An Estimation Method of the Kinetic Rates of Transcription Initiation by $E\sigma^{70}$ and $E\sigma^{38}$ from Measurements of Individual RNA Productions

Huy Tran and Andre S. Ribeiro
Laboratory Biosystem Dynamics/Department of Signal Processing/Tampere University of Technology, Tampere, Finland
Email: huy.tran@tut.fi

***Abstract*—One of the global regulators of transcription dynamics in *Escherichia coli* is the intracellular population of σ factors, due to their role in gene selection for transcription. It is unknown to which degree σ factors affect the dynamics of transcription initiation, following the binding between the RNAP holoenzyme ($E\sigma$) and the promoter, and the closed complex formation. Proposed here is a new method to study the kinetics of the underlying steps in transcription initiation from time-lapse imaging of transcription events at the single RNA level in live cells. Namely, assuming a promoter that can be transcribed by $E\sigma^{70}$ or $E\sigma^{38}$, the researchers make use of *in silico* data from a stochastic model of transcription dynamics of that promoter, to show that the method estimates consistently and effectively the kinetics rates of closed and open complex formation by $E\sigma^{70}$ and $E\sigma^{38}$. In the end, the necessary measurement procedures for acquiring the data needed to apply this new methodology are described.**

Index Terms*—gene expression, computational biology, single molecule, *in vivo

I. Introduction

Escherichia coli can implement different gene expression profiles to cope with different stress conditions. One of the global regulators of gene expression is the intracellular population of σ factors. In normal growth conditions, most RNA polymerase core enzymes (E) are bound by the house-keeping σ factor, σ^{70} [1], to form holoenzymes ($E\sigma^{70}$), which only allow the expression of genes whose promoter region is recognized by σ^{70}. Meanwhile, in the stationary growth phase, the number of stress-responding σ^{38} sub-units increases, enhancing the competition with σ^{70} for the limited pool of E [2], thus altering the distribution of the holoenzymes ($E\sigma$) carrying each factor [3]. This change induces the expression of genes recognized by σ^{38} and hinders the expression of the remaining genes [4]-[7].

Despite the only known role of σ factors being to aid E to find specific sequences at the promoter regions [8], it is also known that they remain bound to E during the whole transcription initiation process and are only released stochastically after the start of transcription elongation [9], [10]. It is unknown, whether σ factors, while present in the RNA polymerase - promoter complex, affect significantly the dynamics of transcription initiation, following the closed complex formation.

An *in vivo* study of the role of σ factors on transcription initiation kinetics is expected to face challenges. First, it is not possible to observe transcription dynamics *in vivo* in the absence of σ^{70} or when overexpressing other σ factors to levels that would silence house-keeping genes. Second, to properly compare the dynamics of transcription as a function of the σ factor, one needs target promoters that can be transcribed, with comparable probabilities, by $E\sigma$ carrying different σ factors. Finally, altering σ factors numbers may lead to intracellular changes that will indirectly also affect the kinetics of transcription.

Here, we propose a method to study the *in vivo* dynamics of the underlying steps of transcription initiation when performed by $E\sigma$ carrying two different σ factors, namely σ^{70} and σ^{38}. In particular, we aim to quantify how fast $E\sigma$ finds and binds to the target promoter (i.e. the closed complex formation) and the average time of the open complex formation as function of the σ factor it carries. For that, we implement a stochastic model of single gene expression dynamics in the presence of both $E\sigma^{70}$ and $E\sigma^{38}$ and then present a method to estimate the kinetic rates of transcription initiation of the model from the measured distributions of time intervals between consecutive transcription events in individual, in different conditions. The data provided by the model mimics data that is possible to acquire using a well-known RNA fluorescence tagging method (the MS2d-GFP tagging method system [11]), which has been used recently to study the *in vivo* kinetics of transcription initiation of several promoters in *E. coli* has a function of induction scheme, temperature, stress conditions [12]-[15].

II. Methods

A. Full-Sized Camera-Ready (CR) Copy Model of Transcription Assuming Two Functional σ Factors

We follow the modeling strategy of transcription described in [12], [14], [16], [17], which was based on

Manuscript received June 16, 2015; revised August 9, 2015.

doi: 10.18178/ijpmbs.4.3.151-157

previous *in vitro* measurements of transcription kinetics [18], [19] and has been recently validated by *in vivo* measurements [12]-[14]. We model transcription performed by each of the two types of holoenzyme considered ($E\sigma^{70}$ and $E\sigma^{38}$) as a two rate-limiting step process as follow (1-4):

$$\mathrm{Pr} \xrightarrow{ind.P(E\sigma^{70}).k_{cc70}} \mathrm{Pr}_{cc70} \quad (1)$$

$$\mathrm{Pr} \xrightarrow{ind.P(E\sigma^{38}).k_{cc38}} \mathrm{Pr}_{cc38} \quad (2)$$

$$\mathrm{Pr}_{cc70} \xrightarrow{k_{oc70}} \mathrm{Pr}_{oc70} \xrightarrow{\infty} \mathrm{Pr} + RNA \quad (3)$$

$$\mathrm{Pr}_{cc38} \xrightarrow{k_{oc38}} \mathrm{Pr}_{oc38} \xrightarrow{\infty} \mathrm{Pr} + RNA \quad (4)$$

Reaction (1) and (2) model the binding of $E\sigma$ ($E\sigma^{70}$ and $E\sigma^{38}$ respectively) to the promoter region (Pr) to form the closed complex (Pr_{cc70} and Pr_{cc38}). $P(E\sigma^{70})$ and $P(E\sigma^{38})$ are the probabilities that E is bound by σ^{70} and σ^{38} respectively. Meanwhile, *ind*, which can take values ranging from 0 to 1, is the promoter's induction level, which can be regulated by varying external inducer concentrations.

We assume that only the duration of the closed complex formation is affected by the inducer level (as in the case of *lac* promoters induced by IPTG [14], [20]). Note that, since $E\sigma$ and the target promoter's repressors exist in cells in high copy numbers [2], [20], and given their fast association/disassociation to the promoter region [21], [22] when compared to the rate-limiting steps in transcription initiation [12]-[14], their numbers should only affect the mean rate of closed complex formation and therefore can be accounted for in the values of $P(E\sigma^{70})$, $P(E\sigma^{38})$ and *ind*. Finally, k_{cc70} and k_{cc38} are the rates of closed complex formation when the promoter is fully induced (*ind=1*) and the holoenzymes are mostly bound by one of the two σ factors ($P(E\sigma^{70})=1$ while $P(E\sigma^{38})=0$ and $P(E\sigma^{70})=1$ while $P(E\sigma^{38})=1$, respectively).

Reaction (3) and (4) describe the formation of open complex (Pr_{oc70} or Pr_{oc38}) at the rates k_{oc70} and k_{oc38}, respectively. The open complex is quickly followed by promoter escape (and return of the promoter to the primary state, *Pr*) and transcription elongation [23], [24]. The latter process ends with the release of a complete RNA. Elongation, being of the order of tenths of seconds [23], is assumed to be instantaneous, since initiation is of the order of 10^2-10^3 seconds [11]-[13].

We assume that, in the conditions tested, other σ factors exist in cells only in small copy number [3] and therefore occupy a negligible proportion of $E\sigma$. Therefore:

$$P(E\sigma^{38}) = 1 - P(E\sigma^{70}) \quad (5)$$

Meanwhile, from reaction (1) and (2), the rate of closed complex formation, regardless of the σ factor present in the transcribing $E\sigma$, is given by:

$$k_{cc} = ind.(P(E\sigma^{70}).k_{cc70} + P(E\sigma^{38}).k_{cc38}) \quad (6)$$

Once a closed complex is formed, the probability that the transcribing $E\sigma$ is $E\sigma^{70}$ or $E\sigma^{38}$ is given by, respectively:

$$P(\mathrm{Pr}_{cc70}) = \frac{P(E\sigma^{70}).k_{cc70}}{P(E\sigma^{70}).k_{cc70} + P(E\sigma^{38}).k_{cc38}} \quad (7)$$

$$P(\mathrm{Pr}_{cc38}) = \frac{P(E\sigma^{38}).k_{cc38}}{P(E\sigma^{70}).k_{cc70} + P(E\sigma^{38}).k_{cc38}} = 1 - P(\mathrm{Pr}_{cc70}) \quad (8)$$

Finally, the distribution of time intervals between RNA releases (Δt) is given by:

$$\Delta t = t_{cc} + t_{oc} \quad (9)$$

where t_{cc} and t_{oc} are the combined distributions of the durations of closed complex and open complex formations by $E\sigma^{70}$ and $E\sigma^{38}$. The probability distributions of t_{cc} and t_{oc} are given by (10) and (11), respectively:

$$P_{t_{cc}}(t) = k_{cc}.e^{-k_{cc}.t} \quad (10)$$

$$P_{t_{oc}}(t) = P(\mathrm{Pr}_{cc70}).k_{oc70}.e^{-k_{oc70}.t} + P(\mathrm{Pr}_{cc38}).k_{oc38}.e^{-k_{oc38}.t} \quad (11)$$

From (7), (8), (9), (10), and (11), with k_{cc70}, k_{cc38}, k_{oc70}, and k_{oc38} being promoter specific, it is possible to calculate the distribution of time intervals between transcription events for each pair of values of $P(E\sigma^{70})$ and *ind*.

Here, given the set of experimental procedures assumed, we make use of the fact that the value of $P(E\sigma^{70})$ depends on the host strain (e.g whether the gene encoding σ^{38} is deleted or not [25]) and the growth phase of the cells (i.e. whether cells are in the exponential or in the stationary growth phase [3]).

B. Inference on the Kinetic Parameters of Transcription Initiation

1) Kinetic rates of transcription by $E\sigma^{70}$ in the mutant strain lacking σ^{38}

To infer the kinetic parameters of transcription initiation by $E\sigma^{70}$, we fit the model of transcription to the data obtained from a deletion mutant strain lacking σ^{38}. Given the deletion, we assume that all core enzymes are bound by σ^{70}, that is $P(E\sigma^{70})_{MT}=1$. As such, the distributions of durations of the sequential steps in transcription initiation are given by:

$$P_{t_{cc}}(t) = \frac{k_{cc70}}{ind} \times e^{-\frac{k_{cc70}}{ind} \times t} \quad (12)$$

$$P_{t_{oc}}(t) = k_{oc70} \times e^{-k_{oc70} \times t} \quad (13)$$

First, we make use of measurements of consecutive RNA production intervals under full induction (*ind =1*)

and for partial induction strength ($ind = ind_{partial} <1$). The inducer levels to achieve full and partial induction are to be extracted from the induction curve of the promoter for varying inducer concentrations. For the case of partial induction, for simplicity, we choose the inducer concentration corresponding to 50% of the RNA level under full induction.

Next, we fit the model to the data using maximum likelihood, in order to infer the values of k_{cc70}, k_{oc70} and $ind_{partial}$. This method has been applied previously to infer the rates of closed and open complex formation in transcription initiation of σ^{70}-dependent promoters [12]-[14], [26].

2) Kinetic rates of transcription by $E\sigma^{38}$ in the wild type strain

Here, we make use of measurements of consecutive RNA production intervals in individual cells with the same inducer concentrations as above (ind=1 and ind=$ind_{partial}$) for the wild type strain, where both σ^{70} and σ^{38} are present. In the WT strain, we assume that $P(E\sigma^{70})_{WT} <1$ (this assumption holds when cells are in the stationary phase [3]). Next, given the values of k_{cc70}, k_{oc70} and $ind_{partial}$ estimated above, we fit the data of the wild type strain to the model using maximum likelihood and, finally, estimate the values of k_{cc38}, k_{oc38} and $P(E\sigma^{70})_{WT}$.

3) Test data

We tested the estimator's performance on *in silico* data generated from the model using the software SGNS2 [27], which uses the Stochastic Simulation Algorithm (SSA) [28] or the delay SSA [29] to simulate the dynamics of the model (depending on whether the model contains, or not, time delays [16]). For each set of promoter kinetic parameters (k_{cc70}, k_{oc70}, k_{cc38}, and k_{oc38}) we simulate the transcription activity in four conditions, differing in the value of $P(E\sigma^{70})$ (between $P(E\sigma^{70})_{WT}$ and *1*) and value of ind (between $ind_{partial}$ and 1). To achieve 50% of the maximum RNA level (under full induction), $ind_{partial}$ is set to:

$$ind_{partial} = \frac{1}{2 + k_{cc70} / k_{oc70}} \qquad (14)$$

In each case, we collect N samples, each of which corresponding to a time interval between two consecutive RNA productions. Due to the finite sampling frequency and limited measurement time, each sample is rounded to the nearest multiple of 30 s and samples with values greater than 14400 s (4 hours) are discarded.

For each set of data (4×N samples), we estimate the values of both the promoter-specific parameters (k_{cc70}, k_{oc70}, k_{cc38}, k_{oc38}) and the host-specific parameters ($P(E\sigma^{70})_{WT}$ and $ind_{partial}$). In search for the parameter set that maximizes the likelihood functions, the search range for k_{cc70}, $k_{oc70,}$ k_{cc38} and k_{oc38} is set between 1/3000 s^{-1} and 1/30 s^{-1} and $ind_{partial}$ is set between 0 and 1. Finally, the search range for $P(E\sigma^{70})_{WT}$ is set between 0.5 and 0.9, based on previous reports on the level of σ factors [2], [3] and their binding affinity to *E* [1], [30].

4) Assessment of the performance of the estimator

To assess the estimator's performance, we investigate how accurate the estimator predicts, from the *in silico* data, the values of t_{cc70}, t_{oc70}, t_{cc38}, *and* t_{oc38}, which equal $1/k_{cc70}$, $1/k_{oc70}$, $1/k_{cc38}$, and $1/k_{oc38}$ respectively. To the level of error and bias in the prediction of the simulation parameters from the simulation data in regards to the magnitude of the parameters, we calculate the Normalized Mean Squared Error (NMSE) and the Normalized Bias (NB) of the estimator, which are given by:

$$NMSE(\theta, x) = MSE(\theta, x) / \|\theta\|^2 \qquad (15)$$

$$NB(\theta, x) = B(\theta, x) / \|\theta\| \qquad (16)$$

in which $MSE(\theta, x)$ and $B(\theta, x)$ are respectively the mean squared error and the bias of the estimator given the sample set x and the parameter being estimated θ. The normalization coefficient $\|\theta\|$ equals θ for the estimated parameter being $P(E\sigma^{70})_{WT}$ and $ind_{partial.}$ If the estimated parameter is t_{oc70} or t_{oc70}, $\|\theta\|=t_{cc70}+t_{oc70}$ (i.e. the mean transcription interval when $P(E\sigma^{70})=1$). If the estimated parameter is t_{oc38} or t_{oc38}, $\|\theta\|=t_{cc38}+t_{oc38}$ (i.e. the mean transcription interval when $P(E\sigma^{70})=0$).

III. Results

A. Kinetic Rates of Transcription by $E\sigma^{70}$

We assessed the performance of the estimator of the kinetic parameters of transcription by $E\sigma^{70}$ and the induction level ind using simulated data produced by the stochastic model. The tested values of k_{cc70} and k_{oc70} are $1/t_{cc70}$ and $1/t_{oc70}$ respectively, with t_{cc70} and t_{oc70} varying from 2 minutes to 30 minutes with the increment of 2 minutes. The range of values of t_{cc70} and t_{oc70} are in agreement with reported values in previous studies using the single RNA tagging system [12]-[14]. The evaluation results of the estimator are shown in Fig. 1, for N=500.

From Fig. 1, the inference of t_{cc70}, t_{oc70} using maximum likelihood from the data of different induction levels achieves high accuracy (NMSE smaller than 0.1) in most of the parameter space tested. The normalized biases in the estimation of t_{oc70} and t_{cc70}, as shown in Fig. 1B and 1C, are smaller than 0.1, ensuring that given an infinite number of samples, the errors in the estimation of t_{cc70} and t_{oc70} are less than 10% of the mean production interval measured under full induction (i.e. $t_{cc70}+t_{cc70}$).

The error level is significant only when $t_{cc70}<<t_{oc70}$ and when t_{cc70} is greater than 30 minutes. In this regime, t_{cc70} is over estimated. t_{oc70} is under estimated. When $t_{cc70}<<t_{oc70}$, the changes in the induction level ind should not result in significant changes in the dynamics of transcription, since step affected (the closed complex formation) is much shorter in duration than other step. For t_{cc70} greater than 30 minutes, the transcription intervals at partial induction are subject to right censoring due to the limited measurement time, which affects negatively the accuracy of the inference.

We also tested the performance of the estimator for greater values of N (i.e. 700, 1000). The results show that

the estimator is consistent in most of the parameter space tested, except in the regime when $t_{cc70} << t_{oc70}$. When t_{cc30}=2 mins and t_{oc70}=30 mins, NMSE(t_{cc70}, t_{oc70})~1.50 for all N>500, suggesting that these errors in the estimation are mostly due to the limitations imposed by the sampling frequency and the measurement time.

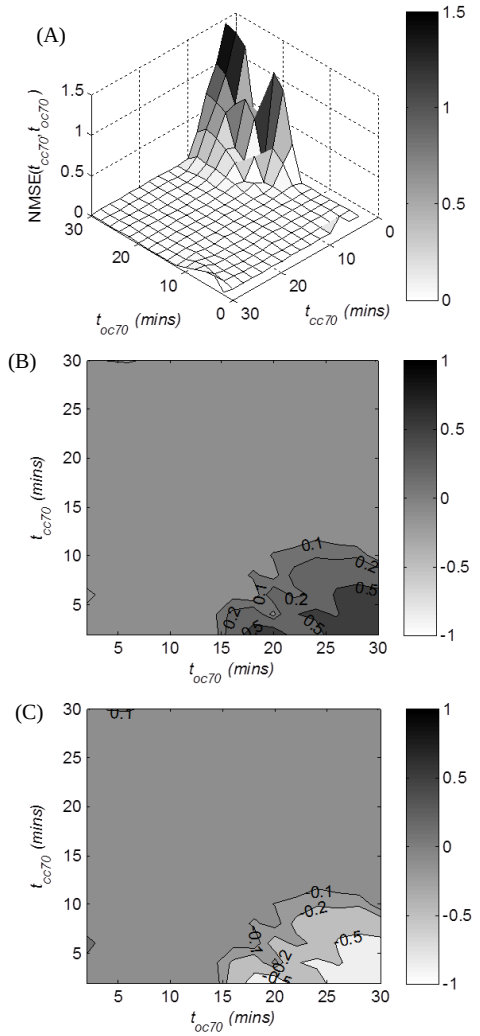


Figure 1. Estimation of kinetic rates of transcription by $E\sigma^{70}$. (A) NMSE with the estimated parameter θ being [t_{cc70} t_{oc70}]. (B)(C) Normalized Bias for the estimation of t_{cc70} and t_{oc70} respectively.

B. Kinetic Rates of Transcription by $E\sigma^{38}$

With k_{cc70}, $k_{oc70,}$ $ind_{partial}$ estimated from the condition $P(E\sigma^{70})=P(E\sigma^{70})_{MT}$ = 1, we evaluate the performance of the estimator of t_{cc38} and t_{oc38} assuming that $P(E\sigma^{70})=P(E\sigma^{70})_{WT}$ =0.75 [3]. Here, we set t_{cc70}=*10 min* and t_{oc70}=*10 min*, following the measured duration of these steps in the cases of lac-ara1 and BAD promoters [13], [14] in live cells. The ranges of k_{cc38} and k_{oc38} tested are from 2 minutes to 30 minutes, with the increment of 2 minutes. The estimator's performance (with N=500) is shown in Fig. 2.

From Fig. 2, it is shown that the performance of the estimator on t_{cc38} and t_{oc38} is not better than that on t_{cc70} and $t_{oc70.}$ This is expected due to the existing errors in the estimated values of t_{cc70} and t_{oc70}. The estimation of t_{cc38} and t_{oc38} has NMSE smaller than 0.2 within most of the parameter space tested, except in the regime of small $t_{cc38} << t_{oc38}$. From Fig. 2(A) and 2(B), t_{cc38} is overestimated with increasing t_{cc38} and t_{oc38} is overestimated with increasing $t_{oc38.}$ However, in most cases, the biases in the estimation of t_{cc38} and t_{oc38} is less than 30% except when $t_{cc38} >> t_{oc38}$ and when $t_{cc38} >> t_{oc38.}$

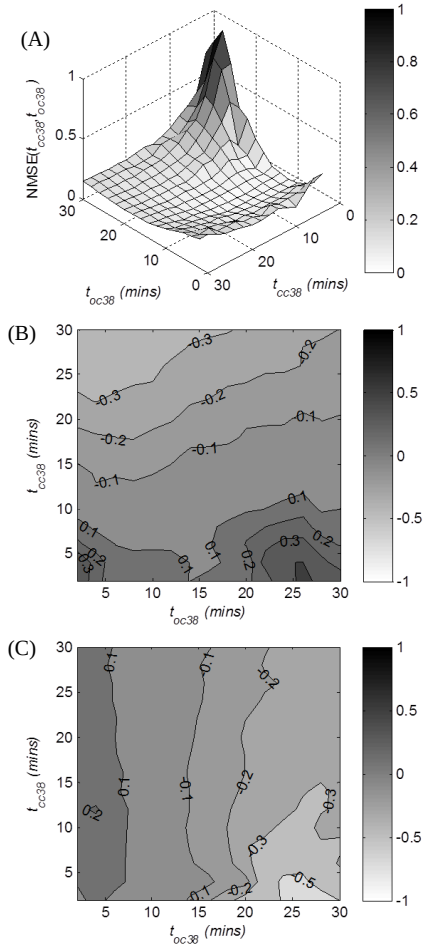


Figure 2. Estimation of kinetic rates of transcription by $E\sigma^{38}$. (A) NMSE with the estimated parameter θ being [t_{cc38} t_{oc38}]. (B)(C) Normalized Bias for the estimation of t_{cc38} and t_{oc38} respectively.

Interestingly, we also found that, despite of the accurate prediction of k_{cc38} and k_{oc38}, the inferred value of $P(E\sigma^{70})$ is close-to-uniformly distributed over the search change (data not shown). This indicates that the estimator can work well without knowing the real value of $P(E\sigma^{70})$.

We also tested the performance of the estimator for greater values of N (i.e. 700, 1000). The results show that the estimator is consistent throughout most of the parameter space tested and thus can be improved with increasing sample number.

IV. Implementation

In this section we describe a possible, realistic implementation of the methods proposed.

In order to compare the transcription dynamics when performed by $E\sigma^{70}$ and $E\sigma^{38}$, it is required that the target promoter can be recognized by either σ factor and that k_{cc70} is on the same order of magnitude as k_{cc38}. Studies of the sequence-dependence of promoter's selectivity for σ

factors [31], [32] show that both σ^{70} and σ^{38} can recognize a consensus sequence at the -10 position from the transcription starting site and that a promoter's preference for σ^{38} can be enhanced with the degeneration of the consensus at the -35 position [8], [31], [32]. Provided this, and given that some natural promoters can be transcribed by $E\sigma^{70}$ and $E\sigma^{38}$ [33], [34], it is reasonable to assume that it is possible to construct synthetic promoters that can be recognized by both σ^{38} and σ^{70}, e.g. by scrambling the -35 region of a known σ^{70}-dependent promoter. We propose the use of, e.g., the well-studied lac-ara1 [20] promoter as the template promoter given its effective regulation mechanism by IPTG and arabinose [14], [20]. To degenerate the consensus sequence at -35 region, while maintaining the promoter strength within reasonable levels, we propose the strategy described in [35], [36], for adjusting the "interaction energy" of the consensus boxes at position -35 and -10 with RNA polymerases.

Given this, the *E. coli* strains employed could be DH5α-PRO (generously provided by I. Golding, University of Illinois, USA) and its deletion mutant lacking the gene encoding σ^{38}. The mutant strain is created by following the protocol in [25]. DH5α-PRO is a genuine producer of araC and lacR [20], lac-ara1's repressors, which allows tight and homogeneous regulation of lac-ara-1 by IPTG and arabinose [14].

Finally, the MS2d-GFP RNA tagging system [11] can be used to monitor the dynamics of transcript production (diagram in Fig. 3(A)). The tagging system requires a high copy plasmid carrying $P_{LtetO-1}$-MS2d-GFP and a single-copy target plasmid coding for an RNA sequence with 96 target binding sites (96bs) for MS2d-GFP, controlled by the promoter of interest. Upon transcribed, the binding sites are bound by MS2d-GFP proteins and the target RNA appears on the confocal microscope as a bright spot.

For all experiments cells should be in the stationary growth phase, so that σ^{38} is expressed and occupies a significant amount of $E\sigma$ [4], [5], [37], allowing the observation of transcription activity with a mixture of $E\sigma^{70}$ and $E\sigma^{38}$ ($P(E\sigma^{70})_{WT}$<1). In the mutant strain lacking σ^{38}, most E will be bound by σ^{70} [4], [5], [37] ($P(E\sigma^{70})_{MT}$=1). The strategy to induce the stationary growth is described in [38]. The induction level of the promoter (*ind*) is varied by changing IPTG concentrations in the media. The IPTG concentrations to achieve full and partial induction can be determined from the induction curve obtained from qPCR measurements.

The measurement protocol and data analysis to be followed is described in [14], [15], [39]. Microscopy measurements are typically 4 hour long, with cells being imaged every 30s. To maintain, during microscopy, stable growth conditions and induction of the promoters controlling the production of MS2d-GFP and of the RNA target for MS2d-GFP, a peristaltic pump will be used to introduce a constant flow of phase-inducing media and inducers. An example of the results of applying these methods is shown in Fig. 3(B) and 3(C).

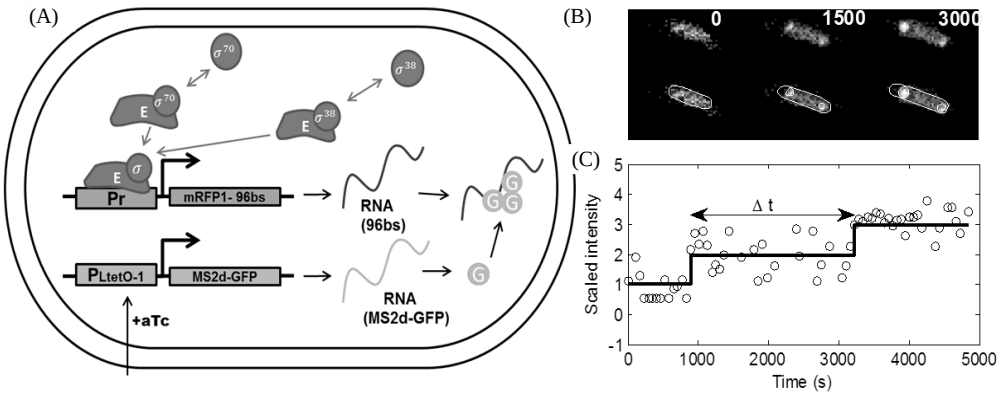


Figure 3. (A) Diagram of the σ factors - RNA polymerase (E) and RNA polymerase - target promoter (Pr) interactions, and the RNA tagging system by MS2d-GFP. (B) Unprocessed frames and segmented cells and RNA spots. (C) Examples of time series of scaled spot intensity levels from one cell (circles) and the corresponding estimated RNA numbers (solid lines), from which the transcription intervals (Δt) are extracted.

V. Conclusions

In this work, we proposed a method that, from the *in vivo* dynamics of RNA production at the single event level, allows estimating the kinetic rates of the closed complex and the open complex formation when performed by $E\sigma^{70}$ and $E\sigma^{38}$, provided that the promoter of interest can be transcribed by $E\sigma$ carrying either σ^{70} or σ^{38}.

From the estimator's performance on the simulation data, it is shown that, given a realistic number of measurements, this method estimates effectively the kinetics rates of closed and open complex formation by $E\sigma^{70}$ and $E\sigma^{38}$, provided that these rates are on the same order of magnitude. Also described here is the necessary measurement procedures for acquiring the data needed to apply this new methodology.

Relevantly, the collection of data on gene expression should be performed in well-controlled conditions. That is, one should verify that when switching between strains and inducer concentrations, the total levels of RNA polymerases, repressors (either active or inactive) and other σ factors are not significantly affected.

While it is known that the binding affinity of Eσ to the promoter region depends on the σ factor [32]

(consequently, the duration for the closed complex formation is σ factor dependent), it remains unknown whether the kinetics of the open complex formation is, or not, σ factor dependent. The method proposed here will allow addressing this question. Also, provided such a dependency, the method will allow us to determine whether the degree of influence of the choice of σ factor is promoter sequence-dependent.

REFERENCES

[1] H. Maeda, N. Fujita, and A. Ishihama, "Competition among seven *Escherichia coli* σ subunits: Relative binding affinities to the core RNA polymerase," *Nucleic Acids Res.*, vol. 28, pp. 3497-3503, 2000.

[2] I. L. Grigorova, N. J. Phleger, V. K. Mutalik, and C. A. Gross, "Insights into transcriptional regulation and sigma competition from an equilibrium model of RNA polymerase binding to DNA," *Proc. Natl. Acad. Sci.* U. S. A., vol. 103, pp. 5332-5337, 2006.

[3] M. Jishage, A. Iwata, S. Ueda, and A. Ishihama, "Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: Intracellular levels of four species of sigma subunit under various growth conditions," *J. Bacteriol.*, vol. 178, pp. 5447-5451, 1996.

[4] A. Farewell, K. Kvint, and T. Nystro, "Negative regulation by RpoS: A case of sigma factor competition," *Mol. Microbiol.*, vol. 29, pp. 1039-1051, 1998.

[5] T. Dong and H. E. Schellhorn, "Control of RpoS in global gene expression of *Escherichia coli* in minimal media," *Mol. Genet. Genomics*, vol. 281, pp. 19-33, 2009.

[6] M. Rahman, M. R. Hasan, O. Takahiro, and S. Kazuyuki, "Effect of rpoS gene knockout on the metabolism of *Escherichia coli* during exponential growth phase and early stationary phase based on gene expressions, enzyme," *Biotechnol. Bioeng.*, vol. 94, pp. 585-595, 2006.

[7] D. Chang, D. J. Smalley, and T. Conway, "Gene expression profiling of *Escherichia coli* growth transitions: An expanded stringent response model," *Mol. Microbiol.*, vol. 45, pp. 289-306, 2002.

[8] T. M. Gruber and C. A. Gross, "Multiple sigma subunits and the partitioning of bacterial transcription space," *Annu. Rev. Microbiol.*, vol. 57, pp. 441-466, 2003.

[9] R. A. Mooney, S. A. Darst, and R. Landick, "Sigma and RNA polymerase: An on-again, off-again relationship?" *Mol. Cell.*, vol. 20, pp. 335-345, 2005.

[10] M. Raffaelle, E. I. Kanin, J. Vogt, R. R. Burgess, and A. Z. Ansari, "Holoenzyme switching and stochastic release of sigma factors from RNA polymerase *in vivo*," *Mol. Cell*, vol. 20, pp. 357-366, 2005.

[11] I. Golding and E. C. Cox, "RNA dynamics in live *Escherichia coli* cells," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 101, pp. 11310-11315, 2004.

[12] A. B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, *et al*., "Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells," *Nucleic Acids Res.*, vol. 40, pp. 8472-8483, 2012.

[13] J. Mäkelä, M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, *et al*., "*In vivo* single-molecule kinetics of activation and subsequent activity of the arabinose promoter," *Nucleic Acids Res.*, vol. 41, pp. 6544-6552, 2013.

[14] M. Kandhavelu, J. Lloyd-Price, A. Gupta, A. B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro, "Regulation of mean and noise of the *in vivo* kinetics of transcription under the control of the lac/ara-1 promoter," *FEBS Lett.*, vol. 586, pp. 3870-3875, 2012.

[15] M. Kandhavelu, A. Häkkinen, O. Yli-Harja, and A. S. Ribeiro, "Single-molecule dynamics of transcription of the lar promoter," *Phys. Biol.*, vol. 9, p. 026004, 2012.

[16] A. S. Ribeiro, R. Zhu, and S. A. Kauffman, "A general modeling strategy for gene regulatory networks with stochastic dynamics," *J. Comput. Biol.*, vol. 13, pp. 1630-1639, 2006.

[17] H. Buc and W. R. McClure, "Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps," *Biochemistry*, vol. 24, pp. 2712-2723, 1985.

[18] W. R. McClure, "Mechanism and control of transcription initiation in prokaryotes," *Annu. Rev. Biochem.*, vol. 54, pp. 171-204, 1985.

[19] R. Lutz, T. Lozinski, T. Ellinger, and H. Bujard, "Dissecting the functional program of *Escherichia coli* promoters: The combined mode of action of Lac repressor and AraC activator," *Nucleic Acids Res.*, vol. 29, pp. 3873-3881, 2001.

[20] R. Lutz and H. Bujard, "Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements," *Nucleic Acids Res.*, vol. 25, pp. 1203-1210, 1997.

[21] L. J. Friedman, J. P. Mumm, and J. Gelles, "RNA polymerase approaches its promoter without long-range sliding along DNA," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 110, pp. 9740-9745, 2013.

[22] J. Elf, G. W. Li, and X. S. Xie, "Probing transcription factor dynamics at the single-molecule level in a living cell," *Science*, vol. 316, pp. 1191-1194, 2007.

[23] K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, *et al*., "Sequence-resolved detection of pausing by single RNA polymerase molecules," *Cell*, vol. 125, pp. 1083-1094, 2006.

[24] L. M. Hsu, "Promoter clearance and escape in prokaryotes," *Biochim. Biophys. Acta.*, vol. 1577, pp. 191-207, 2002.

[25] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, *et al*., "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: The Keio collection," *Mol. Syst. Biol.*, vol. 2, p. 2006.0008, 2006.

[26] H. Mannerstrom, O. Yli-Harja, and A. S. Ribeiro, "Inference of kinetic parameters of delayed stochastic models of gene expression using a Markov chain approximation," *Eurasip. J. Bioinforma. Syst. Biol.*, vol. 2011, pp. 11-15, 2011.

[27] J. Lloyd-Price, A. Gupta, and A. S. Ribeiro, "SGNS2: A compartmentalized stochastic chemical kinetics simulator for dynamic cell populations," *Bioinformatics*, vol. 28, pp. 3004-3005, 2012.

[28] D. T. Gillespie, "A general method for numerically simulating the stochastic time evolution of coupled chemical reactions," *J. Comput. Phys.*, vol. 22, pp. 403-434, 1976.

[29] M. R. Roussel and R. Zhu, "Stochastic kinetics description of a simple transcription model," *Bull. Math. Biol.*, vol. 68, pp. 1681-1713, 2006.

[30] A. Ganguly and D. Chatterji, "A comparative kinetic and thermodynamic perspective of the σ-competition model in *Escherichia coli*," *Biophys. J.*, vol. 103, pp. 1325-1333, 2012.

[31] G. Becker and R. Hengge-Aronis, "What makes an *Escherichia coli* promoter sigma(S) dependent? Role of the -13/-14 nucleotide promoter positions and region 2.5 of sigma(S)," *Mol. Microbiol.*, vol. 39, pp. 1153-1165, 2001.

[32] R. Hengge-Aronis, "Stationary phase gene regulation: What makes an *Escherichia coli* promoter σS -selective," *Curr Opin Microbiol.*, vol. 5, pp. 591-595, 2002.

[33] C. Peano, J. Wolf, J. Demol, E. Rossi, L. Petiti, G. De Bellis, *et al*., "Characterization of the *Escherichia coli* σS core regulon by Chromatin Immunoprecipitation-sequencing (ChIP-seq) analysis," *Sci. Rep.*, vol. 5, p. 10469, 2015.

[34] B. K. Cho, D. Kim, E. M. Knight, K. Zengler, and B. O. Palsson, "Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states," *BMC Biol.*, vol. 12, p. 4, 2014.

[35] J. B. Kinney, A. Murugan, C. G. Callan, and E. C. Cox, "Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence," *Proc. Natl. Acad. Sci. U. S. A.*, vol. 107, pp. 9158-9163, 2010.

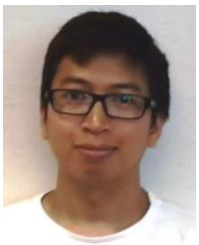
[36] R. C. Brewster, D. L. Jones, and R. Phillips, "Tuning promoter strength through RNA polymerase binding site design in *Escherichia coli*," *PLoS Comput. Biol.*, vol. 8, 2012.

[37] T. Dong, R. Yu, and H. Schellhorn, "Antagonistic regulation of motility and transcriptome expression by RpoN and RpoS in *Escherichia coli*," *Mol. Microbiol.*, vol. 79, pp. 375-386, 2011.

[38] G. Sezonov, D. Joseleau-Petit, and R. D'Ari, "*Escherichia coli* physiology in Luria-Bertani broth," *J. Bacteriol.*, vol. 189, pp. 8746-8749, 2007.

[39] A. B. Muthukrishnan, A. Martikainen, R. Neeli-Venkata, and A. S. Ribeiro, "*In vivo* transcription kinetics of a synthetic gene

uninvolved in stress-response pathways in stressed *Escherichia coli cells*," *PLoS One*, vol. 9, p. e109005, 2014.



Huy Tran obtained his BSc degree in electronics and telecommunication in Hanoi University of Technology (2010) and MSc degree in signal processing in Tampere University of Technology (TUT) (2013). He is currently a PhD student in the Laboratory of Biosystem Dynamics (LBD) at the Department of Signal Processing (DSP), TUT, Finland, since 2013. His topics of study include the dynamic of single gene expression and small gene networks, and the control of their dynamics using environment cues, such as temperature, inducer and nutrient availability.

Andre S. Ribeiro graduated in Physics, University of Lisbon, Portugal (1999). He obtained his PhD in physics engineering from IST, University Tecnica de Lisboa, Portugal (2004). From 2004-2007, he was a Postdoc at the University of Calgary, Canada. Since 2008, he is the PI of the Laboratory of Biosystem Dynamics (LBD) at the DSP, TUT, Finland. He is also an associate professor at the DSP, TUT. The LBD focuses on studies of regulatory mechanisms of gene expression and genetic circuits in *Escherichia coli* from single-cell, single-molecule *in vivo* measurements and stochastic models, combined with advanced signal processing techniques.