Journal of Theoretical Biology 364 (2015) 355–363

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Stochastic gene expression with delay

Martin Jansen [a], Peter Pfaffelhuber [b,*]

[a] University Medical Center Freiburg, Institute of Clinical Chemistry and Laboratory Medicine, Hugstetter Straße 55, 79106 Freiburg, Germany
[b] University of Freiburg, Abteilung für Mathematische Stochastik, Eckerstr. 1, 79104 Freiburg, Germany

HIGHLIGHTS

- We extend the standard model for stochastic gene expression by adding a delay to both, transcription and translation.
- Our results are precise formulas for the first two moments of the (random) number of RNA and protein in equilibrium.
- We find that delay distributions lead to a less bursty transcription than without delay.

ARTICLE INFO

Article history:
Received 22 March 2014
Received in revised form
7 August 2014
Accepted 17 September 2014
Available online 5 October 2014

MSC:
92C42
92C40
60K35

Keywords:
Chemical reaction network
Poisson process

ABSTRACT

The expression of genes usually follows a two-step procedure. First, a gene (encoded in the genome) is transcribed resulting in a strand of (messenger) RNA. Afterwards, the RNA is translated into protein. We extend the classical stochastic jump model by adding delays (with arbitrary distributions) to transcription and translation.

Already in the classical model, production of RNA and protein comes in bursts by activation and deactivation of the gene, resulting in a large variance of the number of RNA and proteins in equilibrium. We derive precise formulas for this second-order structure with the model including delay in equilibrium.

1. Introduction

The central dogma of molecular biology is that a gene (encoded within the genome) is transcribed into (messenger) RNA (also abbreviated mRNA), which in turn is translated into protein, the whole process also being called gene expression. Mathematical models for this process have by now been studied for a long time; see e.g. Rigney and Schieve (1977), Berg (1978), McAdams and Arkin (1997), Swain et al. (2002), Paulsson (2005), Cottrell et al. (2012), Bokes et al. (2012), Pendar et al. (2013), and Fromion et al. (2013).

Within a single cell, gene expression often comes with stochastic fluctuations; see e.g. Raser and O'Shea (2005), Raj and van Oudenaarden (2008), and Balazsi et al. (2011). There are either one or two copies of the genome, and only a few genes code for the same protein. As reviewed by Jackson et al. (2000) the majority of expressed RNA species in mammalian cells have less than 10 copies, though there are also RNA species present at an order of 10,000 copies. Guptasarma (1995) observed that for 80% of genes in *E. Coli* genome the copy number of many proteins is less than 100. Hence in many cases there are only a small copy numbers of RNA and protein molecules, making them a noisy (i.e. stochastic) quantity. While this stochasticity has been assumed to be detrimental to the cellular function, it can also help a cell to adapt to fluctuating environments, or help to explain genetically homogeneous but phenotypically heterogeneous cellular populations (Kaern et al., 2005).

In order to consider stochasticity in gene expression, Swain et al. (2002) distinguish between *intrinsic* and *extrinsic* noise. The latter accounts for changing environments of the cell, while the former accounts for the stochastic process of transcription and translation. Let us look at the possible sources of intrinsic noise in more detail; see e.g. Zhu et al. (2007), Roussel and Zhu (2006a).

(i) Various mechanisms for gene expression require random events to occur. In order to understand this let us have a closer look at the mechanisms of gene expression. Transcription starts when RNA polymerase (which are enzymes helping in the synthesis of RNA) binds to the promoter region of the gene, forming an

* Corresponding author.
E-mail addresses: martin.jansen@uniklinik-freiburg.de (M. Jansen), p.p@stochastik.uni-freiburg.de (P. Pfaffelhuber).

http://dx.doi.org/10.1016/j.jtbi.2014.09.031
0022-5193/

elongation complex. This elongation complex is then ready to start walking along the DNA, reading off DNA and making RNA. Before the transcript is released, a *ribosome binding site* (which is needed for translation) is being produced on the transcript. Then follows translation which starts when a free ribosome binds to the ribosome binding site of the transcript and again is a complex process involving many chemical reactions, which lead to fluctuations.

(ii) Another source of noise comes from turning genes on and off. This means that transcription factors can bind to promoter regions of the gene and only bound (or unbound) promoters can initiate transcription. This process has been found to be the most important source of randomness for gene expression (see e.g. Swain et al., 2002; Kaern et al., 2005; Zhu and Salahub, 2008; Raj and van Oudenaarden, 2008; Iyer-Biswas et al., 2009). The effect of this activation and inactivation of genes is a burst-like behavior of protein production, already apparent in McAdams and Arkin (1997). When considering the amount of RNA within the cell during the production of a specific protein, it is hence not surprising that production of RNA comes in bursts, which are related to times when the gene is turned on. This burst-like behavior is inherited to protein formation, which also comes in bursts during translation.

The classical model of stochasticity in gene expression uses exponential waiting times between transcription and translation events, and once produced, RNA and protein molecules are immediately available to the system. The latter contradicts several biological facts, valid in prokaryotes as well as in eukaryotes, e.g.: Production of RNA consists of many enzymatic reactions (Roussel and Zhu, 2006a). In the translation process another set of reactions unbinds RNA from the ribosome. For eukaryotes, post-transcriptional modification of RNA and the transport of RNA out of the nucleus to the ribosomes, as well as folding of proteins, requires time. Taking such issues into account, it makes sense to model a (random) time delay before an RNA or protein molecule can be used by the system. In our paper, we are studying the effect of (random) delay on the noise in gene expression. In real-life applications, models for such gene expression delays have been considered e.g. by Lewis (2003), Monk (2003), Barrio et al. (2006), and Bratsun et al. (2005).

While our modeling approach only takes a single gene/RNA/protein triple into account, the field of systems biology aims at unraveling interactions between genes in so-called pathways. It seems clear that randomness as well as delays can accumulate in such networks of interacting genes and proteins. As a simple example, the transcription factor regulating the expression of gene A is coded by a gene B which in turn may be regulated by gene A (or by itself), which can lead to a bi-modal distribution of the number of proteins encoded by gene A or B; see e.g. Kaern et al. (2005). Although such feedback systems are highly interesting, we are not touching on this level of complexity.

Today, delays in biochemical reaction networks also serve as a tool for model reduction. Barrio et al. (2013) and Leier et al. (2014) argue that lumping together certain reactions effectively leads to a delay for other reactions. At least for first order reactions, they compute the resulting delay times which serve for a precise model reduction.

Simulation of chemical systems, or *in silico modeling*, today paves the way to understanding complex cellular processes. While the Gillespie algorithm is a classical approach for stochastic simulations (Gillespie, 1977) – see also the review (Gillespie et al., 2013) – chemical delay models have as well been algorithmically studied. Various explicit simulation schemes for delay models – in particular in the field of stochasticity in gene expression – have been given; see Bratsun et al. (2005), Roussel and Zhu (2006b), Barrio et al. (2006), Cai (2007), Tian et al. (2007), Anderson (2007), Ribeiro (2010), Tian (2013), and Zavala and Marquez-Lago (2014).

The goal of this paper is to give a quantitative evaluation of delay in the standard model of stochastic gene expression. We do this in a general way in which the delay – both for transcription and translation – can have an arbitrary distribution. Although we give a full description of the stochastic processes of the total number of RNA and protein molecules, our quantitative results are restricted since we only address the calculation of the first two moments (expectation, variance and autocovariances) of the number of RNA and protein molecules.

Outline: In Section 2, we introduce our delay model using a classical approach of stochastic time-change equations as well as a description of the system in equilibrium. Then, we present our main results in Section 3. Basically, Theorems 3.2 and 3.4 give the second-order structure of the number of RNA and protein in equilibrium under the delay model, respectively. In Section 4, we give several examples (uniformly and exponentially distributed delay, and delay with small variance). We end our paper with a discussion and connections to previous work in Section 5.

2. The model

In order to be able to model gene expression in a sophisticated way, we now give our delay model. Using the terminology from Roussel (1996), we may write

$$\begin{aligned}
&\text{inactive gene} \underset{\lambda_1^+}{\overset{\lambda_1^-}{\rightleftharpoons}} \text{active gene}\\
&\text{active gene } (t) \xrightarrow{\lambda_2} \text{active gene}(t)+\text{RNA transcript}(t+G_2)\\
&\text{RNA transcript}(t) \xrightarrow{\lambda_3} \text{RNA transcript}(t)+\text{protein}(t+G_3)\\
&\text{RNA transcript} \xrightarrow{1/T_2} \varnothing\\
&\text{protein} \xrightarrow{1/T_3} \varnothing.
\end{aligned} \tag{1}$$

Essentially, (1) is an extension of the well-studied model of gene expression, as e.g. given in Paulsson (2005). Gene expression of $n_{\max}$ similar genes is studied. Each gene is activated and deactivated at rates λ_1^+ and λ_1^-, respectively. (Additionally, we will set $\tau_1 = 1/(\lambda_1^+ + \lambda_1^-)$). Every active gene creates the RNA transcript at rate λ_2, which is degraded at rate $1/\tau_2$. However, a RNA molecule is available for the system (i.e. can be translated) only some random delay time G_2 after its creation, where G_2 is an independent random variable with distribution μ. Then, each RNA transcript available for the system initiates translation of protein at rate λ_3 which in turn degrades at rate $1/\tau_3$. Again, it takes a delay of a random time G_3, distributed according to ν and independent of everything else, that the protein molecule is available for the system (i.e. for other downstream processes).

We note that (1) is a special case of a model studied in Zhu et al. (2007). Since they consider the ribosome binding site as an own chemical species, their model requires more delay random variables. Moreover, they distinguish gene expression in prokaryotes (bacteria) and eukaryotes (higher organisms), the main difference being that only eukaryotes have a cellular core. As a consequence, in prokaryotes translation can already be initiated when transcription is not complete yet. (The ribosome can bind to the ribosome binding site while the RNA transcript is still being produced.) The simplification (6) and (7) in Zhu et al. (2007) for gene expression in prokaryotes (both, the time the promoter region of the gene is