Effects of the Dynamics of the Steps in Transcription Initiation on the Asymmetry of the Distribution of Time Intervals between Consecutive RNA Productions

Sofia Startceva[1], Vinodh Kumar Kandavalli[1], Ari Visa[2], and Andre S. Ribeiro[1]

[1]Laboratory of Biosystem Dynamics, BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, 33101, Tampere, Finland
[2]Signal Processing Unit, Faculty of Computing and Electrical Engineering, Tampere University of Technology, 33101, Tampere, Finland
andre.ribeiro@tut.fi

Abstract. Asymmetries in the distribution of time intervals between consecutive RNA productions from a gene can play a critical role in, e.g., allowing/preventing the RNA and, thus, protein numbers to cross thresholds involved in gene network decision making. Here, we use a stochastic, multi-step model of transcription initiation, with all rate constants empirically validated, and explore how the kinetics of its steps affect the temporal asymmetries in RNA production, as measured by the skewness of the distribution of intervals between consecutive RNA productions in individual cells. From the model, first, we show that this skewness differs widely with the mean fraction of time that the RNA polymerase spends in the steps preceding open complex formation, while being independent of the mean transcription rate. Next, we provide empirical validation of these results, using qPCR and live, time-lapse, single-molecule RNA microscopy measurements of the transcription kinetics of multiple promoters. We conclude that the skewness in RNA production kinetics is subject to regulation by the kinetics of the steps in transcription initiation and, thus, evolvable.

Keywords: Transcription Initiation, Asymmetries in RNA production; Stochastic Models; Single-RNA measurements.

Gene expression regulation in bacteria occurs mostly in transcription initiation [1]. In *Escherichia coli*, this process is sequential [2], starting with an RNA polymerase (R) binding to an active promoter (P_{ON}) and forming a closed complex (RP_{cc}). Next, the open complex (RP_{oc}) forms. Relevantly, the subsequent steps of RNA elongation [3], termination, and RNA and R release are much faster. Thus, dynamically, transcription can be approximately modeled as:

$$R + P_{ON} \xrightarrow{k_{cc}} RP_{cc} \xrightarrow{k_{oc}} RP_{oc} \xrightarrow{\infty} P_{ON} + R + \text{RNA} \quad (1)$$

Here, RNA production kinetics is controlled by k_{cc} and k_{oc}. The probability density function (pdf) of the distribution of intervals between transcription events is the convolution of their pdfs: $f_{\Delta t}(t) = \frac{k_{cc} \cdot k_{oc}}{k_{oc} - k_{cc}} (e^{-k_{cc} \cdot t} - e^{-k_{oc} \cdot t})$. To measure asymmetries

in this distribution, we use skewness, $S = \frac{m_3}{{m_2}^{3/2}}$, where $m_r = \frac{1}{n}\Sigma(x_i - \bar{x})^r$ [4]. We estimate the sample skewness $S_s = \frac{\sqrt{n(n-1)}}{n-2} \cdot S$, where n is the sample number [5]. To obtain confidence boundaries for S_s we use non-parametric bootstraps as in [6].

In (1), k_{cc} is the inverse of the mean time for R to bind the promoter and complete a closed complex (τ_{cc}), while k_{oc} is the inverse of the mean time for an open complex to form (τ_{oc}). The mean time between transcription events: $\Delta t = \tau_{cc} + \tau_{oc}$.

To validate the model predictions of skewness, we collected empirical data for Δt and $\tau_{cc}/\Delta t$ for various promoters (P_{TetA}, P_{BAD}, $P_{Lac\text{-}ara\text{-}1}$, and $P_{Lac\text{-}ara\text{-}1}$ under oxidative stress) [7-9] (Fig. 1). Next, given the mean Δt of each promoter, we varied $\tau_{cc}/\Delta t$ (from 0 to 1) while maintaining Δt constant. Then, for each value of $\tau_{cc}/\Delta t$, we calculated S from the pdf of the distribution of intervals between transcription events (solid line, Fig. 1). Interestingly, we observed that S is independent of the mean value of Δt. Finally, from Fig. 1, we find that the model predictions of S fit the empirical data.

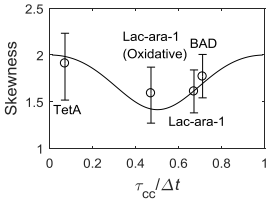

Fig 1. Predicted skewness of Δt distributions with given $\tau_{cc}/\Delta t$ (solid line) and sample skewness of the empirical Δt distributions (with 95% confidence intervals) for the studied promoters. For each promoter, 100 or more Δt intervals were extracted from a total of 100 or more cells.

Importantly, as S is tunable by τ_{cc} and τ_{oc}, which are sequence dependent and subject to regulation, we expect it to be evolvable and adaptive to environment shifts.

References

1. Kaern, M. et al.: Stochasticity in gene expression: from theories to phenotypes. Nat. Rev. Genet. 6, 451–464 (2005).
2. McClure, W.R.: Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. 54, 171–204 (1985).
3. Uptain, S.M., et al.: Basic mechanisms of transcript elongation and its regulation. Annu. Rev. Biochem. 66, 117–172 (1997).
4. H . L . MacGillivray: Skewness and Asymmetry : Measures and Orderings. Ann. Stat. 14, 994–1011 (1986).
5. D. N. Joanes, C. A. Gill: Comparing Measures of Sample Skewness and Kurtosis. R. Stat. Soc. 47, 183–189 (1998).
6. Carpenter, J., Bithell, J.: Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. Stat. Med. 19, 1141–1164 (2000).
7. Muthukrishnan, A.B. et al.: In Vivo Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in Stressed Escherichia coli Cells. PLoS One. 9, e109005 (2014).
8. Lloyd-Price, J. et al.: Dissecting the stochastic transcription initiation process in live Escherichia coli. DNA Res. 23, 203–214 (2016).
9. Kandavalli, V.K., et al.: Effects of σ factor competition are promoter initiation kinetics dependent. Biochim. Biophys. Acta - Gene Regul. Mech. 1859, 1281–1288 (2016).