A strategy for dissecting the kinetics of transcription repression mechanisms

Cristina S.D. Palma[1,2], Sofia Startceva[1], Ramakanth Neeli-Venkata[1], Marzieh Zare[1], Nadia S.M. Goncalves[1], Jose M. Fonseca[2], Samuel M.D. Oliveira[1], and Andre S. Ribeiro[1,2]

[1] Laboratory of Biosystem Dynamics, BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, 33101, Tampere, Finland

[2] Computational Intelligence Group of CTS/UNINOVA. Faculdade de Ciências e Tecnologia da Univ. Nova de Lisboa, Portugal

Abstract— **Promoters in *Escherichia coli* include an 'OFF' state, during which transcription is halted. Here, we propose a novel empirical method for assessing the time-length spent by promoters in this state. It relies on direct measurements of RNA production kinetics at the single molecule level at different induction levels, followed by an estimation of the RNA production rate under infinite induction, which is then compared to this rate under real, maximum induction. We apply it to the LacO3O1 promoter and infer that, under full induction, on average, 15% of the time between successful transcription events is spent in the OFF state. We verify this result by comparing the kinetics of a mutant strain lacking repressor molecules with that of the inferred rate under infinite induction. We expect this strategy of dissecting the kinetics of transcription repression to be applicable to a wide number of promoters in *E. coli*.**

Keywords— **Transcription, Induction, τ Plot, OFF state.**

I. Introduction

Novel experimental techniques of microscopy and fluorescent molecular probes have led to the rapid acquisition of invaluable data on the dynamics of gene expression in live cells. One particularly valuable development has been the engineering of the MS2-GFP protein that has the ability to bind specific RNA sequences and, thus, provided multiple binding sites, detecting individual RNA molecules as soon as they are produced in live cells [1]. This technique allows both estimating RNA numbers in individual cells of a population at a given time [1] as well as obtaining RNA production time intervals [2]. This data greatly increased our knowledge on the *in vivo* dynamics of transcription.

Recently, a technique was developed for dissecting the dynamics of active transcription [3]. This method uses measurements of the RNA production dynamics from cells with differing RNA polymerase (RNAP) concentrations, which allows estimating what would be the rate of transcription in cells with infinite RNAP concentration.

Here, we propose a novel, similar methodology that uses data from cells with differing intracellular inducer concentrations, to further dissect the kinetics of transcription initiation. In particular, we focus on the promoter OFF state.

II. Methods

A. Cells, Plasmids, and Chemicals

We use *E. coli* strain BW25113 ($lacI^+$ $rrnB_{T14}$ $\Delta lacZ_{WJ16}$ *hsdR514* $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$) [4], which have the constitutive promoters P_{lacI+} and P_{araC} producing, respectively, LacI repressors for the LacO3O1 promoter [5] and AraC repressors for the BAD promoter. We also use the deletion mutant strain JW0336 (BW25113 *ΔlacI*), lacking the ability to express LacI repressor molecules.

In both strains, we placed a single-copy plasmid pBELOBAC11 carrying a $P_{lacO3O1}$ promoter [6] controlling the production of an RNA coding for 48 binding sites for MS2-GFP proteins (48BS). We also introduced a medium-copy plasmid, pZA25, with the reporter gene, P_{BAD}-MS2-GFP, responsible for producing the fusion protein MS2d-GFP, generously provided by Orna Amster-Choder (Hebrew University of Jerusalem, Israel) [7]. The activity of $P_{lacO3O1}$ is regulated by the repressor LacI and the inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG). The activity of P_{BAD} is regulated by AraC and the inducer L-arabinose.

This system has been used to measure the distribution of time intervals between RNA production events due to its ability to detect individual RNAs, as the MS2-GFP proteins rapidly bind to newly formed RNAs, which can be seen as fluorescent foci under a fluorescence microscope [1-3].

B. Growth Conditions

Cells were grown overnight in LB medium supplemented with appropriate antibiotics (34 μg/ml of Chloramphenicol and 50 μg/ml of Kanamycin) with shaking at 250 rpm. We subsequently made subcultures by diluting the stationary-phase culture into fresh M9 medium, supplemented with Glycerol (0.4% final concentration) along with the appropriate antibiotics. Cells were placed in the incubator until reaching OD600 of ~0.25. For the reporter plasmid activation, we added 0.4% of L-arabinose to the culture, which was incubated at 37°C for 60 minutes. Next, for the target plasmid, specific concentrations of IPTG (0, 5, 25, 50, 100, 250, 500, and 1000 μM) were added. Cells were then incubated for 120

minutes. In the end, cells were collected by centrifugation at 8000 × g for 1 minute, and diluted in fresh M9 medium. From this, 5 µL of cells were added to an M9 glycerol agarose gel pad, prior to microscopy observation.

C. Microscopy

Cells were imaged by a 488 nm argon laser (Melles-Griot), and an emission filter (HQ514/30, Nikon), using a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with a 100× Apo TIRF (1.49 NA, oil) objective. To obtain single-time-point images, we used a C2+ (Nikon) confocal laser-scanning system. Meanwhile, in time-lapse measurements, cells were imaged by Highly Inclined and Laminated Optical sheet (HILO) microscopy, using an EMCCD camera (iXon3 897, Andor Technology). In both cases, phase-contrast images were acquired by a CCD camera (DS-Fi2, Nikon), for purposes of cell segmentation. The software for image acquisition was NIS-Elements (Nikon, Japan).

In time-lapse microscopy, cells were constantly supplied with fresh media with IPTG/L-arabinose during image acquisition, at the same concentration as in liquid culture, by a micro-perfusion peristaltic pump (Bioptechs) at 0.3 ml/min. Images were captured for 2 hours, 1 per minute in the case of fluorescence and 1 per 5 minutes in the case of phase-contrast. Also, cells were kept in a temperature-controlled chamber (FCS2, Bioptechs) at optimal temperature (37 °C).

D. Image Analysis

Microscopy images were processed as in [3]. First, cells were detected from phase contrast images and then aligned with the confocal images. Fluorescent spots and their intensities were detected from the confocal images as in [8].

For population analysis, the intensity of one spot was calculated from the histogram of total intensity of fluorescent spots per cell, normalized by the intensity of the first spike on this histogram, as in [1].

For time series analysis, jumps in each cell's spot intensity over time were detected using a least-deviation jump-detection method [8]. Given the noise in the time series, the intensity of 'one RNA' was first selected automatically, and then corrected by manual inspection of the total foreground spot intensities, as in [3]. To avoid spurious jumps, we disregarded jumps within 5 min of the beginning or end of a cell's lifetime [3]. Censored intervals were calculated as the time from the last RNA production in a cell until the last time point when a jump could have been observed [8].

E. Model of transcription

RNA production starts with the freeing of the promoter for transcription, followed by the closed complex formation, which includes multiple binding events of RNAPs until the promoter commits to the next step, the open complex formation, and finally to the promoter escape and RNA production. We represent this in reactions (1) and (2). In these, the various k's represent the rate constants at which these events take place:

$$P_{OFF} \underset{k_{off}*[\mathrm{Rep}]}{\overset{k_{on}}{\rightleftarrows}} P_{ON} \quad (1)$$

$$P_{ON} \underset{k_{-1}}{\overset{k_1*[\mathrm{RNAP}]}{\rightleftarrows}} P_{cc} \xrightarrow{k_2} P_{oc}{}^{i} \xrightarrow{k_3} P_{oc}{}^{f} \xrightarrow{k_4} P_{ON} + RNA \quad (2)$$

In (1), it is represented the reversibility between ON and OFF states of the promoter P (represented by P_{ON} and P_{OFF}, respectively), due to, e.g., unbinding/binding of repressors (respectively). Another cause for OFF states could be the accumulation/release of local positive DNA super-coiling in the chromosomal integrated gene, generated by transcription events and the release by gyrases, respectively [9,10].

In (2), a promoter in the ON state proceeds to form a closed complex (P_{cc}), as an RNAP finds the start site. As such, this event depends both on the concentration of freely available RNAPs and on the rate with which one RNAP binds to the start site (k_1) At this stage, there is a significant chance, explicitly represented, that the promoter reverts to the previous stage. After several attempts, the promoter reaches, first, the initial stage of open complex formation (P_{oc}^{i}). From here onwards the process is nearly irreversible and, in the presence of Mg^{2+}, a fully formed open complex (P_{oc}^{f}) is created [11]. Subsequently, the RNAP escapes the promoter and proceeds with elongation, which leads to the production of a fully formed RNA. Finally, note that elongation is not considered explicitly, since it is a fast process relative to the previous events and since it does not affect the mean duration of intervals between consecutive RNA productions.

F. τ Plots

The mean time-length between transcription events can be altered by changing the free RNAP concentration, as demonstrated both *in vitro* and *in vivo* [3, 12-14]. Only the duration of the closed complex formation is affected by these changes in free RNAP concentration [3]. As such, from a set of measurements of mean interval durations between consecutive RNA production events in individual cells whose RNAP concentrations differ, it is possible to estimate the rate of RNA production for an infinite RNAP concentration. This rate should correspond to the inverse of the duration of the open complex formation [3].

Here, we use a similar strategy. However, instead of altering intracellular RNAP concentrations, we alter inducer concentrations. As shown below, this produces data that allows estimating the time spent in both the closed and open complex formation. Further, assuming the model of transcription in (1) and (2), confronting this data with the intervals between

consecutive RNA productions informs on the mean time that promoters spend in the OFF state.

III. Results

We first derive, assuming the model described by (1) and (2), the equations supporting our method of dissection of the transcription repression kinetics. From (1) and (2), the mean duration of the intervals between consecutive RNAs, Δt, is:

$$\Delta t = \tau_{OFF} + \tau_{cc} + \tau_{oc} \quad (3)$$

In (3), τ_{OFF} is the mean time spent in P_{OFF} state between two RNA production events (note that this mean time can, and likely does, result from multiple 'passages' through the OFF state between two consecutive RNA production events). Meanwhile, τ_{cc} is the mean time spent in P_{ON} state until forming a successful closed complex (again, the promoter will likely be several times in this state between two consecutive RNA production events). Finally, τ_{oc} is the mean time for a closed complex to successfully form an open complex. As such, it includes the time to change from P_{cc} to P_{oc}^{i} (open complex in initial state) along with the time to change from P_{oc}^{i} to P_{oc}^{f} (open complex in fully formed state). Since the open complex formation is physically nearly irreversible once initiated, usually it only occurs once between two consecutive RNA production events.

Of these, only τ_{OFF} is expected to differ with the inducer concentration, *provided that, e.g. the inducer acts by inactivating the repressor*, as in the case of IPTG [5]. Given this, we define Δt_{ind} as the mean time between transcription events aside from the OFF period, and rewrite (3) as:

$$\Delta t = \tau_{OFF} + \Delta t_{ind} \quad (4)$$

From the above definitions, changing IPTG levels will alter only τ_{OFF} (and thus Δt). I.e. assuming a new condition, differing in [IPTG] (within the range where changes in [IPTG] result in changes in the RNA production rate):

$$\Delta t^{new} = \tau_{OFF}{}^{new} + \Delta t_{ind} \quad (5)$$

Next, one can plot the mean Δt versus the inverse of the [IPTG], for various IPTG concentrations. On this data, one can do a linear fit to estimate Δt^{inf} (i.e. the expected Δt assuming infinite [IPTG]), as it is given by the height at which the line of the fit crosses the y-axis (see Fig. 2).

Infinite [IPTG] implies that all the repressor molecules in the cell should be inactive. As such, the validity of the inference can be tested, e.g., by measuring the transcription kinetics in cells lacking the ability to produce the repressor.

To implement this strategy, we must first find the induction levels that significantly differ in transcription rates. We thus measured RNA numbers in individual cells at various induction levels of LacO3O1, both to determine its maximum induction level, as well as to find the region of the induction curve where the transcription rate is sensitive to small changes in inducer concentration.

From the results in Figure 1, maximum induction occurs from 50 μM IPTG and beyond, while the region of the induction curve where the RNA production rate is most sensitive to changes in [IPTG] is between 0 and 50 μM.

Thus, we performed microscopy time-series of cells subject, respectively, to 5, 25 and 50 μM IPTG (Methods). From the images, we obtained the mean Δt for each condition (Methods). We also imaged deletion mutants lacking the repression mechanism (i.e. unable to express lacI repressor molecules). Results are shown in Table 1.

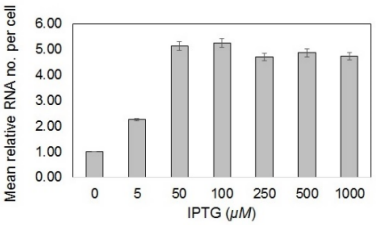

Fig. 1 Mean relative RNA numbers in individual cells. Images taken 2 hours after activating the target gene. The error bars are the standard error of the mean. In all conditions, more than 350 cells were observed.

Table 1 Empirical mean and uncertainty of the intervals between transcription events in individual cells for various induction levels

Condition	No. of cells	No. of intervals	Mean inferred interval and uncertainty (s)
5 μM	360	156	4362 ± 647
25 μM	92	30	2024 ± 709
50 μM	72	54	1922 ± 588
Deletion Mutant	44	23	1805 ± 757

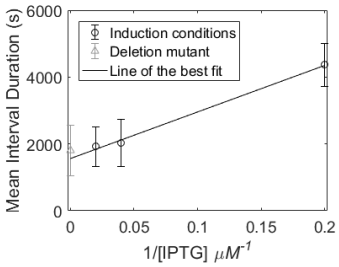

Fig. 2 τ -plot for LacO3O1, showing Δt of each induction level (circles) and standard uncertainties, along with best-fit line by least-squares fit. Also shown is the data from a mutant strain (triangle, not used in the fit).

From Table 1, following the strategy in [3], we obtained a τ-plot (Fig. 2), where the X-axis is the inverse of [IPTG], while the Y-axis is the mean duration of transcription intervals in individual cells. Next, we fitted a line to the data by the least squares method and extrapolated this linear fit to the

'infinite [IPTG]' condition. The intercept of the best fitted line with the Y axis is Δt^{inf}.

From Fig. 2, the best fit line intercepts the y-axis at *1559 s*. Since, under full induction (50 μM IPTG), the mean interval between transcription events is 1839 s *(according to the best fit line in Fig. 2)*, we expect the promoter to spend ~280 s in the OFF state between transcription events (15% of the time), under full induction.

To validate our result, we used a mutant strain lacking the ability to produce the repressor of the promoter of interest. The data from these cells is also in the τ plot and in Table 1, and, visibly, their RNA production rate is in agreement with the estimated rate for 'infinite' induction, from which we find the estimation to be reliable.

IV. Conclusions

We proposed a novel methodology for the estimation of the mean time spent by promoters in the OFF state between transcription events. To our knowledge, this is the first time this is achieved from data from live, individual cells.

At the moment, the method has several requirements that depend on the repression mechanism. E.g., it requires prior knowledge or measurements of how the inverse of the RNA production rate changes with the inverse of the inducer concentration. Here, this change was assumed to be linear within a certain range (which the data did not disprove), which allowed estimating the kinetics of RNA production under infinite induction levels. Also, it requires an inducer or activator that acts by freeing the promoter from an OFF state, rather than by affecting the kinetics of subsequent steps in transcription, in which case the outcome of the method will inform on different parameters than those in the present work.

We applied the methodology to LacO3O1 and its repression system. For validation, we measured the transcription kinetics in a deletion mutant for the repressor of LacO3O1, and found it to be in agreement with our method's estimation.

We believe that, in its present state, this method can already be of use to dissect the kinetics of repression mechanisms of transcription of various promoters of *E. coli*. We expect that several improvements can be further made to the present method, to allow its application to a broader range of repression (as well as induction) mechanisms.

Acknowledgment

Work supported by Academy of Finland (295027 and 305342 ASR), Jane & Aatos Erkko Foundation (610536 ASR), Tampere University of Technology President's Graduate Program (S.S. and R.N.-V.), Finnish Academy of Science and Letters (SO), Doctoral Programme of Computing and Electrical Engineering of TUT (NG) and Erasmus+ program 2919(713)2915/2016/SMS (C.P.). The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Golding I, Paulsson J, Zawilski S, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123:1025–36.
2. Muthukrishnan A-B et al (2014) *In vivo* Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in Stressed *Escherichia coli* cells. PLoS ONE 9(9): e109005
3. Lloyd-Price J et al. (2016) Dissecting the stochastic transcription initiation process in live *Escherichia coli*. DNA Res. 23:203–14.
4. Baba T et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol.2:1-11.
5. Glascock CB and Weickert MJ (1998) Using chromosomal lacI Q1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. Gene 223, 221–23.
6. NSM Goncalves et al. (2016) In vivo single-molecule dynamics of transcription of the viral T7 Phi 10 promoter in *E. coli*. In Proc. of the BIOTECHNO 2016, June 26-30, 2016 Lisbon, Portugal.
7. Nevo-Dinur K et al. (2011) Translation-independent localization of mRNA in *E. coli*. Science 331(6020):1081-4.
8. Häkkinen A and Ribeiro AS (2015) Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. Bioinformatics 31 (1): 69-75.
9. Chong S, Chen C, Ge H, and Xie XS (2014) Mechanism of Transcriptional Bursting in Bacteria. Cell 158, 314–326.
10. Liu L and Wang J (1987) Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. 84, 7024–7.
11. DeHaseth PL, Zupancic ML and Record MT (1998) RNA polymerase-promoter interactions: The comings and goings of RNA polymerase. J. Bacteriol. 180, 3019–25.
12. Liang S et al. (1999) Activities of constitutive promoters in *Escherichia coli*. J. Mol. Biol. 292:19–37.
13. McClure WR (1985) Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. 54:171–204
14. Ehrenberg M et al. (2013) Medium-dependent control of the bacterial growth rate. Biochimie. Elsevier Masson 95:643–58.

Corresponding Author: Andre S. Ribeiro
Institute: Tampere University of Technology
Street: Korkeakoulunkatu 10, 33720 Tampere
City: Tampere
Country: Finland
Email: andre.ribeiro@tut.fi