BBA - Gene Regulatory Mechanisms 1862 (2019) 119–128

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

journal homepage: www.elsevier.com/locate/bbagrm

Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression

Sofia Startceva[a], Vinodh K. Kandavalli[a], Ari Visa[b], Andre S. Ribeiro[a,*]

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

[b] Faculty of Computing and Electrical Engineering, Tampere University of Technology, Tampere 33101, Finland

ARTICLE INFO

Keywords:
Single-cell time-lapse microscopy
Transcription initiation
RNA and protein numbers
Asymmetry and tailedness
Threshold crossing

ABSTRACT

Genetic circuits change the *status quo* of cellular processes when their protein numbers cross thresholds. We investigate the regulation of RNA and protein threshold crossing propensities in *Escherichia coli*. From in vivo single RNA time-lapse microscopy data from multiple promoters, mutants, induction schemes and media, we study the asymmetry and tailedness (quantified by the skewness and kurtosis, respectively) of the distributions of time intervals between transcription events. We find that higher thresholds can be reached by increasing the skewness and kurtosis, which is shown to be achievable without affecting mean and coefficient of variation, by regulating the rate-limiting steps in transcription initiation. Also, they propagate to the skewness and kurtosis of the distributions of protein expression levels in cell populations. The results suggest that the asymmetry and tailedness of RNA and protein numbers in cell populations, by controlling the propensity for threshold crossing, and due to being sequence dependent and subject to regulation, may be key regulatory variables of decision-making processes in *E. coli*.

1. Introduction

The gene regulatory networks of bacteria, such as *Escherichia coli*, include network motifs [1,2]. Some of these are responsible for decision-making processes that assist cells in adapting to environmental changes [3,4]. Significant behavioural changes in these motifs usually occur when the numbers of one or more of the component proteins cross thresholds [3]. The underlying mechanisms that define the propensity for the protein numbers of a given gene to cross a specific threshold are not yet fully understood.

In *E. coli*, it is common for the protein numbers to follow the corresponding RNA numbers [5,6]. These are determined by the rates of RNA production and degradation. Interestingly, RNA degradation in *E. coli* appears to be largely independent from the RNA sequence, abundance and metabolic function [7–9], suggesting that little regulation occurs at this stage. Meanwhile, various regulatory mechanisms of transcription have been identified, which usually act at the stage of initiation, suggesting that control over the RNA numbers is exerted at this stage [10–12].

From the dynamics point of view, the regulation of transcription initiation kinetics occurs via the tuning of the time-length of the rate-limiting steps of initiation, respectively, the events prior and after committing to open complex formation [13–17]. In particular, recent studies [14,16–18] have shown that, under full induction, the in vivo kinetics of these rate-limiting steps, along with supercoiling buildups [19], define, to a great extent, the distribution of time intervals between consecutive RNA production events (here referred to as 'Δt distribution'). Further, it was shown that not only the first moment (mean), but also the second moment of this distribution (variance) can be tuned by the kinetics of these steps [16,18].

Given this, we hypothesise that, by tuning the kinetics of these rate-limiting steps, one can also tune the third and fourth moments of the Δt distribution (respectively, the skewness and kurtosis). Further, we hypothesise that these two moments can be tuned independently from the mean and coefficient of variation. To test these hypotheses, we perform in vivo time-lapse microscopy employing single-RNA detection by MS2-GFP tagging [20–22], from which we extract the Δt distributions for various promoters, media, induction schemes, growth phases, mutants and a stress condition. Next, for each condition, we estimate their mean, coefficient of variation, skewness and kurtosis. Subsequently, we estimate the kinetics of the rate-limiting steps in each condition and assess their influence on the skewness and kurtosis. Finally, to test whether changing the skewness and kurtosis of the Δt distribution has functional consequences, we measure the corresponding values of the

* Corresponding author.
E-mail address: andre.ribeiro@tut.fi (A.S. Ribeiro).

https://doi.org/10.1016/j.bbagrm.2018.12.005
Received 29 October 2018; Received in revised form 4 December 2018; Accepted 5 December 2018
Available online 14 December 2018

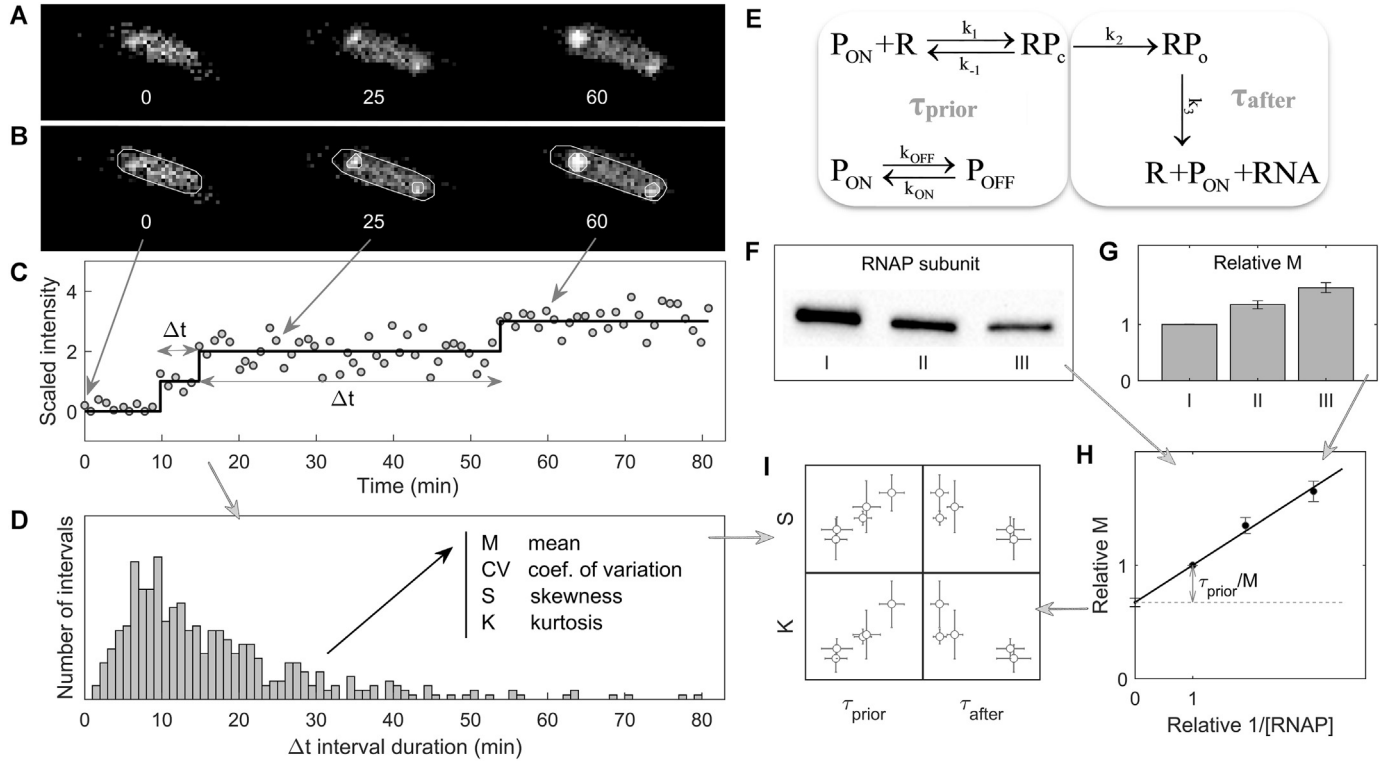

Fig. 1. Schematic representation of the steps for the analysis of the dynamics of RNA production in individual cells, from in vivo single-RNA, single-cell measurements. (A) Example confocal microscopy images over time of a cell expressing MS2-GFP and the target RNAs. (B) Segmentation of a cell and the MS2-GFP tagged RNA spots within (white lines). (C) Scaled RNA spots intensity over time (grey circles) of the example cell, along with the best-fitting monotonic piecewise-constant curve (black line) from which Δt intervals are estimated. (D) The distribution of time intervals between consecutive RNA production events in individual cells (Δt) from which mean (M), coefficient of variation (CV), skewness (S) and kurtosis (K) are extracted. (E) Model of transcription initiation. The first box contains the reactions occurring before commitment to open complex formation, with their mean time-length denoted as τ_{prior}. The second box contains the reactions occurring after commitment to open complex formation, with their mean time-length equals τ_{after}. For a detailed description of these reactions and parameters see Supplementary materials and methods, Section 1.6. (F) Western blot image of the RNA polymerase (RNAP) subunit in different media richness. (G) Relative inverse transcription rate of the target gene, measured by qPCR. (H) Relative τ plot (Lineweaver–Burk plot [25] of the inverse of the RNA production rate versus the inverse of the RNAP concentration, [RNAP]) for estimating τ_{prior} relative to M. (I) S and K versus τ_{prior} and τ_{after} in different conditions.

skewness and kurtosis of the distributions of single-cell protein expression levels.

2. Materials and methods

Fig. 1 informs on the models and methods used. In short, the main empirical data (Δt distributions) are obtained by measuring when each RNA appears in each cell. Also, we measure the average intracellular RNAP concentration. From these concentrations and the corresponding mean of the Δt distribution in each condition, we estimate the time spent in transcription initiation prior and after commitment to open complex formation (τ_{prior} and τ_{after}, respectively, with their sum equalling Δt) (model in Fig. 1E).

In summary, we first estimate τ_{prior}/M from τ plots [23]. For this, the inverse of the RNA production rate relative to the control (as measured by qPCR) is plotted against the inverse of the RNAP concentration relative to the control (as measured by Western blot, Supplementary materials and methods, Section 1.4). Next, a line is fitted to the data. The point where this line intersects the Y axis equals the extrapolated value of the inverse of the transcription rate for an 'infinite' RNAP concentration. As such it should equal τ_{after}/M, according to the model in Fig. 1E. From this and the value of M, one can calculate τ_{after} and τ_{prior} (Supplementary materials and methods, Section 1.5). Next, from the same Δt distributions, we extract the coefficient of variation, skewness and kurtosis in each condition.

Note that, although genes replicate during the cells lifetime by a process that is not absent of noise and many variables control when each specific gene is replicated [24], we assume that the rate constants controlling the kinetics of RNA production of our gene of interest (Fig. 1E), which is on a single-copy F-plasmid, do not change significantly during the lifetime of the cells. To validate this assumption we compared the distributions of time intervals (between consecutive RNA production events) that started and ended in the first half of the lifetime with intervals that started and ended in the second half (Supplementary results, Section 2.1). From the comparisons of these distributions in each condition (Table 1) we conclude that the assumption is sufficiently accurate.

2.1. Bacterial strains, plasmids, growth conditions, MS2-GFP tagging system, induction of the reporter and target genes, and measurement conditions

The *E. coli* strain used was DH5α-PRO (identical to DH5αZ1) [26] whose genotype is: deoR, endA1, gyrA96, hsdR17(rK− mK+), recA1, relA1, supE44, thi-1, Δ(lacZYA-argF)U169, Φ80δlacZΔM15, F-, λ-, PN25/tetR, PlacIq/lacI and SpR. This strain produces, from the chromosome and in abundance, the necessary regulatory proteins for their constructs, namely, LacI, AraC and TetR [26]. E.g. LacI, the main repressor of the control promoter ($P_{lac/ara-1}$), exists in a concentration much higher than the wild type (~3000 copies vs ~20 in wild type [26]). These characteristics allow tight regulation of both target and reporter genes, ensuring that the observed RNAs are due to active transcription and not the result of transcription leakiness (i.e. in the absence of activation). In particular, we measured leaky expression of

Table 1
Description of conditions. Shown are the name by which the condition is identified, the target plasmid and corresponding inducer, the reporter plasmid and corresponding inducer, and the media.

Conditions	Target promoter	Target inducers	Reporter promoter	Reporter inducer	Growth media
LA	$P_{lac/ara-1}$	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1×
LA(75)	$P_{lac/ara-1}$	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	0.75×
LA(50)	$P_{lac/ara-1}$	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	0.5×
LA(ara)	$P_{lac/ara-1}$	1% ara	$P_{LtetO-1}$	100 ng aTc	1×
LA(IPTG)	$P_{lac/ara-1}$	1 mM IPTG	$P_{LtetO-1}$	100 ng aTc	1×
LA(oxi)	$P_{lac/ara-1}$	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1× + 0.6 mM H_2O_2
Mut1	$P_{lac/ara-1}$ (Mut-1)	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1×
Mut2	$P_{lac/ara-1}$ (Mut-2)	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1×
Mut3	$P_{lac/ara-1}$ (Mut-3)	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1×
Mut4	$P_{lac/ara-1}$ (Mut-4)	1 mM IPTG + 1% ara	$P_{LtetO-1}$	100 ng aTc	1×
tetA	P_{tetA}	–	P_{lac}	1 mM IPTG	1×
tetA(st)	P_{tetA}	–	P_{lac}	1 mM IPTG	Stationary phase
BAD	P_{BAD}	0.1% ara	P_{lac}	1 mM IPTG	1×
BAD(st)	P_{BAD}	0.1% ara	P_{lac}	1 mM IPTG	Stationary phase

$P_{lac/ara-1}$, in the absence of IPTG and arabinose, and found only ~5% or less cells with an MS2-GFP tagged RNA, 2 h after inducing the reporter expressing MS2-GFP.

We also use BW25113, whose genotype is F-, DE(araD-araB)567, lacZ4787(del)::rrnB-3, LAM-, rph-1, DE(rhaD-rhaB)568, hsdR514, which expresses LacI and AraC from the genotype. The absence of TetR allows the Tet promoter to express constitutively.

All cells carry two plasmids: a multi-copy reporter plasmid coding for MS2-GFP under the control of an inducible promoter and a single-copy F-based target plasmid coding for the transcript with multiple MS2-GFP binding sites under the control of another promoter (Table 1). Also, in all target plasmids, we inserted a sequence coding for a red fluorescent protein, between the target promoter and MS2 binding sites. Promoter sequences are specified in Supplementary Fig. S1. Tagged RNAs can be visualized as fluorescent spots [14,20–23] (Fig. 1A).

In general, to observe RNAs tagged by MS2-GFP proteins, cells were grown overnight in LB media with the respective antibiotics at 30 °C in an orbital shaker with aeration of 250 rpm. From the overnight culture, cells were diluted using fresh LB media (unless stated otherwise in Table 1) to an initial OD_{600} of 0.05 (measured with a spectrophotometer, Ultrospec 10; GE Healthcare) and incubated at 37 °C at 250 rpm to allow growth until reaching an OD_{600} of 0.25. In general, the reporter gene was induced 1 h prior to the target gene, to allow for sufficient MS2-GFP proteins to be produced prior to the appearance of the target RNAs. For a detailed description, see Supplementary materials and methods, Section 1.1. Inducers of target and reporter genes are described in Table 1.

The MS2-GFP RNA tagging technique, proposed in [27], is at present the only direct method to measure time intervals between RNA production events in live, individual cells [14,16,21,22]. This is possible because, first, once appearing, each tagged RNA spot exhibits 'full' fluorescence (assuming 1 min interval between microscopy images) [22]. This removes uncertainty in the process of RNA counting as it reduces the possibility for 'partially fluorescent RNAs'. This uncertainty is further reduced in that, once tagged, the fluorescence of the spots remains near constant for longer than our measurement time (2 h or more) [22]. This provides significant reliability to the quantification of the time-length of intervals between consecutive RNA production events [21].

MS2-GFP tagging affects the spatial organization of the RNAs inside the cell [28]. However, this does not affect the precision of quantification of the intervals between consecutive RNA production events, which are based solely on the total intensity of the MS2-GFP tagged RNAs in a cell, not on their location.

To assess whether this technique has a negative impact on cell physiology, we compared cell growth rates and morphology with and without activating the expression of the MS2-GFP reporter. Supplementary results in Section 2.2 show that growth rates and cell morphology are not significantly affected by expression of MS2-GFP, in agreement with previous studies [14,23].

Finally, it is also reasonable to assume that MS2-GFP tagging could affect the protein expression levels of the target gene, due to partially interfering with the target RNA (albeit in a different region from the one coding for the red fluorescent protein). We tested this by comparing protein expression levels when and when not activating the expression on MS2-GFP (Supplementary results, Section 2.3). The results confirm that the expression levels of the red fluorescent protein are not perturbed significantly by MS2-GFP tagging (Fig. S9).

Meanwhile, to measure the single cell distributions of RNAP concentration, we used *E. coli* RL1314 strain with fluorescently tagged β' subunits (a kind gift from Robert Landick, University of Wisconsin-Madison) [29]. From the overnight culture, we diluted the cells to an OD_{600} of 0.1 in various media richness (Materials and methods) and allowed them to grow to an OD_{600} of 0.5 at 37 °C at 250 rpm. Cells were then pelleted by centrifugation and visualized under the microscope.

The plasmids (Table 1) construction and transformation were performed using standard molecular cloning techniques [30]. To construct $P_{lac/ara-1}$-mCherry-48 binding sites (bs) mutants, we used a plasmid carrying mCherry followed by a 48bs array in the pBELO vector backbone, originally constructed in [31]. To obtain the mutant promoters (Supplementary Fig. S1), we synthesized new promoter sequences of $P_{lac/ara-1}$ with specific point mutants with support from Gene Script, USA. Next, we inserted them into the pBELO vector backbone by Gibson Assembly [32], to obtain a single copy F-based plasmid carrying the target region $P_{lac/ara-1}$-mCherry-48bs mutants. This product was transferred into competent *E. coli* host cells. The recombinants were selected by antibiotic screening and confirmed with sequence analysis. It is noted that the mutant promoters were selected solely based on that their Δt distributions differed from the one of $P_{lac/ara-1}$.

2.2. Chemicals

The chemical components of LB media are Tryptone, Yeast extract and NaCl, purchased from LabM (Topley House, Bury, Lancashire, UK). The antibiotics used are Kanamycin 34 μg/ml, Ampicillin 50 μg/ml and Chloramphenicol 35 μg/ml, purchased from Sigma-Aldrich (St. Louis, MO). The inducers used are isopropyl β-D-1-thiogalactopyranoside (IPTG), anhydrotetracycline (aTc) and arabinose (ara), purchased from Sigma-Aldrich. Agarose (Sigma-Aldrich) was used for preparing the microscope gel pads. For PCR, Phusion high-fidelity polymerase and other PCR reagents were purchased from Finnzymes (Finland). Qiagen kits (USA) were used for plasmid isolation. For qPCR, cells were treated with RNA protect bacteria reagent (Qiagen, USA). iScript Reverse Transcription Supermix for cDNA synthesis and iQ SYBR green

supermix for qPCR were purchased from Biorad (USA).

2.3. Growth media

In all experiments, we used the LB media and its altered versions, first described in [14]. Namely, we used the following media compositions per 100 ml: 1 g tryptone, 0.5 g yeast extract and 1 g NaCl (pH 7.0), referred to as '1 ×' (Table 1); 0.75 g tryptone, 0.375 g yeast extract and 1 g NaCl (pH 7.0), referred to as '0.75 ×'; 0.5 g tryptone, 0.25 g yeast extract and 1 g NaCl (pH 7.0), referred to as '0.5 ×'; 0.25 g tryptone, 0.125 g yeast extract and 1 g NaCl (pH 7.0), referred to as '0.25 ×'. These four media are used to attain various mean intracellular RNA polymerase concentrations ([RNAP]) in cell populations, while not affecting normal cell physiology and morphology [14,16,23] (Supplementary Fig. S2A). Additionally, in two conditions, as in [23], we used the stationary phase media obtained by centrifuging the overnight culture of LB media at 10000 rpm for 10 min followed by filtration [23] (growth rates shown in Supplementary Fig. S2B).

2.4. qPCR measurements

Cells with target plasmids were harvested by centrifuging them at 8000 ×*g* for 5 min. To the pelleted cells, twice the amount of RNA protect reagent (Qiagen) was added, followed by the enzymatic lysis with Tris EDTA lysozyme buffer (pH 8.0). Total RNA was isolated using RNeasy kit (Qiagen) according to the kit instructions. The concentration of RNA was quantified using the Nanovue plus spectrophotometer (GE Healthcare). The RNA samples were treated with DNase to remove the residual DNA, followed by cDNA synthesis, using the iSCRIPT reverse transcription super mix. The cDNA samples were mixed with the qPCR master mix containing iQ SYBR Green Supermix (Biorad) with primers for the target and reference genes. The reaction was carried out in triplicates with the total reaction volume of 20 µl. For quantifying the target gene we used following primers: for mRFP1 (Forward: 5′ TACG ACGCCGAGGTCAAG 3′ and Reverse: 5′ TTGTGGGAGGTGATGTCCA 3′), for mCherry (Forward: 5′ CACCTACAAGGCCAAGAAGC 3′ Reverse: 5′ TGGTGTAGTCCTCGTTGTGG 3′). For the reference gene, 16S RNA primers (Forward: 5′ CGTCAGCTCGTGTTGTGAA 3′ and Reverse: 5′ GGACCGCTGGCAACAAAG 3′) were used. The qPCR experiments were performed by a MiniOpticon Real- time PCR system (Biorad). The following conditions were used during the reaction: 40 cycles of 95 °C for 10 s, 52 °C for 30 s and 72 °C for 30 s for each cDNA replicate. We used no-RT controls and no-template controls to crosscheck non-specific signals and contamination. PCR efficiencies of these reactions were > 95%. The data from CFX Manager TM Software was used to calculate the relative gene expression and its standard error [33].

2.5. Microscopy

Measurements of integer-valued numbers of RNAs or of the moments when a new RNA appears in individual cells were conducted using microscopy. For this, a few µl of cells carrying the induced reporter and target plasmids were placed between a coverslip and agarose gel pad (2.5%), with the respective inducers and antibiotics. Next, an FCS2 chamber (Bioptechs) was heated to 37 °C and placed under the microscope. Cells were visualized using a Nikon Eclipse (Ti-E, Nikon) inverted microscope, equipped with a 100× Apo TIRF (1.49 NA, oil) objective. Confocal images were obtained by a C2+ (Nikon) confocal laser-scanning system. For measuring GFP fluorescence (to visualize MS2-GFP 'spots' or RNAP-GFP), we used a 488 nm laser (Melles-Griot) and an emission filter (HQ514/30, Nikon). For time series, confocal images were taken every 1 min for 2 h. Previous studies [14] have shown that these microscopy settings do not cause significant phototoxicity in this strain. Finally, phase-contrast images were obtained simultaneously, with an external phase-contrast system and CCD camera (DS-Fi2, Nikon), every 5 min. Images were extracted using Nikon Nis-Elements software.

2.6. Image and data analysis

Microscopy images were analysed using the software 'CellAging' [34]. For details see Supplementary materials and methods, Section 1.2. From these analysed time-lapse images, we extracted intervals between consecutive RNA production events in individual cells, from which empirical distributions of these intervals (Δt distributions) were obtained (Fig. 1A–D). Data analysis was conducted using tailored algorithms implemented in MATLAB R2017b (MathWorks).

2.7. Flow cytometry

Measurements of protein expression levels were conducted using flow cytometry (FC). For this, cells from 5 ml of bacterial culture were diluted 1:10,000 into 1 ml PBS vortexed for 10 s. We performed measurements under various conditions. In each condition, a total of 50,000 cells were observed. Measurements were performed using an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA) with a yellow laser (561 nm) for excitation and the PE-Texas Red (mCherry) fluorescence detection channel (615/20 nm filter) for emission, at a flow rate of 14 µl/min and a core diameter of 7.7 µM. The PMT voltage of 584 was used for mCherry. To avoid background signal from particles smaller than bacteria, the detection threshold was set to 5000 in FSC-H analyses.

We applied unsupervised gating [35] (implemented in Python 3.6) to the flow cytometry data. We set the fraction of the cells whose data is used in the analysis (α) to 0.9, as it was sufficient to remove data points produced by debris, cell doublets and other undesired events. Reducing α further did not change the results qualitatively.

3. Results

3.1. Mean, coefficient of variation, skewness and kurtosis of the distributions of time intervals between consecutive RNA productions in individual cells differ with promoter sequence, regulatory factors and growth conditions

First, we obtained empirical data on the Δt distributions in 14 conditions (see Table 1 for details). These conditions were selected so as to test if the promoter sequence (conditions LA, Mut1, Mut2, Mut3, and Mut4, see Supplementary Fig. S1), regulatory factors such as RNAP and inducer concentrations (conditions LA, LA(75), LA(50), LA(ara), LA(IPTG)), and variables associated to the environment (e.g. media and stress) affect the skewness and kurtosis of the Δt distribution.

Results are shown in Supplementary Fig. S3. From these distributions, we estimated their mean (M), coefficient of variation (CV), skewness (S) and kurtosis (K) (Supplementary materials and methods, Section 1.3). The data was produced from at least 3 repeats per condition. Since no significant differences were found between repeats, the data for each condition were merged. Noteworthy, all target genes used have identical sequences upstream and downstream of the promoter region (Materials and methods). Also, as noted above, as they are integrated into single-copy F-plasmids, not anchored to the membrane, they are not expected to be significantly influenced by transcription halting due to positive supercoiling buildup [19,36].

From Fig. 2A, M and CV differ between conditions. S and K also differ between conditions, but do so following a similar trend to one another. Importantly, changes in S and K seem uncorrelated with the values of M and CV. These results suggest that altering the promoter sequence and/or the active regulation allows altering M, CV and S independently.

Observing only subsets of this data, we find it to be in accordance with the model considered (Fig. 1E). E.g., consider the conditions LA, LA(75) and LA(50), which differ only in [RNAP] [14]. In these, as

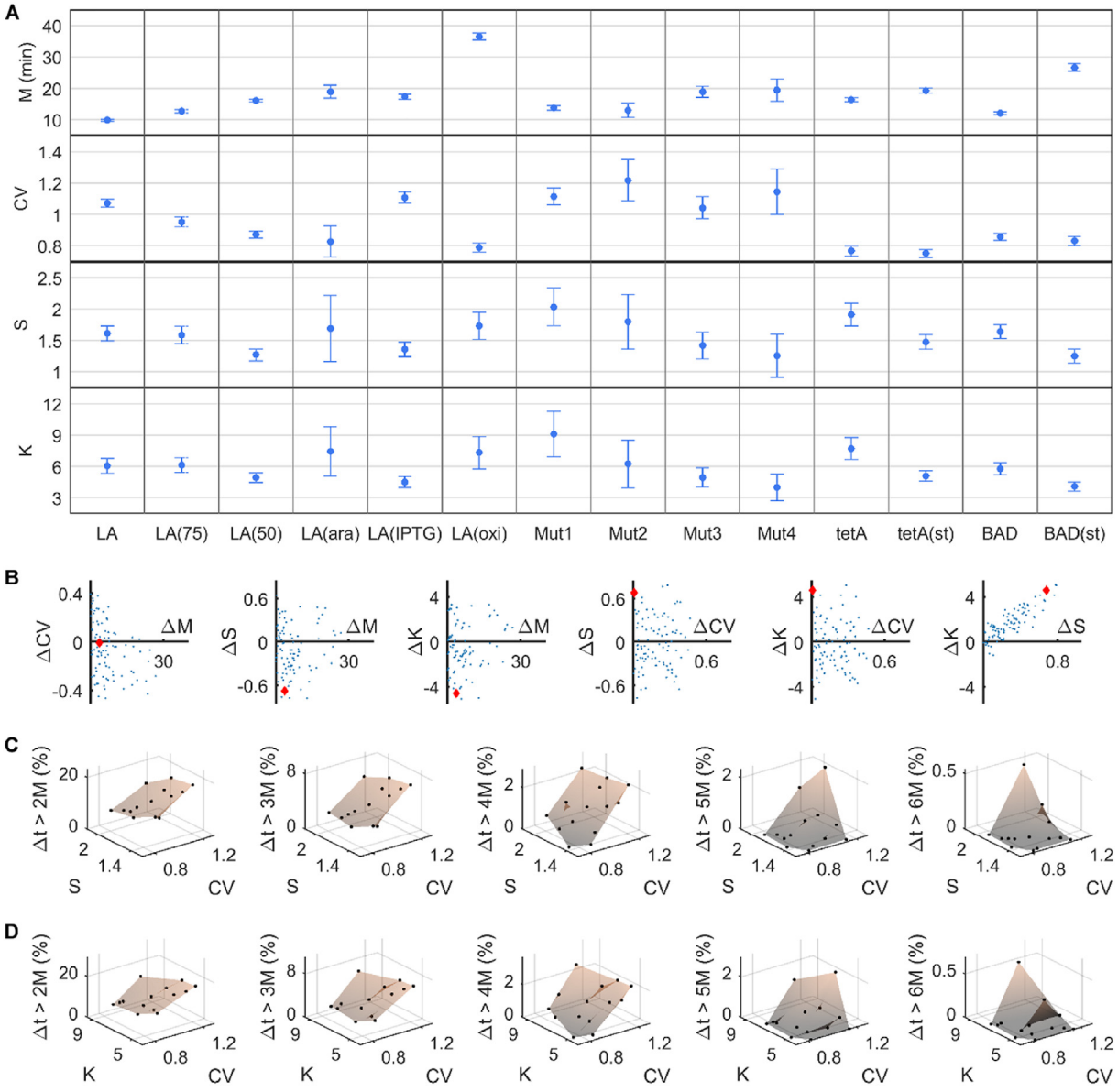

Fig. 2. Skewness (S) and kurtosis (K) affect the probability of crossing upper-bound thresholds in the time length of the intervals between consecutive RNA production events in individual cells (Δt). (A) Mean (M), coefficient of variation (CV), S and K of the distribution of Δt intervals (~600 cells per condition). S and K vary independently from M and CV. Error bars denote SEM. (B) Pairwise differences (Δ) in M, CV, S and K between conditions (blue dots). The red diamond is the difference between LA(IPTG) and Mut1 conditions that illustrates how changes in S and K can be independent from changes in M and CV. (C and D) Percentage of Δt intervals (black dots) that are longer than a given threshold (from 2 M to 6 M) against (C) CV and S, and (D) CV and K. Also shown is the natural neighbour interpolation surface.

[RNAP] decreases, M increases and CV decreases. Meanwhile, S and K decrease (weakly) as [RNAP] decreases. This change is weak enough so that, as shown in the next section, the only significant difference in S is between the two extreme conditions, LA and LA(50), and differences in K are not statistically significant (Supplementary Table S1).

Mutations in $P_{lac/ara-1}$ (Supplementary Fig. S1) also cause significant behavioural changes. Namely, M, CV and S differ between the mutants independently from each other, and only changes in S and K appear to be correlated. The same is observed when considering only the induction schemes of $P_{lac/ara-1}$ (LA, LA(ara) and LA(IPTG) conditions). Oxidative stress also affects M, CV, S and K significantly, when compared to the control. Further, comparing the three promoters tested here ($P_{lac/ara-1}$, P_{tetA} and P_{BAD}), again M, CV and S differ in an independent way, and only the differences between conditions in S and K exhibit a similar trend.

Finally, comparing P_{tetA} and P_{BAD} in the exponential and stationary growth phases (Supplementary Fig. S2A,B), we find that both differ significantly in M, S and K with the growth phase. This agrees with the findings in [23], which reported that the kinetics of rate-limiting steps in transcription changes with σ^{38} numbers (even in σ^{70}-dependent promoters). Interestingly, the differences in M, CV, S and K between growth phases are, qualitatively, the same in both promoters, supporting that they have the same cause.

We also tested whether the differences in M, CV, S and K between conditions could be explained by differences between the distributions of cell lifetimes or between the distributions of intracellular RNAP concentrations. The results of this test indicate that the features of the Δt distribution cannot be explained by the features of either these distributions (Supplementary results, Section 2.4; Supplementary Figs. S4A and S5).

Table 2
Pearson's correlation coefficient r (with the corresponding two-tailed *p*-value) for all conditions, for the subset 'Mutants', where only the promoter sequence differs between conditions, and for the subset 'Regulatory factors', where only the inducers or RNA polymerase concentrations differ between conditions. For *p*-values ≤ 0.05, the null hypothesis that there is no correlation is rejected.

	M vs CV	M vs S	M vs K	CV vs S	CV vs K	S vs K
All conditions	−0.44 (0.12)	−0.19 (0.52)	−0.08 (0.80)	0.01 (0.98)	−0.10 (0.73)	0.94 (< 0.001)
Mutants	−0.12 (0.85)	−0.64 (0.24)	−0.56 (0.32)	0.27 (0.66)	0.07 (0.91)	0.96 (< 0.01)
Regulatory factors	−0.47 (0.43)	−0.24 (0.70)	0.02 (0.98)	−0.17 (0.79)	−0.54 (0.34)	0.91 (0.03)

3.2. Promoter sequence and regulatory factors suffice to alter skewness and kurtosis of RNA production kinetics independently from its mean and coefficient of variation

To determine whether changes in M, CV, S and K between conditions are uncorrelated in a statistical sense, we first calculated linear correlations between each pair of these features when considering all 14 conditions (Fig. 2A). Results in Table 2 show no significant correlation between all pairs, except between S and K. The result holds also when applying the Bonferroni-Holm correction for multiple comparisons (the corrected *p*-value in the case of S and K is < 0.001). Tests for non-linear correlations (Kendall's and Spearman's rank correlation coefficients) give the same qualitative results. While this could be due to the lack of significant changes in M and CV, results in Fig. 2A reject this hypothesis. We thus conclude that all features can differ between conditions in an uncorrelated way, aside from S and K.

We also performed pairwise comparisons of M, CV, S and K between each pair of the 14 conditions. The results (Supplementary Table S1) show statistically significant differences between many pairs of conditions, indicating that all features differ widely between conditions. In detail, one observes that it is possible to alter S and K significantly, while CV is kept unchanged (e.g. between LA(IPTG) and Mut1). Similarly, the same is possible keeping M unchanged (e.g. between LA(50) and tetA).

Next, we quantified the degree with which each feature can differ between conditions while another feature is kept constant. In Fig. 2B we show all pairwise differences in M, CV, S and K between conditions. In all cases, we find that a feature can differ widely while the others remain mostly unchanged, except between S and K.

Finally, we investigated how S and K change as a function of the promoter sequence and the regulatory factors. For this, we considered two subsets of the data above. The first subset ('Mutants') includes the original $P_{lac/ara-1}$ promoter (LA) and the 4 mutants, specifically 1 single-point mutant (Mut1) and 3 three-point mutants (Mut2, Mut3 and Mut4) (Supplementary Fig. S1). The second subset ('Regulatory factors') includes the control (LA), two conditions with different [RNAP] (LA(75) and LA(50)) and two induction schemes (LA(IPTG) and LA(ara)). From Table 2, we conclude that changes in S (and K), due to point mutations and/or due to altering the concentrations of the regulatory factors, are not correlated to the changes in CV and M.

As before, for both subsets, we tested whether the differences in M, CV, S and K between conditions could be explained by differences between the distributions of cell lifetimes. Again, the results showed that the features of the cell lifetimes distributions cannot explain the features of the Δt distribution (Supplementary results, Section 2.4; Supplementary Fig. S4B,C).

3.3. Increasing the skewness and kurtosis of RNA production kinetics enhances the probability of crossing upper bound thresholds in intervals between consecutive RNA production events

Stochastic models of gene expression assuming transcription initiation as a two-step process predict that changing these steps' kinetics can alter the noise in RNA production without changing the mean rate of RNA production [37]. If the intrinsic noise in transcription changes, so will the probability of crossing thresholds based on RNA numbers. Here we quantify this noise by the CV of the Δt distribution [17,18], because this distribution is not affected by noise in RNA degradation.

If this noise was symmetric around the mean of the Δt distribution, the CV would suffice to estimate the probability of threshold crossing. However, recent results [16,17] suggest that it can be significantly asymmetric. As such, a more accurate estimation of threshold crossing probabilities in RNA numbers requires calculating S and K of the Δt distribution.

To test whether S and K differ significantly between the conditions (Supplementary materials and methods, Section 1.3), we first obtained, for each condition, the fraction of individual Δt intervals that are longer than a given threshold. We considered the thresholds 2 M, 3 M, 4 M, 5 M and 6 M, to eliminate influences by the value of M. Results in Supplementary Table S2 indicate that the fraction of intervals that cross a specific threshold differ between conditions, particularly for higher thresholds.

Next, to determine whether it is CV or S (and K) that is responsible for the differences in threshold crossing probabilities between conditions, we plotted the percentage of intervals in each condition that crossed each threshold against CV and S. We also calculated the natural neighbour interpolation surfaces (using MATLAB R2017b function scatteredInterpolant [38]).

Results in Fig. 2C show that for the lower thresholds (2 M and 3 M), varying S does not alter significantly the chance of threshold crossing, while changing CV does. For higher thresholds (4 M and 5 M), both S and CV are relevant. For the highest threshold (6 M), the relevance of S further increases. Equivalent conclusions are reached when considering K instead of S (Fig. 2D).

Overall, tuning S and K of the Δt distribution allows altering significantly the probability of crossing upper-bound thresholds in Δt values and, thus, of crossing lower-bound thresholds of RNA numbers in individual cells.

3.4. Skewness and kurtosis of RNA production kinetics can be tuned by the rate-limiting steps in transcription initiation

Previous studies have established that CV can be tuned by changing the kinetics of the rate-limiting steps in transcription initiation [14,16,17]. In particular, for example, changing the average time spent in the events prior (τ_{prior}) and after (τ_{after}) commitment to open complex formation without changing M, allows tuning noise in RNA production without affecting the rate of this production [16]. We hypothesised that S and K could be similarly regulated.

To test this, for each condition, we first estimated the mean fraction of time spent in the events prior to commitment to open complex formation (τ_{prior}/M) from τ plots (Materials and methods, paragraphs 1–2). Namely, we plotted the inverse of the relative RNA production rate, as measured by qPCR, against the inverse of the relative RNAP concentration, as measured by Western blot (Supplementary materials and methods, Section 1.4). Then, we fitted a line to the data from which we obtain τ_{prior}/M (Fig. 3A and Supplementary Table S3). Finally, from this and the value of M (Fig. 2A), we obtained the absolute values of τ_{prior} and τ_{after} for each condition (Supplementary Table S3).

Cells in the stationary phase (conditions tetA(st) and BAD(st)) are

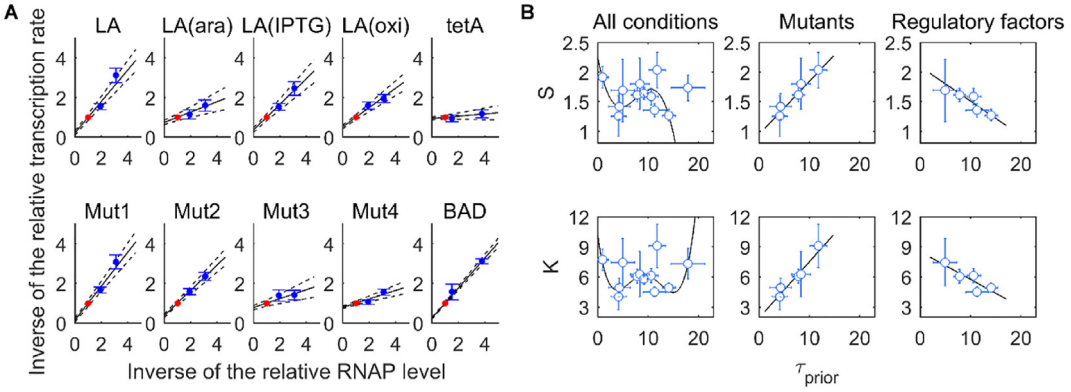

Fig. 3. Skewness (S) and kurtosis (K) of the distribution of intervals between consecutive RNA production events in individual cells change linearly with the fraction of time spent in events prior to commitment to the open complex formation (τ_{prior}). (A) Relative τ plots. Transcription rates are measured by qPCR, and RNA polymerase (RNAP) levels are measured by Western blot (Supplementary Fig. S2C). Values are shown relative to the control condition (red dot). Error bars denote the standard error. The solid line is the best-fitting line, and the dashed lines denote the standard error of the fit. (B) S and K plotted against τ_{prior}. Values plotted for all conditions and for subsets ('Mutants' and 'Regulatory factors'). Error bars denote SEM. The black line is the best-fitting model. The linear relationships are statistically significant when the set of variables allowed to change between conditions is restricted to either the sequence of the promoter or the regulatory factors. When all variables are allowed to differ simultaneously, the best-fitting model is a polynomial of the third or fourth degree.

not considered since, in these conditions, σ^{38} numbers are sufficiently high for the amount of core RNAP enzymes to become a less accurate proxy of the RNAP-σ^{70} holoenzymes levels [23]. Additional factors that may differ include potential sRNA regulation [39,40], ppGpp [41], cAMP (see e.g. [42]) contribute to these differences.

We assessed whether S and K change with τ_{prior}. For this, we plotted S and K against τ_{prior} in each condition (Fig. 3B) and performed likelihood ratio tests (at significance level of 0.05) between the best-fit polynomial models (using weighted total least squares approach [14,43]) with degrees ranging from 0 to N-1, with N being the number of conditions (*p*-values are shown in Supplementary Table S4). We also tested whether the data can be better explained by a model where τ_{prior} does not differ between conditions, by performing a likelihood ratio test between this model and the selected best-fitting polynomial (Supplementary Table S4). For both S and K, the zero-degree and the first-degree polynomial models, as well as the models where τ_{prior} is constant, are rejected in favour of higher-degree polynomials.

The fact that S and K are best fit by, respectively, third and fourth degree polynomials (that still do not explain all data points) illustrates the level of complexity of the data. This is likely due to the conditions differing in several factors (promoter, induction scheme, etc.). We thus next consider, as above, the subsets 'Mutants' and 'Regulatory factors'. For each, we perform, also as above, likelihood ratio tests to determine the best fitting models (Supplementary Table S4). In both subsets, a 1st degree model is preferred.

Meanwhile, from the Pearson's correlation coefficient (with the corresponding two-tailed *p*-value) between τ_{prior} and skewness (S) and kurtosis (K), for the subsets 'Mutants' and 'Regulatory factors', we find a significant correlation in all cases (absolute correlation values above 0.85 and *p*-values ≤ 0.05), except for K in 'Regulatory factors', where the *p*-value equals 0.06. Overall, the results suggest that, similarly to M and CV, tuning τ_{prior} can regulate S and K. This implies that the lower bound threshold crossing probability of RNA numbers over time can be tuned.

Next, we performed the same analysis for changing τ_{after} and τ_{prior}/M. Contrary to when considering τ_{prior}, the results (Supplementary Fig. S6 and Supplementary Tables S5-S6) do not allow establishing statistically significant relationships (also the *p*-values from the Pearson's correlation were larger than 0.05).

Interestingly, the linear relationships of S and K with τ_{prior} are positive in the subset 'Mutants' and negative in the subset 'Regulatory factors'. This strongly indicates that τ_{prior} is not the only parameter defining these features. Namely, we hypothesise that these relationships may depend on what causes τ_{prior} to differ between the conditions. For instance, in one subset, the difference may be due to differences in the mean time required by the RNAP to complete a closed complex formation, while in the other subset the differences may be in the number of times that the RNAP fails to commit to the open complex formation. These potential differences could be accounted for in the model by tuning k_1, k_{-1} and k_2 (Supplementary materials and methods, Section 1.6), but cannot be detected by the measurements conducted here. Future work is needed to test this hypothesis.

3.5. Skewness and kurtosis of the RNA production kinetics and of the distribution of protein expression levels in individual cells are negatively correlated

To assess if changes in S and K of the Δt distribution could affect the phenotypic distribution of cell populations, we next investigate whether these changes result in significant changes in the distribution of protein expression levels of a cell population. This is expected given the known coupling between transcription and translation in prokaryotes [44–46]. Nevertheless, it is reasonable to assume that noise in the stochastic process of translation (e.g. on the time to be completed once initiated) would render changes in S and K ineffectual on protein expression levels. A model of gene expression in prokaryotes accounting for the coupling between the two processes is shown in Supplementary materials and methods, Section 1.6.

We first tested whether the mean protein expression levels of the cell populations follow their mean RNA numbers. For that, we measured RNA numbers (by microscopy) and protein mean expression levels (by flow cytometry) produced under the control of $P_{lac/ara-1}$ for various induction conditions. We expect the same relationship in all other constructs used here, as they have identical sequences following the promoter sequence. Results in Supplementary Fig. S7 show that the average number of proteins in a cell population follows the average RNA numbers.

Given this, since M of the Δt distribution is negatively correlated with the mean RNA numbers of the cell population, one can expect it to also be negatively correlated to the mean number of proteins. Using the same promoter as a case-study, we tested whether the skewness and kurtosis of the distribution of protein expression levels of a cell population are sensitive to the induction strength. For this, we measured the total fluorescence intensity level of the proteins expressed by $P_{lac/ara-1}$ in individual cells for various induction levels using flow cytometry (Materials and methods). From these, for each induction level, we

obtained the distribution of fluorescence of individual cells (in arbitrary units). For each of these distributions, we estimated the mean (M_P), skewness (S_P) and kurtosis (K_P) as previously (Supplementary materials and methods, Section 1.3). From Supplementary Fig. S8, we find that S_P and K_P can differ with induction strength. Also, it is possible to have, for similar values of M_P, significantly different values of S_P and K_P (e.g. conditions 0 to 25 μM). Further, conditions differing in M_P can have similar values of S_P and K_P (beyond 100 μM). Overall, we find that, as for the Δt distributions, S_P and K_P can change independently from M_P and vice versa.

Next, we investigate whether changes in S and K of the Δt distribution due to changing the promoter sequence or its regulation reflect on the distribution of protein expression levels, as expected from the model. For this, we consider, respectively, the subsets 'Mutations' and 'Induction schemes'. We note that, within these subsets, the cells are grown under identical culture conditions and do not differ in their fundamental physiology, and are therefore not expected to differ in, e.g., ribosome population and/or in any other global gene expression regulators, such as [RNAP] or σ factors. For these reasons, here we do not consider the other conditions in Table 1, as the translation rate or protein maturation time may differ significantly from the control.

For each condition considered, we measured the fluorescence intensity from the target proteins by flow cytometry (Materials and methods) and obtained the single-cell distributions of protein fluorescence intensity. Next, we estimated its M_P (in arbitrary units), S_P and K_P, as previously. We also measured M_P for cells with an uninduced $P_{lac/ara-1}$ to obtain a reference point for the values of M_P. In this regard, the LA(ara) condition was not included in the subsequent analysis since, for unknown reasons, its protein expression levels were not significantly above those of the uninduced $P_{lac/ara-1}$ (Fig. 4).

In Fig. 4, we show M_P, S_P and K_P plotted against M, S and K, respectively, along with the best-fitting models obtained by likelihood ratio tests (Supplementary Table S7). In all cases, the linear model is preferred. We also calculated the Pearson's correlation coefficient for each case. The results agree with the likelihood ratio tests. Namely, there are strong, statistically significant (p-values ≤ 0.05), negative correlations between M and M_P (−0.82) and between S and S_P (−0.86). Between K vs K_P the negative correlation is also strong (−0.70), but the p-value is 0.12, likely due to higher uncertainty. From the statistically significant linear relationships, we conclude that the differences in skewness and kurtosis of the Δt distribution between conditions result in statistically significant differences between the skewness and kurtosis of the corresponding protein distributions, in a manner that is consistent with the model. As a side note, our data does not allow investigating whether a similar (expected) correlation exist in the case of CV and CV_P, since LA, LA(IPTG), and the mutant promoters have CV values that cannot be distinguished in a statistical sense (Supplementary Table S1).

Finally, to assess if the values of M could explain the values of S_P and K_P, we performed likelihood ratio tests (as above) between M and S_P and between M and K_P. A polynomial model of the 1st order was rejected in both cases (p-values equal 0.04 and 0.02, respectively). Also, we failed to find linear correlations (p-values equal 0.06 and 0.25, respectively). We conclude that M is not correlated with either S_P or K_P, as expected from the lack of the correlation between M and S or K.

4. Discussion and conclusions

Previous research have established that bacterial transcription is mostly regulated at the stage of initiation [10–12,47]. This regulation, e.g. by transcription factors and σ factors, affects the mean and variance in RNA and protein numbers [10–12,19,48]. From the dynamics point of view, these and similar regulatory molecules were shown to have direct effect on the kinetics of the rate-limiting steps in transcription initiation of a gene (assessed here by τ_{prior} and τ_{after}), resulting in changes in the mean and variance of its distribution of intervals between consecutive RNA production events in individual cells (Δt distribution) [14,23].

Here we provided evidence that the fraction of cells that reach high thresholds in RNA and protein numbers of an externally regulated gene can be tuned by altering the skewness and kurtosis of its Δt distribution. Also, we showed that this can be achieved without significantly altering the mean and CV of this distribution. Further, this regulation is possible by tuning τ_{prior} and τ_{after} alone which can be altered by changing the promoter sequence, the induction scheme, or the intracellular RNAP concentration.

On the other hand, we did not find significant evidence that the skewness and kurtosis could be altered independently from one another. Instead, they exhibit a strong positive correlation (Fig. 2B, ΔK vs ΔS, and Table 2). We suggest that this may be due to the variability of the time length between transcription events along with the existence of mechanical constraints imposed by the transcription machinery. This variability is visible in Fig. S3, which shows that the distributions of intervals between transcriptions are broad, with several intervals having a short time-length. This limits how much the kurtosis of this distribution can increase by increasing the tail on the left side. This limit does not exist on the right side. Thus, increasing the kurtosis of one of these distributions by increasing the size of the right tail cannot be easily compensated on the left side so that the skewness remains unaltered.

Regulation of asymmetry and tailedness of gene expression, so far, has only been considered in the context of small genetic circuits or complex regulatory pathways (e.g. [3]). Given the above, our findings suggest that regulatory mechanisms of individual genes suffice for this regulation as well. In particular, based on the data from the conditions in Table 1, we found statistically significant linear relationships between τ_{prior} and the skewness and kurtosis of the Δt distribution, provided that either only the promoter sequence or the regulatory factors (i.e. inducers and RNAP concentrations) differ between the conditions. We hypothesise that relationships more complex than linear are also possible, if more than one parameter is allowed to change. E.g. in the future it would be of interest to investigate whether the data in Fig. 3B

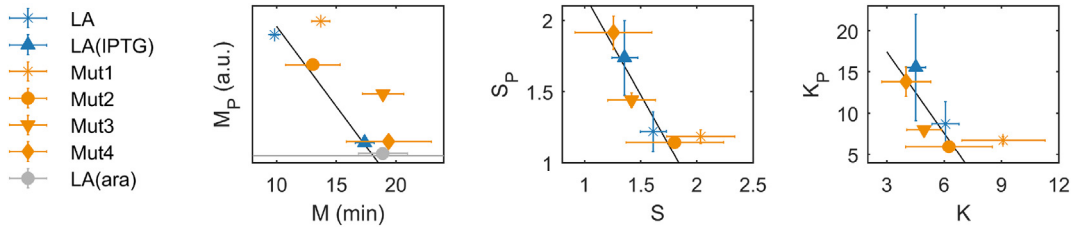

Fig. 4. Mean (M), skewness (S), and kurtosis (K) of the distribution of protein expression levels in individual cells change linearly with the corresponding features of the distribution of time intervals between consecutive RNA production events (Δt distribution). (From left to right) M_P, S_P and K_P of the single-cell distributions of protein levels against the corresponding feature of the Δt distributions (extracted from Fig. 2A). Error bars denote SEM (in some cases, the SEM is too small to produce visible error bars). The solid line is the best fitting model. On the left plot, the horizontal grey line corresponds to M_P for an uninduced $P_{lac/ara-1}$ which is used as a reference point (SEM is too small to be represented). M_P of LA(ara) is not considered in model fitting.

could be better explained by consider both τ_{prior} and τ_{after} simultaneously. Nevertheless, the linear relationships found here are evidence that the skewness and kurtosis are evolvable (i.e. sequence dependent) and adaptable (i.e. subject to regulation). Meanwhile, the strong correlation between RNA production kinetics and single-cell distribution of protein levels suggests that tuning these skewness and kurtosis can have a significant impact on the phenotypic distribution of the cell population.

It is well known that the two rate-limiting steps of transcription initiation here considered (i.e. the events prior and after commitment to open complex formation) are composed of specific 'sub-steps', such as promoter escape [49–51], reversibility of the closed complex formation and isomerization [13,52,53]. Further developments in the dissection techniques of the in vivo kinetics of these sub-steps during transcription initiation should allow characterising, in greater detail, their contributions to the regulation of the skewness and kurtosis of the distributions of RNA production kinetics and corresponding protein numbers. This should also allow establishing precise methods for tuning the skewness and kurtosis of these distributions.

It is worth noting that the findings here reported do not discard the importance of other mechanisms of regulation of protein numbers in *E. coli*, such as regulation by sRNAs [39,40,54]. Here we did not consider this mechanism since all target genes studied shared the same elongation region. It will be of interest to study whether this post-transcription regulation process also allows tuning the skewness and kurtosis of single-cell distributions of protein numbers, particularly given its known effects on the cell-to-cell variability in protein numbers [55,56] and protein numbers' threshold-crossing propensities [39,57].

Finally, while a strict relationship between the skewness and kurtosis in the RNA and protein numbers was established here, the implications of these findings in the context of the qualitative behaviour of genetic circuits remain to be demonstrated. We expect the amplitude of these effects to differ with the circuit topology, as in the case of mean and variance [58–60]. If the effects are significant, direct regulation of these features in genetic circuits (by tuning the rate limiting steps of the component genes) should allow a more precise control of their kinetics, towards enhancing their robustness to fluctuations in molecular numbers or environmental changes, and sensitivity to external signals.

Funding

Work supported by Tampere University of Technology Graduate School Grant (Finland) [to S.S.]; Pirkanmaa Regional Fund [to V.K.K.]; Academy of Finland [295027, 305342 to A.S.R.]; and Jane and Aatos Erkko Foundation [610536 to A.S.R.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Funding for open access charge: Academy of Finland [295027].

Conflict of interest

The authors declare that they have no conflict of interest.

Transparency document

The Transparency document associated with this article can be found, in online version.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2018.12.005.

References

[1] S.S. Shen-Orr, R. Milo, S. Mangan, U. Alon, Network motifs in the transcriptional regulation network of *Escherichia coli*, Nat. Genet. 31 (2002) 64–68, https://doi.org/10.1038/ng881.
[2] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, U. Alon, Network motifs: simple building blocks of complex networks, Science 298 (2002) 824–827, https://doi.org/10.1126/science.298.5594.824.
[3] U. Alon, Network motifs: theory and experimental approaches, Nat. Rev. Genet. 8 (2007) 450–461, https://doi.org/10.1038/nrg2102.
[4] E. Kussell, S. Leibler, Phenotypic diversity, population growth, and information in fluctuating environments, Science 309 (2005) 2075–2078, https://doi.org/10.1126/science.1114383.
[5] Y. Liu, A. Beyer, R. Aebersold, On the dependency of cellular protein levels on mRNA abundance, Cell 165 (2016) 535–550, https://doi.org/10.1016/j.cell.2016.03.014.
[6] C. Vogel, E.M. Marcotte, Insights into the regulation of protein abundance from proteomic and transcriptomic analyses, Nat. Rev. Genet. 13 (2012) 227–232, https://doi.org/10.1038/nrg3185.
[7] J.A. Bernstein, A.B. Khodursky, P.-H. Lin, S. Lin-Chao, S.N. Cohen, Global analysis of mRNA decay and abundance in *Escherichia coli* at single-gene resolution using two-color fluorescent DNA microarrays, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 9697–9702, https://doi.org/10.1073/pnas.112318199.
[8] H. Chen, K. Shiroguchi, H. Ge, X.S. Xie, Genome-wide study of mRNA degradation and transcript elongation in *Escherichia coli*, Mol. Syst. Biol. 11 (2015) 781, https://doi.org/10.15252/msb.20145794.
[9] M.P. Deutscher, Degradation of RNA in bacteria: comparison of mRNA and stable RNA, Nucleic Acids Res. 34 (2006) 659–666, https://doi.org/10.1093/nar/gkj472.
[10] S.M. McLeod, R.C. Johnson, Control of transcription by nucleoid proteins, Curr. Opin. Microbiol. 4 (2001) 152–159, https://doi.org/10.1016/S1369-5274(00)00181-8.
[11] E.F. Ruff, A.C. Drennan, M.W. Capp, M.A. Poulos, I. Artsimovitch, T.M. Record Jr., *E. coli* RNA polymerase determinants of open complex lifetime and structure, J. Mol. Biol. 247 (2015) 2435–2450, https://doi.org/10.1016/j.jmb.2015.05.024.
[12] D.F. Browning, S.J.W. Busby, Local and global regulation of transcription initiation in bacteria, Nat. Rev. Microbiol. 14 (2016) 638–650, https://doi.org/10.1038/nrmicro.2016.103.
[13] P.L. deHaseth, M.L. Zupancic, T.M. Record Jr., RNA polymerase-promoter interactions: the comings and goings of RNA polymerase, J. Bacteriol. 180 (1998) 3019–3025 (PMID: 9620948).
[14] J. Lloyd-Price, S. Startceva, V. Kandavalli, J.G. Chandraseelan, N. Goncalves, S.M.D. Oliveira, A. Häkkinen, A.S. Ribeiro, Dissecting the stochastic transcription initiation process in live *Escherichia coli*, DNA Res. 23 (2016) 203–214, https://doi.org/10.1093/dnares/dsw009.
[15] W.R. McClure, Rate-limiting steps in RNA chain initiation, Proc. Natl. Acad. Sci. U. S. A. 77 (1980) 5634–5638, https://doi.org/10.1073/pnas.77.10.5634.
[16] J. Mäkelä, V. Kandavalli, A.S. Ribeiro, Rate-limiting steps in transcription dictate sensitivity to variability in cellular components, Sci. Rep. 7 (2017) 10588, https://doi.org/10.1038/s41598-017-11257-2.
[17] S.M.D. Oliveira, A. Häkkinen, J. Lloyd-Price, H. Tran, V. Kandavalli, A.S. Ribeiro, Temperature-dependent model of multi-step transcription initiation in *Escherichia coli* based on live single-cell measurements, PLoS Comput. Biol. 12 (2016) e1005174, , https://doi.org/10.1371/journal.pcbi.1005174.
[18] A. Häkkinen, A.S. Ribeiro, Characterizing rate limiting steps in transcription from RNA production times in live cells, Bioinformatics 32 (2016) 1346–1352, https://doi.org/10.1093/bioinformatics/btv744.
[19] S. Chong, C. Chen, H. Ge, X.S. Xie, Mechanism of transcriptional bursting in bacteria, Cell 158 (2014) 314–326, https://doi.org/10.1016/j.cell.2014.05.038.
[20] I. Golding, J. Paulsson, S.M. Zawilski, E.C. Cox, Real-time kinetics of gene activity in individual bacteria, Cell 123 (2005) 1025–1036, https://doi.org/10.1016/j.cell.2005.09.031.
[21] A. Häkkinen, A.S. Ribeiro, Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data, Bioinformatics 31 (2015) 69–75, https://doi.org/10.1093/bioinformatics/btu592.
[22] H. Tran, S.M.D. Oliveira, N. Goncalves, A.S. Ribeiro, Kinetics of the cellular intake of a gene expression inducer at high concentrations, Mol. BioSyst. 11 (2015) 2579–2587, https://doi.org/10.1039/C5MB00244C.
[23] V.K. Kandavalli, H. Tran, A.S. Ribeiro, Effects of σ factor competition are promoter initiation kinetics dependent, Biochim. Biophys. Acta 1859 (2016) 1281–1288, https://doi.org/10.1016/j.bbagrm.2016.07.011.
[24] J.R. Peterson, J.A. Cole, J. Fei, T. Ha, Z.A. Luthey-Schulten, Effects of DNA replication on mRNA noise, Proc. Natl. Acad. Sci. U. S. A. 112 (2015) 15886–15891, https://doi.org/10.1073/pnas.1516246112.
[25] H. Lineweaver, D. Burk, The determination of enzyme dissociation constants, J. Am. Chem. Soc. 56 (1934) 658–666, https://doi.org/10.1021/ja01318a036.
[26] R. Lutz, H. Bujard, Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I_1-I_2 regulatory elements, Nucleic Acids Res. 25 (1997) 1203–1210, https://doi.org/10.1093/nar/25.6.1203.
[27] I. Golding, E.C. Cox, RNA dynamics in live *Escherichia coli* cells, Proc. Natl. Acad. Sci. U. S. A. 101 (2004) 11310–11315, https://doi.org/10.1073/pnas.0404443101.
[28] A. Gupta, J. Lloyd-Price, R. Neeli-Venkata, S.M.D. Oliveira, A.S. Ribeiro, In vivo kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*, Biophys. J. 106 (2014) 1928–1937, https://doi.org/10.1016/j.bpj.2014.03.035.
[29] B.P. Bratton, R.A. Mooney, J.C. Weisshaar, Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*, J. Bacteriol. 193 (2011) 5138–5146, https://doi.org/10.1128/JB.00198-11.
[30] J. Sambrook, E.F. Fritsch, T. Maniatis, Molecular Cloning: A Laboratory Manual, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York,

0879693096, 1989.
[31] N.S.M. Goncalves, L. Martins, H. Tran, S.M.D. Oliveira, R. Neeli-Venkata, J.M. Fonseca, A.S. Ribeiro, *In vivo* single-molecule dynamics of transcription of the viral T7 Phi 10 promoter in *Escherichia coli*, The 8th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies (BIOTECHNO 2016), 2016 978-1-61208-488-6, pp. 9–15.
[32] D.G. Gibson, L. Young, R.-Y. Chuang, J.C. Venter, C.A. Hutchison III, H.O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases, Nat. Methods 6 (2009) 343–345, https://doi.org/10.1038/nmeth.1318.
[33] K.J. Livak, T.D. Schmittgen, Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method, Methods 25 (2001) 402–408, https://doi.org/10.1006/meth.2001.1262.
[34] A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J.M. Fonseca, A.S. Ribeiro, CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*, Bioinformatics 29 (2013) 1708–1709, https://doi.org/10.1093/bioinformatics/btt194.
[35] M. Razo-Mejia, S.L. Barnes, N.M. Belliveau, G. Chure, T. Einav, M. Lewis, R. Phillips, Tuning transcriptional regulation through signaling: a predictive theory of allosteric induction, Cell Syst. 6 (2018) 456–469.e10, https://doi.org/10.1016/j.cels.2018.02.004.
[36] J.D. Boeke, P. Model, A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene *III* protein retains it in the membrane, Proc. Natl. Acad. Sci. U. S. A. 79 (1982) 5200–5204, https://doi.org/10.1073/pnas.79.17.5200.
[37] S.M.D. Oliveira, M.N.M. Bahrudeen, S. Startceva, A.S. Ribeiro, Estimating effects of extrinsic noise on model genes and circuits with empirically validated kinetics, in: M. Pelillo, I. Poli, A. Roli, R. Serra, D. Slanzi, M. Villani (Eds.), Artificial Life and Evolutionary Computation. WIVACE 2017, Springer, 2018, pp. 181–193, , https://doi.org/10.1007/978-3-319-78658-2_14.
[38] I. Amidror, Scattered data interpolation methods for electronic imaging systems: a survey, J. Electron. Imaging 11 (2002) 157–176, https://doi.org/10.1117/1.1455013.
[39] E. Levine, Z. Zhang, T. Kuhlman, T. Hwa, Quantitative characteristics of gene regulation by small RNA, PLoS Biol. 5 (2007) e229, , https://doi.org/10.1371/journal.pbio.0050229.
[40] E.G.H. Wagner, P. Romby, Small RNAs in bacteria and archaea: who they are, what they do, and how they do it, Adv. Genet. 90 (2015) 133–208, https://doi.org/10.1016/bs.adgen.2015.05.001.
[41] C. Condon, C. Squires, C.L. Squires, Control of rRNA transcription in *Escherichia coli*, Microbiol. Rev. 59 (1995) 623–645 (PMID: 8531889).
[42] C.M. Johnson, R.F. Schleif, In vivo induction kinetics of the arabinose promoters in *Escherichia coli*, J. Bacteriol. 177 (1995) 3438–3442 (PMID: 7768852).
[43] M. Krystek, M. Anton, A weighted total least-squares algorithm for fitting a straight line, Meas. Sci. Technol. 18 (2007) 3438–3442, https://doi.org/10.1088/0957-0233/18/11/025.
[44] O. Dahan, H. Gingold, Y. Pilpel, Regulatory mechanisms and networks couple the different phases of gene expression, Trends Genet. 27 (2011) 316–322, https://doi.org/10.1016/j.tig.2011.05.008.
[45] C. Yanofsky, Attenuation in the control of expression of bacterial operons, Nature 289 (1981) 751–758, https://doi.org/10.1038/289751a0.
[46] S. Proshkin, A.R. Rahmouni, A. Mironov, E. Nudler, Cooperation between translating ribosomes and RNA polymerase in transcription elongation, Science 328 (2010) 504–508, https://doi.org/10.1126/science.1184939.
[47] D.F. Browning, S.J.W. Busby, The regulation of bacterial transcription initiation, Nat. Rev. Microbiol. 2 (2004) 57–65, https://doi.org/10.1038/nrmicro787.
[48] W.R. McClure, Mechanism and control of transcription initiation in prokaryotes, Annu. Rev. Biochem. 54 (1985) 171–204, https://doi.org/10.1146/annurev.bi.54.070185.001131.
[49] D. Duchi, D.L.V. Bauer, L. Fernandez, G. Evans, N. Robb, L.C. Hwang, K. Gryte, A. Tomescu, P. Zawadzki, Z. Morichaud, et al., RNA polymerase pausing during initial transcription, Mol. Cell 63 (2016) 939–950, https://doi.org/10.1016/j.molcel.2016.08.011.
[50] L.M. Hsu, Promoter escape by *Escherichia coli* RNA polymerase, EcoSal Plus 3 (2008) 1–16, https://doi.org/10.1128/ecosalplus.4.5.2.2.
[51] A.N. Kapanidis, E. Margeat, S.O. Ho, E. Kortkhonjia, S. Weiss, R.H. Ebright, Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism, Science 314 (2006) 1144–1147, https://doi.org/10.1126/science.1131399.
[52] L.M. Hsu, Promoter clearance and escape in prokaryotes, Biochim. Biophys. Acta 1577 (2002) 191–207, https://doi.org/10.1016/S0167-4781(02)00452-9.
[53] T.M. Record Jr., W.S. Reznikoff, M.L. Craig, K.L. McQuade, P.J. Schlax, *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation, in: F.C. Neidhardt, R. Curtiss, J.L. Ingraham, E.C.C. Lin, K.B. Low, B. Magasanik, W.S. Reznikoff, M. Riley, D. Schneider, H.E. Umbarger (Eds.), *Escherichia coli* and *Salmonella typhimurium*: Cellular and Molecular Biology, 2nd ed., ASM press, Washington, DC, 1555810845, 1996, pp. 792–821.
[54] G. Storz, J. Vogel, K.M. Wassarman, Regulation by small RNAs in bacteria: expanding frontiers, Mol. Cell 43 (2011) 880–891, https://doi.org/10.1016/j.molcel.2011.08.022.
[55] R. Arbel-Goren, A. Tal, T. Friedlander, S. Meshner, N. Costantino, D.L. Court, J. Stavans, Effects of post-transcriptional regulation on phenotypic noise in *Escherichia coli*, Nucleic Acids Res. 41 (2013) 4825–4834, https://doi.org/10.1093/nar/gkt184.
[56] R. Arbel-Goren, A. Tal, B. Parasar, A. Dym, N. Costantino, J. Muñoz-García, D.L. Court, J. Stavans, Transcript degradation and noise of small RNA-controlled genes in a switch activated network in *Escherichia coli*, Nucleic Acids Res. 44 (2016) 6707–6720, https://doi.org/10.1093/nar/gkw273.
[57] P. Mehta, S. Goyal, N.S. Wingreen, A quantitative comparison of sRNA-based and protein-based gene regulation, Mol. Syst. Biol. 4 (2016) 221, https://doi.org/10.1038/msb.2008.58.
[58] E.D. Cameron, J.J. Collins, Tunable protein degradation in bacteria, Nat. Biotechnol. 32 (2014) 1276–1281, https://doi.org/10.1038/nbt.3053.
[59] L.G. Morelli, F. Jülicher, Precision of genetic oscillators and clocks, Phys. Rev. Lett. 98 (2007) 228101, https://doi.org/10.1103/PhysRevLett.98.228101.
[60] R. Zhu, A.S. Ribeiro, D. Salahub, S.A. Kauffman, Studying genetic regulatory networks at the molecular level: delayed reaction stochastic models, J. Theor. Biol. 246 (2007) 725–745, https://doi.org/10.1016/j.jtbi.2007.01.021.