Effects of extrinsic noise are promoter kinetics dependent

Effects of extrinsic noise are promoter kinetics dependent

Mohamed N.M. Bahrudeen
Laboratory of Biosystem Dynamics, Biomeditech, Tampere University of Technology, Finland.
P.O Box 553, 33101 Tampere, Finland.
E-mail: mohamed.mohamedbahrudeen@tut.fi

Sofia Startceva
Laboratory of Biosystem Dynamics, Biomeditech, Tampere University of Technology, Finland.
P.O Box 553, 33101 Tampere, Finland.
E-mail: sofia.startceva@tut.fi

Andre S. Ribeiro
Laboratory of Biosystem Dynamics, Biomeditech, Tampere University of Technology, Finland.
Office TC338, P.O Box 553, 33101 Tampere, Finland.
Phone: +358 408490736.
E-mail: andre.ribeiro@tut.fi

ABSTRACT

Studies in *Escherichia coli* using *in vivo* single-RNA detection and time-lapse confocal microscopy showed that transcription is a multiple rate-limiting steps process, in agreement with previous *in vitro* measurements. Here, from simulations of a stochastic model of transcription validated empirically that accounts for cell-to-cell variability in RNA polymerase (RNAP) numbers, we investigate the hypothesis that the cell-to-cell variability in RNA numbers due to RNAP variability differs with the promoter rate-limiting steps dynamics. We find that increasing the cell-to-cell variability in RNAP numbers increases the cell-to-cell diversity in RNA numbers, but the degree with which it increases is promoter kinetics dependent. Namely, promoters whose open complex formation is relatively longer lasting dampen more efficiently this noise propagation phenomenon. We conclude that cell-to-cell variability in RNA numbers due to variability in RNAP numbers is promoter-sequence dependent and, thus, evolvable.

CCS Concepts

• CCS → Computing methodologies → Modeling and simulation → **Model development and analysis**

<ccs2012>
<concept>
<concept_id>10010147.10010341.10010342.10010343</concept_id>
<concept_desc>Computing methodologies~Modeling methodologies</concept_desc>
<concept_significance>300</concept_significance>
</concept>
</ccs2012>

Keywords

Phenotypic diversity; gene expression; extrinsic noise; transcription initiation; rate-limiting steps; stochastic models.

1. INTRODUCTION

In *Escherichia coli*, major behavioral changes including metabolic, are driven by changes in the numbers of the molecules composing the transcriptional and translational machineries, such as RNA polymerase (RNAP) core enzymes and σ factors [1]. E.g., changes in σ factor numbers allow *E. coli* cells to quickly, and simultaneously, enhance and/or reduce the transcriptional activity of a large number of genes in a selective fashion [2].

This is made possible by the limited number of RNAP core enzymes [3]. As the numbers of a specific σ factor increase, the RNAPs carrying that σ factor increase, and thus the activity of the promoters associated to that σ factor is expected to increase by direct positive regulation. Meanwhile, the activity of the promoters associated to other σ factors is expected to decrease by indirect negative regulation.

Interestingly, following changes in σ factor numbers, while these 'expectations' based on present models of transcription do, in general, take place on a global scale, some genes' activity is unaffected [3], and those that respond do so heterogeneously, i.e., they differ in degree of change, even in the case of genes associated to the same σ factor.

This 'behavioral' diversity in responses is due to differences in the promoters' selectivity for σ factors [4], the action of transcription factors [3] and, according to a recent study, in the case of indirect negative regulation, due to differing dynamics of the multiple steps in transcription initiation [5][6], which were first observed by *in vitro* measurement techniques (for a review see [7]). In particular, it was shown that in promoters preferentially transcribed by σ^{70}, the smaller the time-scale of the closed complex formation relative to the open complex formation, the weaker is the promoter's responsiveness to changes in σ^{38} numbers. It was thus concluded that, in *E. coli*, a promoter's responsiveness to indirect regulation by σ factor competition is determined by the kinetics of the irate limiting steps in initiation.

Given this observation, validated by various measurement techniques of RNA production dynamics applied to several promoters [6], we here hypothesize that the dynamics of the rate-limiting steps in transcription initiation [7] influences also a gene's degree of responsiveness to extrinsic noise sources.

Here, we investigate this hypothesis by, first, establishing a stochastic model of transcription that accounts for cell-to-cell diversity in RNA polymerase numbers and whose parameter values are taken from state-of-the-art, single-cell RNAP levels and single-RNA microscopy measurements, and then performing stochastic simulations of model cells [8][9] carrying the multi-step stochastic model of transcription [10] and whose RNAP numbers are, while constant in time, initially randomly drawn from a normal distribution. By tuning this model's parameter values, we assess to which extent variability in RNAP numbers, as function of transcription initiation kinetics, affects the cell-to-cell diversity in RNA numbers.

2. METHODS

2.1 Model of Transcription

We consider a dynamically broad model of transcription initiation that allows RNA production kinetics to range from sub-Poissonian to super-Poissonian, depending on the rate constant values. This model was derived from data from multiple studies, including genome-wide studies of RNA numbers variability [11][12] and of the transcription dynamics of individual genes [13,14].

The model includes the steps in transcription initiation in *E. coli* (e.g. open complex formation [15] and ON/OFF process [16]). Rate constant values were obtained by fitting the model to empirical data on RNA production kinetics at the molecule level of the Lac-Ara-1 promoter and from single-cell measurements of intracellular concentration of RNAPs reported in [5].

This model is applicable to common promoters in *E. coli*, differing between promoters in the rate constant values, and it consists of the following reactions:

$$P_{\mathrm{ON}} + R \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} RP_c \xrightarrow{k_2} RP_o \xrightarrow{k_3} R + P_{\mathrm{ON}} + \mathrm{RNA} \quad (1)$$

$$P_{\mathrm{ON}} \underset{k_{ON}}{\overset{k_{OFF}}{\rightleftarrows}} P_{\mathrm{OFF}} \quad (2)$$

Reactions (1) represent the multi-step transcription initiation of an active promoter, P_{ON} [17]. First, the closed complex (RPc) is formed as the RNAP (R) binds to a free promoter [18]. Next, intermediate steps occur to form the open complex (RPo) [17][18]. Finally, elongation begins after promoter clearance [19], after which the promoter, the produced RNA, and the RNAP are released. In (1), k_1 is the rate at which RNAPs find and bind to the promoter successfully. k_2 is the open complex formation rate. Finally, k_3 is the promoter escape rate. k_{-1} is the rate of reversibility of the closed complex. In this model, a promoter occupied by an RNAP is unavailable to other initiation events.

Reactions (2) represent the intermittent transitioning of the promoter to an inactive state (P_{OFF}) due to e.g. binding/unbinding of repressors/activators [20], accumulation of positive DNA supercoiling [21], etc.

As the number of RNAPs differs between live cells (see measurements below), in each model cell the number of RNAPs is constant but initially randomly generated from a normal distribution, N(x,y), where x and y are obtained from empirical data [1]. This is the source of extrinsic noise of the model cell population here considered and is the main innovation of our model when compared to previous stochastic models [5][10][22].

2.2 Stochastic Simulations

Simulations are performed by SGNS [8], a simulator of chemical reaction systems whose dynamics is driven by the Stochastic Simulation Algorithm [9] that allows for multi-time-delayed reactions [10]. SGNS also allows hierarchical, interlinked compartments to be created, destroyed and divided at runtime, a feature used to generate dynamically independent model cells.

3. RESULTS AND CONCLUSIONS

Here, each model cell 'contains' one promoter and RNAP molecules, which interact via reactions (1) and (2). Parameter values of the 'control condition' are shown in Tables 1 and 2.

The parameter values associated with RNAP numbers in individual cells (Table 1) are obtained from measurements of RNAP fluorescence intensities in individual *E. coli* RL1314 cells with fluorescently tagged β' subunits reported in [5]. From these, we have set the mean RNAP fluorescence in individual cells arbitrarily to 1 and then obtained the fraction of cells with a given relative fluorescence level. The resulting distribution of relative RNAP fluorescence levels is shown in Figure 1. Note that the 2.5% cells with lowest and highest total fluorescence intensity were discarded, as they were clear outliers.

Next, to obtain the CV^2 of these RNAP relative levels in individual cells, we fitted a normal distribution to the data using the MATLAB package *Statistics and Machine Learning Toolbox*™ [23] (Figure 1). The CV^2 of the fit is shown in Table 1.

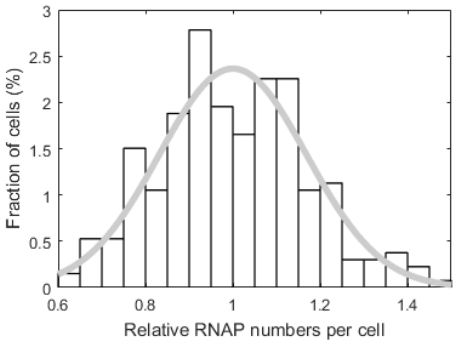

Fig 1. Fraction of cells with a given relative RNAP-fluorescence level as measured by microscopy in *E. coli* RL1314 cells with fluorescently tagged β′ subunits (bars) [1]. The mean absolute RNAP level in individual cells was set to 1. Also shown is the best fitting normal distribution (grey line).

To validate the fitting, we performed a Kolmogorov-Smirnov (KS) test and verified that the two distributions (empirical and best fit) cannot be distinguished in a statistical sense (p-value of 0.69). Thus, we use the best fit distribution to set random RNAP numbers in individual model cells. The mean and CV^2 of RNAP numbers of the model cell population are shown in Table 1.

Table 1. Parameter values of RNAP numbers in model cells in the control condition of the simulations

Parameter	Value	Reference
Mean RNAp available/cell	1*	[5]
CV^2 of RNAp available/cell	0.03	[5]

Table 2 shows the values of k_{ON}, k_{OFF}, k_1, k_{-1}, k_2, and k_3 of the transcription model, which were inferred from empirical distributions of time intervals between consecutive RNA productions in individual cells, under the control of the Lac-Ara-1 promoter in DH5α-PRO *E. coli* cells [5]. For this statistical inference, it was assumed the same model of transcription as here.

Table 2. Parameter values of the transcription model (control)

Parameter	Value	Reference
k_{ON}	0.01 s^{-1}	[5]
k_{OFF}	281 s^{-1}	[5]
k_1	6469 s^{-1}	[5]

Conference'10, Month 1–2, 2010, City, State, Country.

DOI: http://dx.doi.org/10.1145/12345.67890

k_{-1}	1 s^{-1}	[5]
k_2	0.005 s^{-1}	[5]
k_3	∞	[5]

Next, we ran simulations of model cells (control condition) using SGNS2 [8]. Example time series of RNA production events in 5 individual model cells are shown in Figure 2. Visibly, most cells produced 2 RNAs during 2000 s, as expected [5].

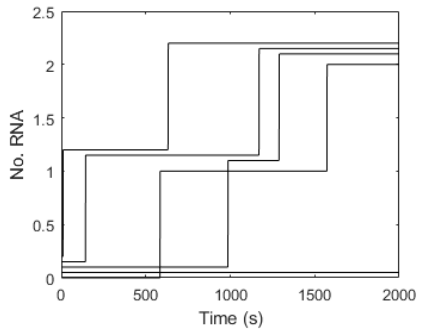

Fig 2. Example time series of the number of RNAs produced by 5 individual model cells in 2000 s. These numbers are offset of each other on the y axis to distinguish between the lines of different cells (only integer RNA numbers are possible).

Next, we study the overall cell to cell diversity in number of produced RNAs as a function of the cell-to-cell variability in RNAP numbers and as a function of the rate constants controlling the kinetics of closed (k_1) and open (k_2) complex formation.

According to the model, e.g., increasing k_1 results in shorter time-length closed complex. Meanwhile, increasing k_2 results in shorter time-length open complex. Here, we change the values of both k_1 and k_2 so that the mean RNA production rate remains unaltered. For that, we use the formula derived in [1], and reproduced here:

$$I(R) = \frac{(k_{ON} + k_{OFF})(k_{-1} + k_2)}{R k_1^* k_2 k_{ON}} + \frac{1}{k_2} + \frac{1}{k_3} \qquad (3)$$

In (3), I(R) is the mean interval between consecutive RNA productions in individual cells. This equation, derived from reactions (1) and (2), assumes infinite cell lifetime. Here, we vary k_1, and then, based on (3), vary k_2 so as to maintain *I(R)* constant.

Given this, we selected 10 values for k_1 and, consequently, k_2 [5]. From these, using (3), we calculate for each case the fraction of time between consecutive RNA production events (Δt) that is spend in closed complex formation ($\tau_{cc}/\Delta t$). The range of these values was set so as to be in agreement with recent measurements made for various promoters subject to various induction levels in cells whose RNAP numbers distribution is similar to that in Table 1. These empirical values are shown in Table 3.

Table 3. Empirical values of $\tau_{cc}/\Delta t$ of various promoters subject to different induction levels

Promoter and induction	$\tau_{cc}/\Delta t$	Reference
P_{BAD} with 0.1% arabinose	0.71	[6]
P_{BAD} with 0.01% arabinose	0.55	[6]
P_{BAD} with 0.001% arabinose	0.17	[6]
$P_{lac\text{-}O1O3}$ with 1 mM IPTG	0.55	[6]
$P_{lac\text{-}O1O3}$ with 0.05 mM IPTG	0.46	[6]
$P_{lac\text{-}O1O3}$ with 0.005 mM IPTG	0.12	[6]
P_{tetA} with no inducers	0.07	[6]
$P_{lac\text{-}O1}$ with 1 mM IPTG	0.05	[6]
$P_{lac\text{-}ara1}$ with 1 mM IPTG and 0.1% arabinose (full induction)	0.49	[6]

Also, we selected 7 different values of CV^2 in RNAP numbers in individual cells, around the empirical value of 0.03 (Table 1). Based on these sets of parameter values, we produced 70 models of cells, combining in all possible ways the two parameter sets. For each model, we simulated 1000 model cells for 10000 s each, and extracted the number of produced RNAs per cell. The mean number of RNAs produced per cell in the various models equaled ~10. In Figure 3, we show the values of the CV^2 of the number of produced RNAs in individual cells in all conditions.

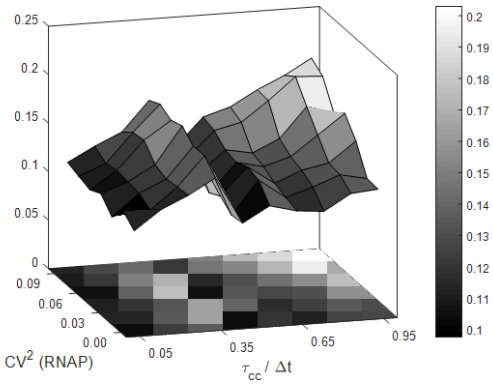

Fig 3. CV^2 of number of produced RNAs in model cells during their lifetime as a function of relative durations of closed and open complex formation and of the cell-to-cell variability in RNAP numbers.

From Figure 3, as $\tau_{cc}/\Delta t$ increases, so does the cell-to-cell variability in RNA numbers. Similarly, the higher the cell-to-cell variability in RNAP numbers, the higher the CV^2 in RNA numbers. Finally, increasing both these two parameters leads to a much higher increase in the cell-to-cell variability in RNA numbers, than if changing only one of these parameters.

The conclusion from these results is that, while as expected the cell-to-cell variability in RNAP numbers 'propagates' to the cell-to-cell diversity in RNA numbers, the degree with which it propagates is heavily promoter kinetics dependent.

Also, there is an unexpected decrease in CV^2 in RNA numbers at $\tau_{cc}/\Delta t$ ~ 0.35, that will require further research to explain.

4. DISCUSSION

We explored the effects of cell-to-cell variability in RNAP numbers in the cell-to-cell variability in RNA production rates, as a function of the kinetics of transcription initiation of a promoter.

For this, we simulated the dynamics of RNA production in model cells, making use of a detailed stochastic model that combines multiple steps in transcription initiation with cell-to-cell variability in RNAP numbers. All parameter values of the model were inferred from single-cell microscopy measurements.

We observed that as the cell-to-cell variability in RNAP numbers increases, so does the variability in RNA numbers. However, genes are not entirely void of 'filters' of this phenomenon. Namely, within the range of realistic parameter values, we observed that different promoter kinetics results in different degrees of variability in RNA numbers in individual cells. Specifically, RNAs whose production is controlled by promoters with relatively slow closed complex formation will exhibit much wider variability in numbers between cells.

As the initiation dynamics of promoters is both sequence-dependent and subject to regulation (e.g. inducers and repressors), we expect the level of cell-to-cell diversity in RNA (and protein) numbers of a gene due to the variability in RNAP numbers to be both evolvable as well as adaptable.

In addition, given the observed degree of changes in variability in RNA numbers as a function of the two parameter values changed in the course of the simulations, we expect this phenomenon to also be observable at the level of small genetic circuits. In the future, it would be of interest to investigate the extent to which this effect influences the behavior of such small circuits, such as genetic switches and clocks.

5. ACKNOWLEDGMENTS

Work supported by Academy of Finland (295027 to A.S.R.), Academy of Finland Key Project Funding (305342 to A.S.R.), Jane and Aatos Erkko Foundation (610536 to A.S.R.), and Tampere University of Technology President's Graduate Program (to S.S). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

6. REFERENCES

[1] Jishage, M., Iwata, A., Ueda, S., Ishihama, A. 1996. Regulation of RNA polymerase sigma sub-unit synthesis in Escherichia coli: intracellular levels of four species of sigma subunit under various growth conditions. *J. Bacteriol.* 178, 5447–51.

[2] Rahman, M., Hasan, M.R., Oba, T. , Shimizu, K. 2006. Effect of rpoS gene knockout on the metabolism of Escherichia coli during exponential growth phase and early stationary phase based on gene expressions, enzyme activities and intracellular metabolite concentrations. *Biotechnol. Bioeng.* 94, 585–95

[3] Farewell, A., Kvint, K., Nyström, T. 1998. Negative regulation by RpoS: A case of sigma factor competition, *Mol. Microbiol.* 29, 1039–51.

[4] Hengge-Aronis, R. 2002. Recent insights into the general stress response regulatory network in Escherichia coli. *J. Mol. Microbiol. Biotechnol.* 4, 341–6.

[5] Lloyd-Price, J., Startceva, S., Kandavalli, V., Chandraseelan, J., Goncalves, N., Oliveira, S.M.D., Häkkinen, A. and Ribeiro, A.S. 2016. Dissecting the stochastic transcription initiation process in live Escherichia coli. *DNA Research* 23(3), 203-214.

[6] Kandavalli, V.K., Tran, H. and Ribeiro, A.S. 2016. Effects of σ factor competition on the *in vivo* kinetics of transcription initiation in *Escherichia coli*. *BBA Gene Regulatory Mechanisms* 1859, 1281–1288.

[7] McClure, W.R. 1980. Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. U. S. A.* 77, 5634–5638.

[8] Lloyd-Price, J., Gupta, A, and Ribeiro, A.S. 2012. SGNS2: A Compartmentalized Stochastic Chemical Kinetics Simulator for Dynamic Cell Populations. *Bioinformatics* 28, 3004-5.

[9] Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81(25), 2340–2361.

[10] Roussel, M.R. and Zhu, R. 2006. Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Phys Biol.* 3(4), 274-284.

[11] Taniguchi, Y., Choi, P.J., Li, G.-W., et al. 2010. Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* 329, 533–538.

[12] Bernstein, J.A., Khodursky, A.B., Pei-Hsun, L., Lin-Chao, S. and Cohen, S. N. 2002, Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. *Proc. Natl Acad. Sci. USA* 99, 9697–702.

[13] Muthukrishnan, A.-B., Kandhavelu, M., Lloyd-Price, J., et al. 2012, Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Res.* 40, 8472–83.

[14] Kandhavelu, M., Lloyd-Price, J., Gupta, A., Muthukrishnan, A., Yli-Harja, O. and Ribeiro, A.S. 2012, Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter. *FEBS Lett.* 586, 3870–5.

[15] McClure, W.R. 1980. Rate-limiting steps in RNA chain initiation. *Proc. Natl Acad. Sci. USA* 77, 5634–8.

[16] So, L.-H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R. and Golding, I. 2011. General properties of transcriptional time series in Escherichia coli, *Nat. Genet.* 43, 554–60.

[17] Saecker, R.M., Record, M.T. and Dehaseth, P.L. 2011. Mechanism of bacterial transcription initiation: RNA polymerase - promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J. Mol. Biol.* 412, 754–71.

[18] Chamberlin, M.J. 1974. The selectivity of transcription, *Annu. Rev. Biochem.* 43, 721–75.

[19] DeHaseth, P.L., Zupancic, M.L. and Record, M.T. 1998. RNA polymerasepromoter interactions: The comings and goings of RNA polymerase. *J. Bacteriol.* 180, 3019–25.

[20] Lutz, R., Lozinski, T., Ellinger, T. and Bujard, H. 2001. Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res.* 29, 3873–81.

[21] Chong, S., Chen, C., Ge, H. and Xie, X.S. 2014. Mechanism of transcriptional bursting in bacteria. *Cell* 158, 314–26.

[22] Ribeiro, A.S., Zhu, R. and Kauffman, S.A. 2006. A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics. *J. of Comput. Biol.* 13, 1630-1639.

[23] MATLAB and Statistics and Machine Learning Toolbox Release 2016b, The MathWorks, Inc., Natick, Massachusetts, United States.