Temperature-dependence of the single-cell variability in the kinetics of transcription activation in *Escherichia coli*

Nadia S.M. Goncalves[†], Sofia Startceva[†], Cristina S.D. Palma[†, ‡], Mohamed N.M. Bahrudeen[†], Samuel M.D. Oliveira[†] and Andre S. Ribeiro[†, §, *]

Running Head: Temperature-dependence of transcription activation in *E. coli*

Keywords: Single-cell Intake kinetics; gene expression activation times; *Escherichia coli*; critically low temperatures;

[†] Laboratory of Biosystem Dynamics, BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, 33101, Tampere, Finland.
[‡] CA3 CTS/UNINOVA. Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal.
[§] Multi-scaled Biodata Analysis and Modelling Research Community, Tampere University of Technology, 33101, Tampere, Finland.
[*] Corresponding author. E-mail: andre.ribeiro@tut.fi, Tel: +358408490736.

Abstract

From *in vivo* single-cell, single-RNA measurements of the activation times and subsequent steady-state active transcription kinetics of a single-copy Lac-ara-1 promoter in *Escherichia coli*, we characterize the intake kinetics of the inducer (IPTG) from the media, following temperature shifts. For this, for temperature shifts of various degrees, we obtain the distributions of transcription activation times as well as the distributions of intervals between consecutive RNA productions following activation in individual cells. We then propose a novel methodology that makes use of deconvolution techniques to extract the mean and the variability of the distribution of intake times. We find that cells, following shifts to low temperatures have higher intake times, although, counter-intuitively, the cell-to-cell variability of these times is lower. We validate the results using a new methodology for direct estimation of mean intake times from measurements of activation times at various inducer concentrations. The results confirm that *E. coli*'s inducer intake times from the environment are significantly higher, following a shift to a sub-optimal temperature. Finally, we provide evidence that this is likely due to the emergence of additional rate-limiting steps in the intake process at low temperatures, explaining the reduced cell-to-cell variability in intake times.

Introduction

RNA and protein numbers differ between cells of monoclonal populations due to the stochastic nature of the chemical reactions composing gene expression ('intrinsic' noise) [1,2] and the cell-to-cell variability in the numbers of the molecules involved ('extrinsic' noise) [3].

Besides these 'constant' sources of cell-to-cell variability, recent studies have shown that, following the appearance of an inducer of gene expression in the media, there is an additional transient cell-to-cell diversity in RNA and protein numbers of the target gene [4–6], which cannot be explained by the intrinsic and extrinsic noise of active gene expression. This additional source can be strong enough and the transient long enough to affect the phenotypic diversity of cell lineages for generations [4–12].

The origin of this transient phenotypic diversity has been shown to be the noise in the intake time of the inducers, which causes the time for transcription to be activated (following the introduction of the inducers in the media) to differ widely between cells [5]. At the RNA numbers level, this transient diversity can be higher than the diversity caused by the intrinsic and extrinsic noise in active transcription for long periods of time [5].

Similarly to noise in gene expression, noise in intake times has two sources. One is the stochasticity of the intake process, caused by the random nature of the chemical reactions and the membrane crossing processes [2,6]. The other is likely a non-negligible degree of cell-to-cell heterogeneity in the efficiency of the mechanisms involved in the intake of inducers [5]. This heterogeneity can be caused by, among other, cell-to-cell diversity in the number of transmembrane proteins involved in the active uptake of

inducer/repressor molecules [5]. One example is the lactose permease (LacY), which, while being produced by an all-or-nothing system that minimizes cellular heterogeneity, it nevertheless exhibits significant cell-to-cell diversity in numbers, following the appearance of the inducer (e.g. TMG) in the media [13].

As natural environmental conditions fluctuate and many genes in *E. coli* are only activated in specific conditions, cellular heterogeneity in gene expression activation times is expected to affect significantly the phenotypic diversity of cell populations.

One environmental parameter that we expect to have a tangible impact on both the mean and variability of intake times of external inducers and repressors of gene expression is temperature. This assumption originates from the fact that temperature affects not only proteins functionality and numbers in cells [14], but also the physical properties of cell walls, periplasm and cytoplasm (e.g. the cytoplasm's viscosity is temperature dependent [15]), and these variables are expected to affect the kinetics of intake of inducers from the environment.

However, there is yet no direct experimental validation and, as many variables are involved, model-based predictions of the quantitative degree of changes with temperature in inducers intake times and subsequent transcription initiation times are unreliable.

Here, we characterize quantitatively the changes in cell-to-cell variability in gene expression activation times of the Lac-ara-1 promoter and, more importantly, of the intake times of its inducer, Isopropyl β-D-1-thiogalactopyranoside (IPTG), caused by rapid physical changes following temperature shifts.

For this, we use time-lapse microscopy measurements of RNA production at the single-cell, single-RNA level at various temperatures, along with several recently developed techniques [6,14,16], including a new strategy here proposed to dissect the kinetics of the intake process. Our results provide novel information for the understanding of the effects of temperature shifts of bacterial populations at the single-cell level.

Methods

Bacterial strains and plasmids

We use *E. coli* strain DH5α-PRO, generously provided by I. Golding, University of Illinois, U.S.A. The genotype is deoR, endA1, gyrA96, hsdR17(rK^- mK^+), recA1, relA1, supE44, thi-1, Δ(lacZYA-argF)U169, Φ80δlacZΔM15, F^-, λ^-, P_{N25}/tetR, P_{lacIq} /lacI, SpR. The strain contains two genes, *lacI* and *tetR*, constitutively expressed under the control of $P_{laci}{}^q$ and P_{N25} promoters, respectively [17]. Relevantly, the native lac operon (*lacZYA*) is mutated, to prevent production of permease (*lacY*) and activation of the lactose metabolic system [18]. I.e., these cells lack the native positive feedback mechanism involving lactose [6,19].

In addition to this strain, we also use *E. coli* JW0334 strain. The genotype is F^- (Δ (araD-araB)567 ΔlacY784 ΔlacZ4787(::rrnB-3) λ^-rph-1 Δ(rhaD-rhaB)568 hsdR514) [18].

This strain also lacks the ability to produce lacY [18]). Here, we only make use of this strain to show that the changes in the target gene activation time with temperature are, qualitatively, only weakly strain dependent. Unless stated otherwise, measurements are made using DH5α-PRO cells.

Both strains lack the ability to express lacY permease [18], which is responsible for a feedback response to the intake of IPTG, which would result in more complex, time-dependent single-cell intake times, as they would not be solely determined by the induction level and temperature.

Two constructs were added to DH5α-PRO cells: pROTET-K133 with $P_{LtetO-1}$-MS2d-GFP and pIG-BAC, a single-copy plasmid with $P_{Lac-ara-1}$-mRFP1-MS2d-96bs [20] (Figure 1). In the case of JW0334 cells, another reporter is used (P_{RHAM}-MS2d-GFP), as these cells lack the ability to express TetR.

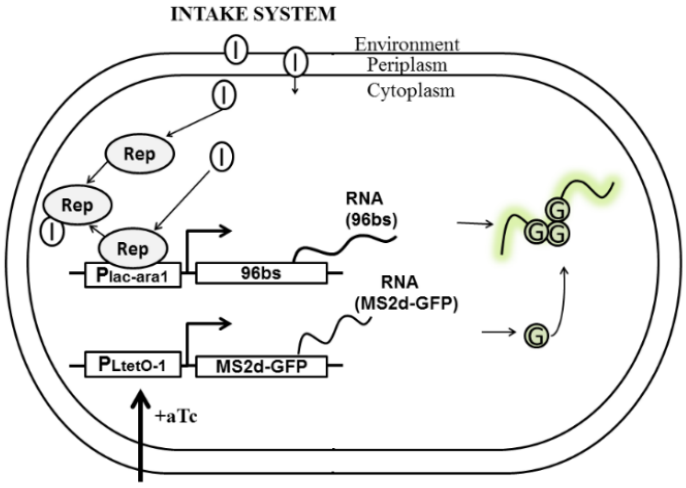

Figure 1. Diagram of the target gene and its RNA tagging system, along with the intake system of inducers of the target gene: IPTG molecules (I) are added to the media and enter the cytoplasm by passing through two membrane layers, with a periplasmic space in between. When in the cytoplasm, they neutralize lacI repressors (R) by forming inducer-repressor complexes (RI). This allows $P_{Lac-ara-1}$ to express RNAs that include an array of 96 MS2d-binding sites. Meanwhile, MS2d-GFP expression is controlled by the $P_{LtetO-1}$ promoter and anhydrotetracycline (aTc). Once produced, each target RNA is rapidly bound by multiple tagging MS2d-GFP proteins (G), and appears as a bright spot, significantly above background fluorescence, under the confocal microscope [6,20]. The tagging provides the RNA a long lifetime, with constant fluorescence, beyond our observation times [6].

Finally, it is noted that previous measurements [6] have shown that, provided full induction of the reporter gene (1 hour) prior to induction of the target gene, any newly produced target RNA molecule becomes 'fully fluorescent' (i.e. its RNA MS2-GFP binding

sites become fully occupied) in less than 1 minute. These measurements were conducted in the same strain and media employed here. Given this, and since our microscopy time-lapse images are separated by 1 minute intervals, it is reasonable to assume that, once a new RNA appears, the full occupation of its MS2-GFP binding sites will take less time than the time between two consecutive images. This is agreement with measurements in [21].

Growth Conditions, Microscopy, Data Extraction on Transcription Activation Times

Cells were grown overnight at 30 °C with aeration and shaking in lysogeny broth (LB) medium, supplemented with the appropriate antibiotics (35 μg/ml Kanamycin and 34 μg/ml Chloramphenicol). From the overnight cultures, cells were diluted into fresh LB medium, supplemented with antibiotics, to an optical density of $OD_{600} \approx 0.05$, and allowed to grow at 37 °C, 250 rpm, until reaching an $OD_{600} \approx 0.3$. Next, 100 $ng.ml^{-1}$ anhydrotetracycline (aTc) was added to induce $P_{LtetO-1}$ and produce MS2d-GFP, and 0.1% L-Arabinose to pre-activate the target gene, controlled by $P_{Lac-ara-1}$ [17,20]. Afterwards, cells were centrifuged (8000 rpm, for 1 minute), and re-suspended in the remaining LB medium. From this, a few microliters of cells were taken and placed between a 3% agarose gel pad and a glass coverslip, before assembling the FCS2 imaging chamber (Bioptechs, see Figure S1). Finally, the chamber was heated to the desired temperature (24 °C, 30 °C, 37 °C and 41 °C) and placed under the microscope.

We observed that, in the absence of IPTG, the cells produce the same (spurious) amount of RNA, with or without Arabinose (data not shown), in agreement with previous studies [20]. However, pre-induction by Arabinose much prior to induction by IPTG, enhances slightly the RNA production rate [16,18]. As such, we pre-induced cells with Arabinose [17,20] 45 minutes prior to introducing IPTG in the media. As such, we pre-induced cells with Arabinose [17]. This implies that, by the time IPTG is added, the cells already contain a constant amount of Arabinose. This is ensured by the presence of Arabinose in the original media and by the constant replenishment of this media during microscopy measurements (Methods and Figure S1). Thus, we do not expect any potential feedback mechanism associated to the Arabinose intake process to influence the transcription activation times measured here, following the introduction of IPTG in the media.

Cells were visualized by a 488 nm argon ion laser (Melles-Griot), and an emission filter (HQ514/30, Nikon) using a Nikon Eclipse (Ti-E, Nikon) inverted microscope with a 100x Apo TIRF (1.49 NA, oil) objective. Fluorescence images were acquired by C2+ (Nikon), a point scanning confocal microscope system, and Highly Inclined and Laminated Optical sheet (HILO) microscopy, using an EMCCD camera (iXon3 897, Andor Technology). The laser shutter was open only during exposure time to minimize photobleaching. All images were acquired with NIS-Elements software (Nikon). While imaging, cells were supplied with a constant flow of fresh LB medium (pre-warmed to the

same temperature as in the chamber), containing 1 mM of IPTG, 0.1% of L-Arabinose, and 100 ng.ml^{-1} of aTc, using a peristaltic pump (Bioptechs), at a rate of 0.1 mL min^{-1}. Images were taken once per minute for 2.5 hours. At each moment, we imaged 6 specific locations, to attain information on multiple lineages.

After performing a semi-automated cell segmentation and lineage construction [22], the moment of production of the first RNA by each cell lineage was obtained by selecting cells absent of RNA spots at the start of the imaging period (i.e., without leaky expression), and then detecting by visual inspection (from fluorescence images) when the first production occurs in each branch of each lineage (Figure 2B), after introducing the inducers.

Aside from visual inspection, fluorescent RNA spots and their intensities were also detected from the confocal images using the Gaussian surface-fitting algorithm proposed in [23] specifically for the purpose of detecting and quantifying MS2-GFP tagged RNAs. We found no significant difference between using this automatic algorithm and the visual inspection of the moment when the first RNA appears.

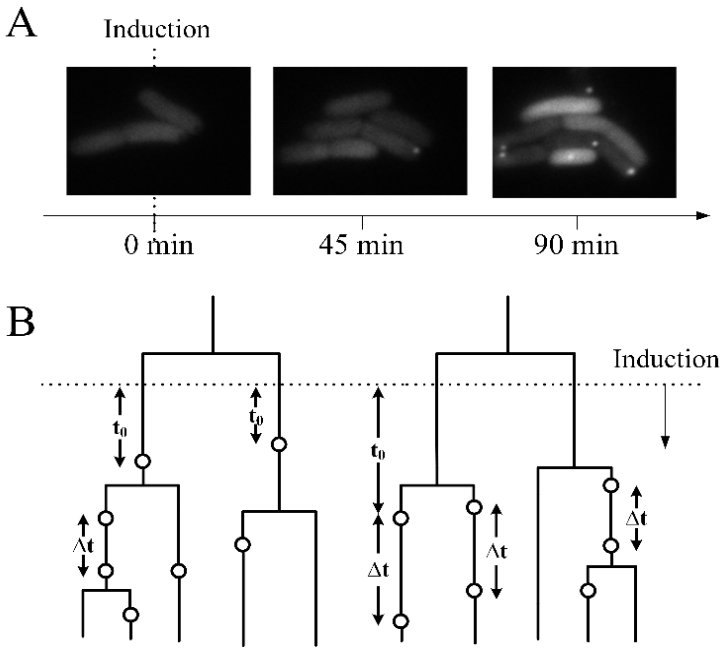

Figure 2. Data collection: (A) Cells are placed under the microscope at t=0 min and continuously supplemented with fresh medium. The reporter system (MS2d-GFP) is induced in liquid culture at t = -45 min. At t = 0 min, with the cells already having sufficient MS2d-GFP proteins for accurate RNA detection, transcription of the target RNA for MS2d-GFP is induced. (B) Illustration of RNA production events (circles) in cell lineages. Shown are the time for the first RNAs to appear (t_0) and the subsequent time intervals between consecutive RNA production events (Δt) in single cells. A dotted line indicates when the inducer of the target promoter is introduced.

As a side note, we found the rate of leaky expression to be very weak (less than 1 spot per ~20 cells prior to induction).

Finally, we note that the data on time intervals between consecutive RNA productions in individual cells used here was entirely obtained from [15]. There, time lapse microscopy was conducted on cells of the same strain, with the same constructs, and under the same induction and growth conditions as the ones used here.

Quantitative PCR for mean RNA quantification

Quantitative PCR (qPCR) was used to attain the induction curve of $P_{Lac\text{-}ara\text{-}1}$ as a function of IPTG concentration at 37 °C (for details, see Supplementary Material). This induction curve is shown in Figure S2. Visibly, for 0.5 mM IPTG and above, $P_{Lac\text{-}ara\text{-}1}$ is fully induced.

Estimation of intake times by deconvolution from empirical data on activation times and active transcription interval duration

The empirical method of MS2-GFP tagging of RNA allows for new RNAs containing multiple MS2-GFP binding sites to be detected shortly after they are produced [20]. From this data, one can directly extract the time intervals between consecutive RNA productions in individual cells following induction, as well as the time for the first RNA to be produced once inducers are added in the media. However, one cannot directly measure the time that inducers take to enter the cells and activate the target promoter. To obtain this information, we next propose a methodology based on deconvolution techniques for extracting this information from the data.

Given the model above, the mean time for the first RNA to appear in a cell following the addition of inducers in the media (here named t_0) depends on the time for inducers to enter the cell (reactions (1-2) in Supplementary Material) [4,5], here denoted as t_{int}. Also, it depends on the time for RNA production by an active promoter (which depends on the rate-limiting steps in transcription) [24,25], determined by reactions (3-6) in Supplementary Material, and here represented by Δt since, under full induction, this time should equal the time between consecutive RNA productions in active promoters [5]. In particular, we have:

$$t_0 = t_{int} + \Delta t \tag{1}$$

As the inducer intake and the production of the first RNA are independent, consecutive processes, one can use deconvolution to obtain a distribution of values of t_{int} (and, thus, mean and variance) from the data. Namely, for each temperature, one can deconvolve the probability density function (PDF) of the Δt distribution from the PDF of the t_0 distribution, provided that these two distributions are known [26].

For this, we estimate the PDFs of Δt and t_0 distributions as their best-fitted gamma distributions to the respective empirical distributions. We choose the gamma distribution as a model, since such distributions allow the mean and the variance to change independently, thus facilitating the fitting to the empirical distribution [14].

First, we use the gamma fits to the empirical Δt distributions reported in a previous work [14]. This fit used censored intervals between productions of consecutive transcripts extracted from live-cell measurements. The censoring accounts for the effects of finite sampling rate (60 s sampling interval), and thus improves the accuracy of the parameter estimation [27]. It also accounts for right-censored intervals, to compensate for the truncation of the right tail of the Δt distribution due to the finite cell division times. This fitting follows the maximum likelihood criteria [14].

Afterwards, to the measured t_0 distributions, we apply the same censored fitting procedure, but without right-censoring (as t_0 durations are not restricted by cell lifetime). Finally, we obtain the PDF of the t_{int} distribution using the Fast Fourier Transform (FFT) deconvolution method, as proposed by Sheu and Ratcliff [26], except that we do not apply frequency filtering, since our estimated t_0 and Δt PDFs do not contain high-frequency noise. As outlined by Sheu and Ratcliff [26], the result of the deconvolution may contain negative values, even though the PDF, by definition, cannot have values below zero. Those negative values should be interpreted as resulting from the uncertainty in the best-fit gamma distributions to t_0 and Δt empirical data, which, in turn, originates from uncertainty in the t_0 and Δt measurements. However, even if the selected models do not precisely depict the PDFs of the corresponding processes, the results of the deconvolution are still interpretable, even though the uncertainty in the deconvolution product is undefined [26]. Here, to allow such interpretation, we set the negative values of the t_{int} PDF to zero.

To estimate the uncertainty of our findings, we constructed bootstrap 95% confidence intervals (CIs) for mean and noise of the t_{int} distribution using non-parametric resampling of t_0 and Δt empirical data [28,29]. For this, for each temperature condition, we perform 2000 random resamples with replacement of the t_0 and Δt empirical distributions (using an original amount of samples), and obtain the t_{int} PDF for each resampled pair of t_0 and Δt distributions, which then allows obtaining the bootstrap distributions of the mean and CV^2 (squared coefficient of variation) of the t_{int} PDFs. We take 0.05 and 0.95 percentiles of those distributions as the 95% CIs of the estimated mean and CV^2 of the t_{int} distribution.

Estimation of intake times using Lineweaver-Burk equation

Aside from the method above, we make use of the Lineweaver-Burk equation [30] to estimate mean intake times. For this, from (1) and the model of gene expression (reactions 1-6 in Supplementary Material), note that as the amount of inducers in the media is increased, in a first stage, the inducers inside the cell will increase in number. As such, during this stage, both t_{int} and Δt will decrease with increasing inducer concentration. However, beyond a certain concentration of inducers in the media, further increases in this concentration will no longer lead to increases in the rate of RNA production (i.e. when the regime of full induction is reached), due to the rate-limiting steps in transcription and the

finite number of RNA polymerases inside the cell (reaction 6 in Supplementary Material). This well-known fact is also demonstrated here by Figure S2, which shows that, beyond a certain inducer concentration (both in the microscopy measurements and the qPCR measurements) the rate of RNA production no longer increases with further increases in the IPTG concentration in the media.

Meanwhile, the time taken by the cell to intake inducers should continue to decrease with increases in inducer concentration in the media, even in the regime of full induction of transcription. Namely, in theory, for an infinite amount of inducers in the media, t_{int} should equal zero. In this regime, following the introduction of infinite number of inducers in the media, the total mean time taken to produce the first RNA will be equal to the duration of subsequent intervals between consecutive RNA productions, i.e.:

$$t_0\left(\left[\mathrm{IPTG}\right]=\infty\right)=\Delta t \tag{2}$$

Thus, provided that the decrease in t_{int} with the decrease of the inverse of the inducer concentration is linear (as assumed in our model reactions (1) and (2) in Supplementary Material), we can derive t_{int} in the 'control condition' using the Lineweaver-Burk equation [30] as follows:

$$t_{int}=\frac{[\mathrm{IPTG}]_2(t_{0_2}-t_{0_1})}{([\mathrm{IPTG}]_1-[\mathrm{IPTG}]_2)} \tag{3}$$

In equation (3), t_{0_1} and $[\mathrm{IPTG}]_1$ are, respectively, the mean t_0 and the inducer concentration in the control condition. Meanwhile, t_{0_2} and $[\mathrm{IPTG}]_2$ are the corresponding values in a condition where the inducer concentration differs from the control, and is above the minimum concentration to achieve maximum RNA production rate.

Also, one can calculate 95% CIs for the obtained mean t_{int} value based on the method of propagation of errors [31].

As a side note, this methodology is similar to the usage of τ plots, from which, by fitting a line to the results of measurements of the transcription rate for increasing RNA polymerase concentrations one can extract the duration of the events following the initiation of the open complex formation [16,24,32].

Inference of the number and duration of the sequential steps in the intake process by fitting with a sum of exponential steps

Our model of intake (reactions (1) and (2) in Supplementary Material) assumes 2 steps, each with a duration following an exponential distribution, in agreement with measurements at optimal temperatures [5,6]. However, as noted, our modelling strategy allows considering the possibility that, at different temperatures, additional or less steps may be rate-limiting.

To determine the number of steps, one can perform fittings of d-steps models (each step following an exponential distribution) for increasing number of steps, until adding a step no longer improves the fitting. In such a model, as more steps are added and if the overall mean duration of the d-steps process is kept constant, the variance of the durations between events will decrease. The closer the d-exponential steps distribution is to a gamma distribution with a shape parameter set to d, the smaller will be its variance.

The d-exponential step model was chosen due to how we model transcription, namely, as a set of consecutive of chemical reactions, each of which having a distribution of intervals between consecutive occurrences that is expected to follow an exponential distribution. Also, there is significant accumulated evidence that, in *E. coli*, this model fits very well, in a statistical sense, the empirical distributions of many promoters [5,6,16,33–35].

Here we perform this fitting to a d-steps model for each temperature condition. For this, by deconvolution of the empirical data, we obtain a distribution of the duration of the intake process. From it, we determine the maximum likelihood fit of a model with *d* statistically independent steps, whose time lengths each follow an exponential distribution, with possibly different rates.

The likelihoods are compared using the likelihood ratio test, and the model with smallest *d* that cannot be rejected at the significance level 0.01 is selected in favor of a higher order model.

Note that this method does not allow determining the order of the steps, only their number and durations. Note also that, while changing temperature may not alter the number of rate-limiting steps, it may instead (or also) cause them to no longer be well modeled by elementary reactions as our model assumes. In that case, we expect the fitting to d exponential steps to require a higher number of steps than if the steps were elementary.

Results and Conclusions

$P_{Lac\text{-}ara\text{-}1}$ transcription activation kinetics is temperature dependent

We first studied, at the single cell level, the temperature dependence of the kinetics of transcription activation of $P_{Lac\text{-}ara\text{-}1}$ by IPTG. All empirical data were obtained from observing individual cells over time, using MS2d-GFP tagging of the target RNA, fluorescence microscopy, and image analysis techniques (Methods).

For this, we placed *E. coli* cells (DH5α-PRO) with a single-copy plasmid coding for the RNA target for MS2d-GFP under the control of $P_{Lac\text{-}ara\text{-}1}$, and fully activated its expression by adding IPTG (1 mM) to the media (Figure S2) while already under microscope observation (Figure 2). The MS2d-GFP reporters, expressed by a multi-copy plasmid, were induced prior to this, so that cells were flooded with MS2d-GFP by the time $P_{Lac\text{-}ara\text{-}1}$ was induced (Methods).

From the time series obtained (~2.5 hour long, with images taken every minute), for each temperature, we extracted t_0, the time taken by individual cells to produce the first RNA, following the addition of inducers in the media (Methods). Note that only one such event per lineage is considered and that cells already with one or more RNAs at the start of the observation period were discarded.

From these data, we calculated the mean, standard error, and CV^2 of t_0 values. Finally, we performed Kolmogorov-Smirnov (KS) tests to compare each distribution of t_0 values with the distribution at 37 °C (named 'control' condition). Results are shown in Table 1.

Table 1. Measurements of t_0 vs temperature. Shown are the number of measurements (N_{t_0}), mean (μ_{t_0}) standard error (SE) and CV^2 of the distribution of t_0 values ($CV^2_{t_0}$). The table also shows the *p*-value from the KS tests comparing the t_0 distributions at each temperature, with the distribution at 37 ˚C (control). For *p*-values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution can be rejected.

T (˚C)	N_{t_0}	$\mu_{t_0} \pm$ SE (s)	$CV^2_{t_0}$	KS-test for t_0 values vs 37 ˚C (*p*-value)
24	93	2743 ± 102	0.13	< 0.01
30	162	3020 ± 119	0.25	< 0.01
37	60	2109 ± 215	0.63	-
41	93	2379 ± 144	0.34	0.19

From the data in Table 1, we find that for temperatures lower than 37 °C, the activation time t_0 differs significantly from the control (in a statistically sense), with its mean (μ_{t_0}) being higher and its $CV^2_{t_0}$ (surprisingly) being lower for lower temperatures.

Qualitatively similar results were obtained (Table S1) using the *E. coli* JW0334 strain (see section 'Bacterial strains and plasmids').

Cell-to-cell variability of t_{int} decreases with decreasing temperature

Next, we investigate how the time for inducers to enter the cell, t_{int}, changes with temperature. For this, besides the data above, we make use of the data from [14], which consists of empirical distributions of intervals between consecutive RNA productions by active promoters in individual cells (Δt), under the same temperature conditions as above. These data therefore informs on the kinetics of active transcription (i.e. is not affected by intake times).

As mentioned in Methods, in accordance to our model (reactions 1-6 in Supplementary Material) and equation 1, the time for the production of the first RNA in each cell, following the introduction of inducers in the media (t_0), should consist of the time for the intake of the inducer by the cell (t_{int}) and the time taken by the active promoter to produce the first RNA (Δt). As these processes are consecutive and independent, it should

be possible to obtain the time-length for intake of the inducers (t_{int}) by deconvolving Δt from t_0.

For this, we performed model fitting with censoring to the data from live-cell measurements of t_0 (Table 1) and used the model fitting of empirical Δt values from [14]. In Figure 3, we show the empirical distribution and the best gamma fits of t_0.

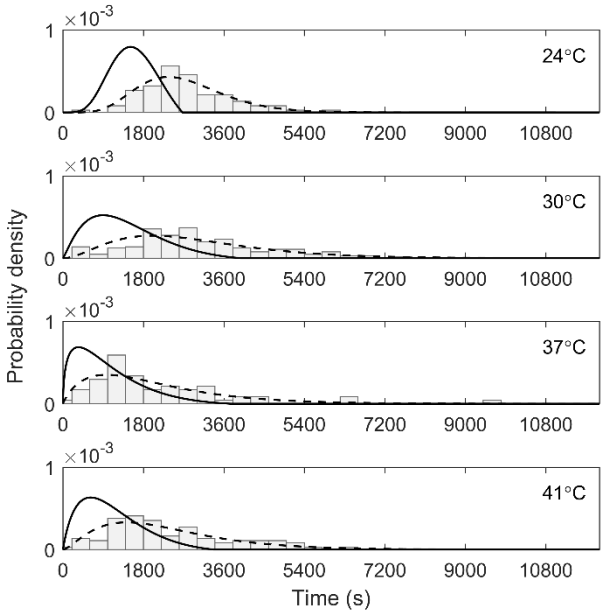

Figure 3. Empirical distribution of t_0 (histogram), along with the best gamma fit to t_0 (dashed line) and the deconvolved t_{int} (solid line), as function of temperature.

Next, we obtained the t_{int} distribution for each temperature condition from the deconvolution of Δt from t_0 (Methods). Results for the mean and CV^2 values of the distributions of t_{int} obtained from this deconvolution are shown in Table 2, along with the 95% CI. It is noted that the values at 37 °C are in agreement with previously reported measurements [5,6].

Meanwhile, the deconvolved distributions are shown in Figure 3. From these, we find a clear change in the shape of the t_{int} distribution as temperature is lowered.

Table 2. Mean and CV^2 of the deconvolved t_{int}, along with the 95% CI for each temperature condition.

T (˚C)	$\mu_{\widehat{t_{int}}}$ (s)	95% CI of $\mu_{\widehat{t_{int}}}$ (s)	$CV^2_{\widehat{t_{int}}}$	95% CI of $CV^2_{\widehat{t_{int}}}$
24	1548	[1316, 1799]	0.10	[0.06, 0.18]
30	1369	[1113, 1671]	0.32	[0.20, 0.48]
37	986	[726, 1329]	0.52	[0.28, 0.95]
41	1083	[807, 1402]	0.37	[0.23, 0.63]

From Table 2, we find that the mean duration of the intake process, $\mu_{\widehat{t_{int}}}$, is the lowest while the variability, $CV^2_{\widehat{t_{int}}}$, is the highest at 37 °C. Meanwhile, at the lowest temperature tested (24 °C) the opposite occurs ($\mu_{\widehat{t_{int}}}$ is the highest and $CV^2_{\widehat{t_{int}}}$ is the lowest).

Also, from the values of t_0 (Table 1) and t_{int} (Table 2), we find that the dynamics of intake plays a major role in the dynamics of transcription activation in all temperature conditions, both regarding the mean duration of activation and its cell-to-cell variability. Thus, it is not a surprise that t_{int} behaves similarly to t_0 with changes in temperature.

Finally, note that the fact that noise is reduced with decreasing temperature suggests that the process becomes more sub-Poissonian, which could occur, e.g., if the number of the rate-limiting steps in the intake process increases with decreasing temperature.

As a side note, we also conducted similar experiments in the absence of IPTG, so as to estimate the level of toxicity due to induction by 1 mM IPTG. We found no difference in cell growth rate between the two conditions, and thus conclude that the levels of toxicity are not significant.

Validation of the inferred mean t_{int} using the Lineweaver-Burk equation

It is possible to empirically validate the mean value of the deconvolved t_{int} using the Lineweaver-Burk equation (Methods). For this, from individual cells at 24 °C, 37 °C and 41 °C, we measured the time between the moment of induction and the moment when the first RNA is produced for IPTG concentrations of 1 mM and of 0.5 mM. Note that both of these concentrations suffice to reach maximum induction in cells under the microscope (as shown in Figure S2). Because of this, Δt does not differ between the two conditions, and only affects t_{int}. From the measurements of t_0 in these two induction levels at a given temperature, using the Lineweaver-Burk equation, one can extrapolate the value of t_0 for infinite inducer concentration, which allows estimating the mean intake time at that temperature (Table 3).

Table 3. Mean t_{int} ($\mu_{t_{int}}$) obtained from the Lineweaver-Burk equation and 95% CI of $\mu_{t_{int}}$ for various temperatures.

T (˚C)	$\mu_{t_{int}}$ (s)	95% CI of $\mu_{t_{int}}$ (s)
24	2434	[1949, 2918]
37	1322	[842, 1801]
41	1459	[1113, 1804]

From Table 3, we find that, in accordance with the results of deconvolution (Table 2), the mean t_{int} is highest at 24 °C, and is similar at 37 °C and 41 °C, being slightly smaller at 37 °C.

Quantitatively, we find that these values are ~ 35% larger (for 37 °C and 41 °C) and ~50% for 24 °C than those in Table 2. This is expected, as the deconvolution method is known to underestimate the peak value of the PDF [26].

Finally, we note that the value at 37 °C is also in clear agreement with a previous estimation of intake times at this temperature [6].

Number and duration of the rate-limiting steps of the intake process differs with temperature

To investigate the hypothesis that temperature affects the number and duration of the rate-limiting steps of the intake process, next, from the deconvolved t_{int} distributions of each temperature condition, we estimated the number and duration of these steps in maximum likelihood sense (Methods).

For this, we generalize the model of intake depicted by reactions (1) and (2) in Supplementary Material to a *d*-steps model, each exponentially distributed in duration, so that the number and duration of the rate-limiting steps are allowed to differ between the temperature conditions.

Results of this estimation are shown in Table 4, where we present the number and duration of the steps of the best fit model, along with the log-likelihood values. Meanwhile, in Table S2, we show the results for each condition when assuming specifically 1, 2, 3, and 4 steps, along with the *p*-values of the tests comparing pairs of models that are used to select the best model. Finally, in Figure 4 we show the best fit to the deconvolved t_{int} for each condition.

Table 4. Rate-limiting steps in the intake process determined by maximum likelihood estimation. Shown are the number of steps, the log-likelihood, the durations of the steps of the inferred models for each condition, and the CV^2 of the best fit. We fit the models to 10^5 random samples from the deconvolved t_{int} distribution. Note that there is no implied temporal order of the steps.

T (˚C)	No. Steps	Steps Durations	Log-likelihood	$CV^2_{\widehat{t_{int}}}$
24	≥4	(387, 387, 387, 386)	-774461	0.25
30	3	(457, 457, 457)	-801201	0.33
37	2	(667, 319)	-783576	0.56
41	3	(532, 532, 20)	-783350	0.48

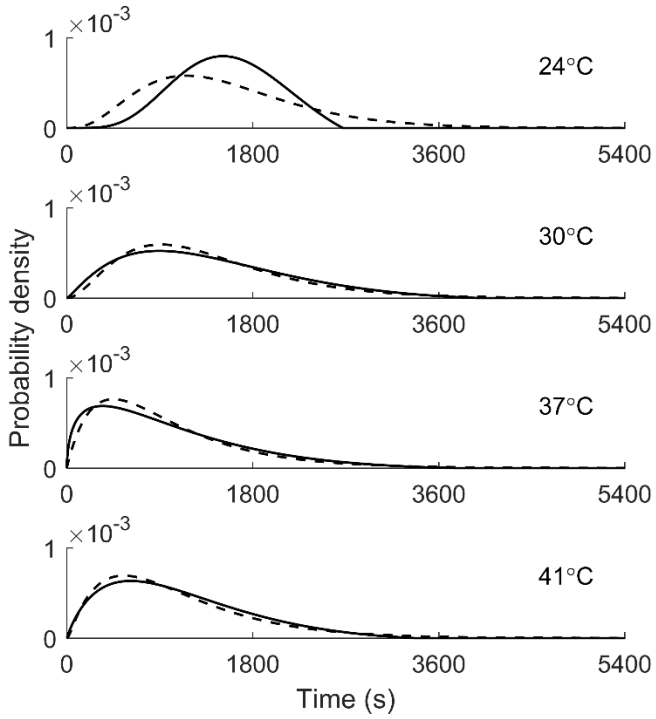

Figure 4. Deconvolved t_{int} distributions (solid line) and their best-fit d-steps model (dashed line).

Importantly, this result is in agreement with previous studies using data from cells at 37 °C [5].

From Tables 4 and S1, for all conditions, the test rejects the 1-step model in favor of a higher order model. This is expected, given the existence of the two membranes in the cell walls of *E. coli* cells and the time that inducers are expected to take to cross the periplasm in between [6].

Also, interestingly, the 2-steps model is the preferred one for cells at 37 °C and 41 °C (the step with a 20 s duration for the 41 °C condition can be disregarded, as the microscopy images are separated by 60 s intervals).

Meanwhile, at lower temperatures, higher order models (3 or more steps) are preferred, indicating that other steps become rate-limiting (in agreement with the deconvolution results), and/or that the steps duration may no longer follow an exponential distribution.

In this regard, we interpret the fact that a 4-steps model did not suffice to model the 24 °C condition (see Figure 4) as evidence for a significant change in the kinetics of intake with temperature, which renders the multi-step, exponentially distributed model incapable of fully capturing the dynamics. We hypothesize that this may be the consequence of increased viscosity of the cytoplasm and periplasm [14], along with changes in the physical properties and functionality of the intake 'machinery' in the cell walls.

Note that the CV^2 values of the best fits for 30 °C, 37 °C and 41 °C match the estimated values of the corresponding t_{int} distributions deconvolved from the fits to the empirical data. While the best fit in 24 °C condition has higher CV^2 than the deconvolved t_{int} (which is expected from the fact that the 4-steps model did not suffice to model the 24 °C condition), the trend in CV^2 of the deconvolved distributions and of their best fits is the same.

Finally, note that, in several cases, the time scales of the steps are identical. This may be due to an unknown artefact of the inference method or be representative of the real kinetics of intake of this inducer.

Discussion

In this work, we studied the single-cell dynamics of intake of IPTG, an inducer of the promoter $P_{Lac\text{-}ara\text{-}1}$, as a function of temperature. Rather than focusing on biological cellular adaptations, we focused solely on rapid physical changes due to temperature shifts in the process of inducer intake and consequent transcription kinetics.

For this, we first measured *in vivo* the time taken by individual cells to produce the first RNA, following the start of induction. From this, and previously collected data on the dynamics of RNA production by $P_{Lac\text{-}ara\text{-}1}$ [14], we applied two novel, independent methods to obtain the single-cell intake kinetics of the inducers, for each temperature condition. These methods' results were consistent with one another.

From this, first, we established that the response of the distribution of intake times of individual cells to temperature changes remains similar to that of the distribution of transcription activation times as temperature is changed, much due to the fact that most of the activation time is spent in the intake process in all conditions. Interestingly, the mean value of these distributions increases while their variability decreases for decreasing temperatures.

Since the intake process is bound to consist of multiple consecutive steps (in the case of IPTG, it was previously shown to be well modeled by a 2-step process for cells at 37 °C [5,6], we hypothesize that the decrease in variability could be the result of the emergence of additional rate-limiting steps in this process with decreasing temperature. The results of the maximum likelihood estimation tests support this view.

Further, they suggest that, at the lowest temperature condition tested here, the process is, from a dynamical point of view, 'too complex' to be well fitted by a sum of a small number (less than 5) of exponential steps. We hypothesize that this clear evidence that the duration of one of the steps of the intake process becomes non-exponential-like at low temperatures. There are several potential causes for this (and perhaps multiple causes), and they are likely not accounted by our model (else, the increase in number of exponential steps would have allowed to fit the data well). We expect these potentials causes to range from malfunctioning of the porins in the membrane responsible for the diffusive intake of

the inducers, increased viscosity of the cytoplasm and periplasm, alteration of the physical properties of the outer and inner membranes, etc.

It is worth noting that the application of the Lineweaver-Burk equation to extract the mean value of the intake times is a methodology that has not been previously used, but we expect it to be of use in future works as well. It requires measuring transcription activation times for various inducer concentrations (at least 2) above the minimum concentration required for maximum induction. It is limited by the fact that the speed of intake is assumed to change linearly with the inverse of the inducer concentration, which may not always be the case. However, we expect this to be the case within certain ranges of inducer concentrations for simpler intake (mostly diffusion-based), mechanisms. Thus, it should be applicable to the study of a wide range of cellular intake mechanisms.

Overall, we conclude that different environmental conditions cause significant changes in the single-cell distributions of intake times of transcription inducers, which is expected to have a significant effect on the degree of heterogeneity in cell populations and cell lineages, due to the longevity of the transients during which this phenomenon has a strong effect in RNA numbers.

In the future, one important aspect that requires further research is the cause for the reduced cell-to-cell diversity in response times with decreasing temperatures, which we believe to be due to the emergence of rate-limiting steps in the intake process. Which steps and how they emerge are open questions, whose answers will help better understanding the robustness of the intake systems of *E. coli*.

Acknowledgements

Work supported by Academy of Finland (295027 and 305342 to ASR), Jane & Aatos Erkko Foundation (610536 to ASR), Tampere University of Technology President's Graduate Program (SS), 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 (CP). The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

References

[1] Arkin A, Ross J and McAdams H 1998 Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells *Genetics* **149** 1633–48

[2] van Kampen NG, Reinhardt WP 1983 Stochastic Processes in Physics and Chemistry Physics Today **36(2)** 78–80

[3] Elowitz B, Siggia D, Levine AJ, Swain PS, Siggia E, and Swain P 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6

[4] Megerle J, Fritz G, Gerland U, Jung K, and Rädler J 2008 Timing and dynamics of single cell gene expression in the arabinose utilization system *Biophys J* **95** 2103–15

[5] Mäkelä J, Kandavalli V, and Ribeiro AS 2017 Rate-limiting steps in transcription dictate sensitivity to variability in cellular components *Sci Rep* **7** 10588

[6] Tran H, Oliveira SMD, Goncalves N, and Ribeiro AS 2015 Kinetics of the cellular intake of a gene expression inducer at high concentrations *Mol Biosyst* **11** 2579–87

[7] Hensel Z, Feng H, Han B, Hatem C, Wang J, and Xiao J 2012 Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis *Nat Struct Mol Biol* **19** 797–802

[8] Rosenfeld N, Young J, Alon U, Swain P, and Elowitz M 2005 Gene regulation at the single-cell level *Science* **307** 1962–5

[9] Robert L, Paul G, Chen Y, Taddei F, Baigl D, and Lindner AB 2010 Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch *Mol Syst Biol* **6** 357

[10] Kiviet D, Nghe P, Walker N, Boulineau S, Sunderlikova V, and Tans SJ 2014 Stochasticity of metabolism and growth at the single-cell level *Nature* **514** 376–9

[11] Yun HS, Hong J, and Lim HC 1996 Regulation of Ribosome Synthesis in *Escherichia coli*: Effects of Temperature and Dilution Rate Changes *Biotechnol Bioeng* **52** 615–24

[12] Gupta A, Lloyd-Price J, Neeli-Venkata R, Oliveira SMD, and Ribeiro AS 2014 *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli Biophysical Journal* **106(9)** 1928–37

[13] Choi PJ, Cai L, Frieda K, and Xie XS 2008 A stochastic single-molecule event triggers phenotype switching of a bacterial cell *Science* **322** 442–6

[14] Oliveira SMD, Häkkinen A, Lloyd-Price J, Tran H, Kandavalli V, and Ribeiro AS 2016 Temperature-Dependent Model of Multi-step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements *PLoS Comput Biol* **12** e1005174

[15] Oliveira SMD, Neeli-Venkata R, Goncalves N, Santinha JA, Martins L, Tran H, Mäkelä J, Gupta A, Barandas M, Häkkinen A, Lloyd-Price J, Fonseca JM and Ribeiro AS 2016 Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli Mol Microbiol* **99** 686–99

[16] Lloyd-Price J, Startceva S, Chandraseelan JG, Kandavalli V, Goncalves N, Häkkinen A, and Ribeiro AS 2016 Dissecting the stochastic transcription initiation process in live *Escherichia coli DNA Res* **23** 203–14

[17] Lutz R, and Bujard H 1997 Independent and tight regulation of trasncriptional units in *Escherichia coli* via the LacR/O, the TetR/O and Arac/I1-I2 regulatory elements *Nucleic Acids Res* **25** 1203–10

[18] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, and Mori H 2006 Construction of *Escherichia coli* K-12 in-frame, single-

gene knockout mutants: the Keio collection *Mol Syst Biol* **2** 1234–44

[19] Marbach A, and Bettenbrock K 2012 Lac operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA *J Biotechnol* **157** 82–8

[20] Golding I, Paulsson J, Zawilski S, and Cox EC 2005 Real-time kinetics of gene activity in individual bacteria *Cell* **123** 1025–36

[21] Golding I, and Cox EC 2004 RNA dynamics in live *Escherichia coli* cells *Proc Natl Acad Sci U S A* **101** 11310–5

[22] Häkkinen A, Muthukrishnan A, Mora A, Fonseca JM, and Ribeiro AS 2013 CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli* *Bioinformatics* **29** 1708–9

[23] Häkkinen A, and Ribeiro AS 2015 Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data *Bioinformatics* **31** 69–75

[24] McClure WR 1985 Mechanism and control of transcription initiation in prokaryotes *Annu Rev Biochem* **54** 171–204

[25] 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

[26] Sheu CF, and Ratcliff R 1995 The application of fourier deconvoution to reaction time data: a cautionary note *Psychol Bull* **118** 285–99

[27] Häkkinen A and Ribeiro AS 2016 Characterizing rate limiting steps in transcription from RNA production times in live cells *Bioinformatics* **32** 1346–52

[28] DiCiccio T and Efron B 1996 Bootstrap Confidence Intervals *Stat Sci* **11** 189–228

[29] Carpenter J and Bithell J 2000 Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians *Stat Med* **19** 1141–64

[30] Lineweaver H and Burk D 1934 The Determination of Enzyme Dissociation Constants *J Am Chem Soc* **56** 658–66

[31] Press W, Teukolsky S, Vetterling W, and Flannery B 1992 *Numerical Recipes in C* (Cambrigde: Cambrigde University Press) 661–5

[32] Bertrand-Burggraf E, Lefèvre J, and Daune M 1984 A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322 *Nucleic Acids Res* **12** 1697–706

[33] Kandavalli VK, Tran H, and Ribeiro AS 2016 Effects of σ factor competition are promoter initiation kinetics dependent *BBA - Gene Regul Mech* **1859** 1281–8

[34] Häkkinen A and Ribeiro A S 2016 Characterizing rate limiting steps in transcription from RNA production times in live cells *Bioinformatics* **32** 1346–52

[35] Muthukrishnan A, Martikainen A, Neeli-Venkata R, and Ribeiro AS 2014 In vivo

transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells *PLoS One* **9** e109005