume for a moment that there is no self-regulation and the gene is switched between its two states with constant rates: from the state 1 to the state 0 with the rate α and from 0 to 1 with the rate β (not βn with n being the number of protein molecules as in the self-regulating gene case). We will show (as it might be expected) that the expected value of the number of protein molecules in a given state is equal to the expected value of the number of molecules

times the frequency of that state, that is $\langle n \rangle_i = A_i \langle n \rangle$; $i = 0, 1$. As in Sect. 2, $f_i(n)$ are probabilities that there are n protein molecules in the system and the gene is in the state i . Now instead of (1), we have the following Master equation (we do not assume here that $k_1 = 0$):

$$\begin{aligned} \frac{d}{dt} f_0(n, t) &= k_0 [f_0(n - 1) - f_0(n)] \\ &\quad + \gamma [(n + 1) f_0(n + 1) - n f_0(n)] - \beta f_0(n) + \alpha f_1(n) \\ \frac{d}{dt} f_1(n, t) &= k_1 [f_0(n - 1) - f_0(n)] \\ &\quad + \gamma [(n + 1) f_1(n + 1) - (n) f_1(n)] \beta f_0(n) - \alpha f_1(n) \end{aligned} \tag{11}$$

The equations for generating functions (see Sect. 2) now read

$$\begin{aligned} \frac{\partial F_0(z, t)}{\partial t} &= (z - 1) \left[k_0 F_0(z, t) - \gamma \frac{\partial F_0(z, t)}{\partial z} \right] - \beta F_0(z, t) + \alpha F_1(z, t) \\ \frac{\partial F_1(z, t)}{\partial t} &= (z - 1) \left[k_1 F_1(z, t) - \gamma \frac{\partial F_1(z, t)}{\partial z} \right] + \beta F_0(z, t) - \alpha F_1(z, t) \end{aligned} \tag{12}$$

As in Sect. 2, we differentiate the above equations with respect to z once and twice, set $z = 1$, time derivatives to zero, and get the following algebraic equations for the moments of the stationary distributions of the number of protein molecules:

$$\begin{cases} A_0 + A_1 = 1 \\ \beta \langle n \rangle_0 - \alpha A_1 = 0 \\ k_0 A_0 - \gamma \langle n \rangle_0 - \beta \langle n \rangle_0 + \alpha \langle n \rangle_1 = 0 \\ k_1 A_1 - \gamma \langle n \rangle_1 + \beta \langle n \rangle_0 - \alpha \langle n \rangle_1 = 0 \\ 2k_0 \langle n \rangle_0 - 2\gamma \langle n(n - 1) \rangle_0 - \beta \langle n(n - 1) \rangle_0 + \alpha \langle n(n - 1) \rangle_1 = 0 \\ 2k_1 \langle n \rangle_1 - 2\gamma \langle n(n - 1) \rangle_1 + \beta \langle n(n - 1) \rangle_0 - \alpha \langle n(n - 1) \rangle_1 = 0 \end{cases} \tag{13}$$

The above system of equations is closed and it can be solved. In particular, we get $A_0 = \frac{\alpha}{\alpha + \beta}$ and $A_1 = \frac{\beta}{\alpha + \beta}$ (this of course follows immediately from the assumption about constant switching rates) and

$$\langle n \rangle_0 = \frac{k_0(\alpha + \gamma)A_0 + k_1\alpha A_1}{\gamma(\alpha + \beta + \gamma)} \tag{14}$$

$$\langle n \rangle = \frac{k_0 A_0 + k_1 A_1}{\gamma} \tag{15}$$

In the limit of infinitely fast switching, that is when $\frac{k_0, \gamma}{\alpha, \beta} \rightarrow 0$, it follows that $\langle n \rangle_0 = A_0 \langle n \rangle$ and then $\langle n \rangle_1 = A_1 \langle n \rangle$. The two-gene case and in general n -gene case can be treated in the same way and the same conclusion follows.

Gene expression models with constant switching rates were discussed in Paulsson (2004, 2005) and Paszek (2007) and formulas for the variance of the number of protein molecules in the stationary state were derived.

Now we discuss self-repressing genes. It is suggested in Hornos et al. (2005) that also in this case, $\langle n \rangle_i = A_i \langle n \rangle$ in the limit of infinitely fast switching. Let us examine this. The second line in (3) reads

$$A_1 = \frac{\beta \langle n \rangle_0}{\alpha} \quad (16)$$

It might also be written as

$$A_1 = \frac{\beta \langle n \rangle_0}{\alpha A_0 + \alpha A_1} = \frac{\beta \langle n \rangle_0}{\alpha A_0 + \beta \langle n \rangle_0} \quad (17)$$

It is easy to see that $\langle n \rangle_0 = A_0 \langle n \rangle$ is equivalent to

$$A_1 = \frac{\beta \langle n \rangle}{\alpha + \beta \langle n \rangle} \quad (18)$$

which is the equilibrium mass action law as discussed in Hornos et al. (2005). However, in the limit of infinitely fast switching, for any fixed n , the gene state is in equilibrium, and hence

$$A_1 = \frac{\beta n}{\alpha + \beta n} \quad (19)$$

In the stationary state, we have to average the above expression, and we get

$$A_1 = \left\langle \frac{\beta n}{\alpha + \beta n} \right\rangle \quad (20)$$

which in general is different from (18). We have also considered a simple cut-off system with maximally two protein molecules allowed. In such a case one can get analytical formulas for the stationary probability distribution. It appeared that in the adiabatic limit, $\langle n \rangle_i \neq A_i \langle n \rangle$ but we are very close to the equality. Numerical calculations of the exact, but not explicit formula presented in Hornos et al. (2005) indicate that $\langle n \rangle_i = A_i \langle n \rangle$; $i = 0, 1$ is a very good approximation.

Now we set $\langle n \rangle_i = A_i \langle n \rangle$; $i = 0, 1$. This closes (3) for the one-gene case and the analogous system of equations in the two-gene case. We see in Fig. 2 that in the fast-switching case, the mean-field and adiabatic approximations practically coincide. We will now show how far is the variance from the mean in the adiabatic approximation.

In the one-gene case, (5) together with $\langle n \rangle_i = A_i \langle n \rangle$; $i = 0, 1$ give us

$$\text{var}(n) = \langle n \rangle - A_1 \quad (21)$$

For the two-gene case, from (10) it follows that

$$\text{var}(n) = \langle n \rangle - A_1 - 2A_2 \quad (22)$$

We can also get that for large mean expression levels, when one may neglect one protein molecule bound to the promoter, in the adiabatic limit $\text{var}(n) = \langle n \rangle$.

5 Slow Switching Gene

In the slow-switching case, we divide (3) by k_0 and k_1 , respectively, and assume that $\frac{\alpha}{k_i} = \frac{\beta}{k_i} = 0$ and get

$$\begin{cases} \langle n \rangle_0 = \frac{k_0}{\gamma} A_0 \\ \langle n \rangle_1 = \frac{k_1}{\gamma} A_1 + A_1 \\ \langle n(n-1) \rangle_0 = \frac{k_0}{\gamma} \langle n \rangle_0 \\ \langle n(n-1) \rangle_1 = \frac{k_1}{\gamma} \langle n \rangle_1 + \langle n \rangle_1 - A_1 \end{cases} \quad (23)$$

The formula for the variance takes the following form:

$$\begin{aligned} \text{var}(n) &= \langle n(n-1) \rangle + \langle n \rangle - \langle n \rangle^2 \\ &= \left(\frac{k_0}{\gamma} \right)^2 A_0 + \left(\frac{k_1}{\gamma} \right)^2 A_1 + 2 \frac{k_1}{\gamma} A_1 + \frac{k_0}{\gamma} A_0 + \frac{k_1}{\gamma} A_1 + A_1 \\ &\quad - \left(\frac{k_0}{\gamma} A_0 + \frac{k_1}{\gamma} A_1 + A_1 \right)^2 \end{aligned} \quad (24)$$

Now we use the conditional variance formula

$$\text{Var}(X) = \text{Var}(E(X|Y)) + E(\text{Var}(X|Y)), \quad (25)$$

where X is the random variable describing the number of protein molecules and Y describes the gene state. For a fixed state of the gene, $Y = i$, the stationary state of production and degradation processes is Poissonian and, therefore, $\text{Var}(X|Y = 0) = E(X|Y = 0) = \frac{k_0}{\gamma}$ and $\text{Var}(X|Y = 1) = \frac{k_1}{\gamma}$, $E(X|Y = 1) = \frac{k_1}{\gamma} + 1$. It is easy to see that we get exactly the same formula as (24).

The approximation of slow switching has been also used in Qian et al. (2009), but only for the one-gene case. It was of course assumed that when the binding and unbinding rates approach 0, we have two Poisson distributions for the unbound and bound states that we may plug into the Master equation and calculate the total probability that there are n proteins in the system.

6 Discussion

We analyzed analytically a simple model of a self-repressing system with one and two gene copies. We showed that the stationary variance and the Fano factor are bigger for the one-gene case than for the two-gene case, and the difference decreases to zero as switching rates increase.

We derived our formulas within the self-consistent mean-field approximation. The approximation was tested in two extreme cases: fast switching and slow switching genes. We discussed the validity of the adiabatic approximation for fast switching genes and showed that both mean-field and adiabatic approximations agree in this

regime. In the slow-switching case, we derived rigorous formulas, which coincide with the mean-field approximation formulas.

We also established the linear dependence of the variance with respect to the mean as the adiabaticity parameter increases; the slope is bigger in the two-gene case than in the one-gene case.

It would be interesting to use mean-field approximation in other regulatory gene systems, like the toggle switch, and in general in systems with bistabilities.

Acknowledgements J.M. would like to thank the Ministry of Science and Higher Education for a financial support under the grant N201 362536. P.S. was supported by the EU through the European Social Fund, contract number UDA-POKL.04.01.01-00-072/09-00.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

References

- Barzel, B., & Biham, O. (2011). Binomial moment equations for stochastic reaction systems. *Phys. Rev. Lett.*, *106*, 150602.
- Barzel, B., Biham, O., & Kupferman, R. (2011). Analysis of the multiplane method for stochastic simulations of reaction networks with fluctuations. *Multiscale Model. Simul.*, *6*, 963–982.
- Becskei, A., & Serrano, L. (2000). Engineering stability in gene networks by autoregulation. *Nature*, *405*, 590–593.
- Hat, B., Paszek, P., Kimmel, M., Piechór, K., & Lipniacki, T. (2007). How the number of alleles influences gene expression. *J. Stat. Phys.*, *128*, 511–533.
- Hornos, J. E., Schultz, D., Innocentini, G. C., Wang, J., Walczak, A. M., Onuchic, J. N., & Wolynes, P. G. (2005). Self-regulating gene: an exact solution. *Phys. Rev. E*, *72*, 051907.
- Huang, K. (1963). *Statistical mechanics*. New York: Wiley.
- Kepler, T., & Elston, T. (2001). Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.*, *81*, 3116–3136.
- Komorowski, M., Mięksiz, J., & Kierzek, A. (2009). Translational repression contributes greater noise to gene expression than transcriptional repression. *Biophys. J.*, *96*, 372–384.
- Lipniacki, T., Paszek, P., Marciniak-Czochra, A., Brasier, A. R., & Kimmel, M. (2006). Transcriptional stochasticity in gene expression. *J. Theor. Biol.*, *238*, 348–367.
- Lipshtat, A., Perets, H., Balaban, N., & Biham, O. (2005). Modeling of negative autoregulated genetic networks in single cells. *Gene*, *347*, 265–271.
- Loinger, A., & Biham, O. (2009). Analysis of genetic toggle switch systems encoded on plasmids. *Phys. Rev. Lett.*, *103*, 068104.
- Ma, S.-K. (1985). *Statistical mechanics*. Singapore: World Scientific.
- Mięksiz, J., & Szymańska, P. (2012). On spins and genes. *Math. Appl.*, *40*(1), 15–25.
- Nasell, I. (2003). An extension of the moment closure method. *Theor. Popul. Biol.*, *64*, 233–239.
- Ohkubo, J. (2010). Approximation scheme based on effective interactions for stochastic gene regulation. *Phys. Rev. E*, *83*, 041915.
- Paszek, P. (2007). Modeling stochasticity in gene regulation: characterization in the terms of the underlying distribution function. *Bull. Math. Biol.*, *69*, 1597–1601.
- Paulsson, J. (2004). Summing up the noise in gene networks. *Nature*, *427*, 415–418.
- Paulsson, J. (2005). Models of stochastic gene expression. *Phys. Life Rev.*, *2*, 157–175.
- Qian, H., Shi, P.-Z., & Xing, J. (2009). Stochastic bifurcation, slow fluctuations, and bistability as an origin of biochemical complexity. *Phys. Chem. Chem. Phys.*, *11*, 4861–4870.
- Ramos, A. F., Innocentini, G. P., & Hornos, J. E. (2011). Exact time dependent solutions for a self-regulating gene. *Phys. Rev. E*, *83*, 062902.
- Simpson, M., Cox, L. C. D., & Saylor, G. S. (2003). Frequency domain analysis of noise in autoregulated gene circuits. *Proc. Natl. Acad. Sci. USA*, *100*, 4551–4556.

- Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. USA*, *99*, 12795–12800.
- Thattai, M., & van Oudenaarden, A. (2001). Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. USA*, *98*, 8614–8619.
- Van Kampen, N. (1997). *Stochastic processes in physics and chemistry* (2nd ed.). Amsterdam: Elsevier.