Transcription closed and open complex formation coordinate expression of genes with a shared promoter region

Antti Häkkinen[1], Samuel M. D. Oliveira[1], Ramakanth Neeli-Venkata[1] and Andre S. Ribeiro[1]

[1]BioMediTech Institute and Department of Signal Processing, Tampere University of Technology, PO box 553, 33101, Tampere, Finland

Abstract

Many genes are spaced closely, allowing coordination without explicit control through shared regulatory elements and molecular interactions. We study the dynamics of a stochastic model of a gene-pair in a head-to-head configuration, sharing promoter elements, which accounts for the rate-limiting steps in transcription initiation. We find that only in specific regions of the parameter space of the rate-limiting steps is orderly coexpression exhibited, suggesting that successful cooperation between closely spaced genes requires the co-evolution of compatible rate-limiting step configuration. The model predictions are validated using *in vivo* single-cell, single-RNA measurements of the dynamics of pairs of genes sharing promoter elements. Our results suggest that, in *E. coli*, the kinetics of the rate-limiting steps in active transcription can play a central role in shaping the dynamics of gene-pairs sharing promoter elements.

1 Introduction

Closely-spaced gene-pairs abound in genomes of all life forms, from human [1, 2] to prokaryotes [3, 4]. Further, they are highly conserved [2, 5], suggesting that they provide functionality that is selectively advantageous.

Gene-pairs can be arranged head-to-head (transcriptionally divergent), with their transcription start sites (TSS) closely located, sharing promoter elements such as transcription factor binding sites [1, 2, 4]. Head-to-tail (tandem) and tail-to-tail (convergent) overlapping gene-pairs are also found, allowing interference between RNA polymerases (RNAP) [6] and/or with transcription factors [7, 8]. Each configuration can vary in several parameters, such as distance between TSSs, which affect transcription of the component genes [9, 10, 11, 4], allowing co-regulation without explicit control mechanisms. The multitude of naturally occurring configurations suggests that each configuration possesses distinct selective advantages.

While some configurations have been identified and their ubiquity established by models and measurements [2, 12, 11, 13], the range of possible behaviors and advantages as a gene regulation mechanism remain largely uncharacterized. Such characterization would benefit understanding the array of tasks that organisms such as *Escherichia coli* perform using closely-spaced promoters, as opposed to using individual genes or genes connected by transcription factors.

Much research has been conducted on the global co-expression patterns of closely located genes, particularly in eukaryotes. This has resulted in the accumulation of evidence for the existence of various, complementary mechanisms at play [14, 15, 16, 17, 18]. For instance, Limi et al. reported that two highly expressed genes, *Cryba4* and *Crybb1*, can be simultaneously transcribed from adjacent bidirectional promoters in humans, despite their close proximity [14]. Meanwhile, Behjati et al. found that bidirectional promoters employ three major architectures across the human genome, varying in their DNA accessibility,

histone modifications, DNA methylation and transcription factor binding profiles [15]. Direct regulatory interference between spatially close genes has also been reported in human, and it has been associated with reduced gene expression noise [16]. Finally, Scruggs et al. established that the distance between the promoters can be used as a regulatory mechanism of the degree of interaction between their dynamics [17].

However, some of these findings may not apply to bacterial genomes, due to various structural differences. For example, eukaryotic genomes are an order of magnitude larger and they are partitioned into linear chromosomes and confined to the nucleus. Meanwhile bacterial genomes are singular, circular, and bacteria lack membrane-bound nuclei [19, 20]. Further, eukaryotic DNA is packed around histones, while prokaryotic DNA is compressed by supercoiling [20, 21, 22], which is expected to cause major differences into how the cells of these two domains access their DNA, for both replication as well as transcription, which may explain why the dynamics of these processes differ so widely between them (e.g. DNA replication is two orders of magnitude slower in prokaryotes [19]). Consequently, we expect the regulatory mechanisms and the dynamics of closely spaced promoters to differ significantly between eukaryotes and prokaryotes, which may alter which mechanism and whether co-expression or interference has a dominant role in each domain.

One aspect that remains unexplored with respect to the bidirectional promoters is the existence of multiple rate-limiting steps during transcription initiation [23, 24]. As only some of these steps are physically involved in the gene-pair interactions, we expect the nature of the rate-limiting steps of each promoter to affect the dynamics of closely-spaced configurations. Importantly, the durations of the open complex formation of a strong and a weak promoter can differ from little to up to two orders of magnitude [25] and live cell single-RNA measurements suggest that different promoters are rate-limited at different stages of transcription initiation [26, 27, 24, 28, 29]. As such, it is plausible that promoters whose initiation kinetics are similar in mean duration but whose rate-limiting step structures differ will feature different dynamics in the bidirectional configuration.

Here, we study the dynamics of a stochastic model of a gene-pair in a head-to-head configuration sharing promoter elements (the most common closely-spaced gene-pair configuration [2, 5]) as a function of the rate-limiting step configuration of each gene. We analyze the models using analytical stochastic methods. Next, we validate the main findings by performing time-lapse microscopy measurements of individual genes and in pairs of genes sharing promoter elements, at the single-RNA level, in live *E. coli*.

2 Methods

2.1 Models

Transcription in *E. coli* starts when an RNAP, recruiting the appropriate σ-factor, specifically binds to a promoter. This creates a closed complex of the RNAP and DNA, which can require several trials before stabilizing [30]. In strong promoters, this step is nearly irreversible [31]. The virtually irreversible open complex formation follows, consisting of e.g. DNA unwinding and compaction [32] and the RNAP clamp assembly [33].

We assume a variant of a model of transcription initiation of the overlapping promoters of the galactose operon in the absence of cAMP-CRP [3]. The transcribed promoter is stochastically selected based on the relative affinities between the two promoters and the RNAP, encoded in the forward rates of the closed complex formation of each promoter. After the selection, the remaining steps of transcription initiation occur at the promoter region [23]. The following stochastic chemical [34, 35] reactions are used to model this:

$$\begin{aligned} P_0 &\xrightarrow{k_1} I_1 \xrightarrow{k_3} P_0 + X \\ P_0 &\xrightarrow{k_2} I_2 \xrightarrow{k_4} P_0 + Y \end{aligned} \tag{1}$$

where P_0 represents a free promoter (unoccupied by an RNAP), I_1 and I_2 represent intermediate transcriptional complexes committed to transcribing genes 1 and 2, respectively, and X and Y represent the messenger RNA products (or, if they closely follow [36], proteins) of genes 1 and 2, respectively. A schematic is provided in Figure 2 (a), and a thorough analysis in the supplement.

If the genes did not share promoter elements, the intervals between productions of X (gene 1) would be [29]:

$$\begin{aligned} & {\tau_X}^{(1)} \sim \mathcal{E}^-(k_1, k_3) \\ \text{with} \quad & \mathbb{E}\left[{\tau_X}^{(1)}\right] = {k_1}^{-1} + {k_3}^{-1} \\ & \mathbb{V}\text{ar}\left[{\tau_X}^{(1)}\right] = {k_1}^{-2} + {k_3}^{-2} \end{aligned} \tag{2}$$

where $\mathcal{E}^-(\lambda_1, \cdots, \lambda_n)$ represents a hypoexponential distribution (i.e. a sum of exponential distributions) with rates $\lambda_1, \cdots, \lambda_n$. Similarly, the production intervals for gene 2 would be ${\tau_Y}^{(1)} \sim \mathcal{E}^-(k_2, k_4)$.

These distributions are of low noise, as measured by the coefficient of variation (standard deviation over the mean), as this quantity equals unity for Poissonian production (exponential production intervals). More specifically, the noise is determined by the ratio of k_1 and k_3. Regardless of the mean, it is minimized for steps of equal duration and maximized when a single step is rate-limiting. The dynamics of an individual gene is unaffected by the step order (i.e. interchanging k_1 and k_3 has no effect on ${\tau_X}^{(1)}$).

Regardless of the configuration, the mean and variance of the production intervals are linked to that of the produced RNAs. In the long-term (infinite time), the mean and variance of produced RNA per unit time are [37]:

$$\begin{aligned} \mu_Z &\doteq \lim_{t\to\infty} \frac{\mathbb{E}[\,Z(t)\,]}{t} = \mathbb{E}[\tau_Z]^{-1} \\ \eta_Z\,\mu_Z &\doteq \lim_{t\to\infty} \frac{\mathbb{V}\text{ar}[\,Z(t)\,]}{t} = \mathbb{V}\text{ar}[\tau_Z]\;\mathbb{E}[\tau_Z]^{-3} \end{aligned} \tag{3}$$

i.e. the mean number of RNAs produced per unit time (μ_Z) equals the inverse interval mean, while the Fano factor (variance over the mean) of the RNA numbers (η_Z) equals the squared coefficient of variation of the production intervals. The cell phenotype is also affected by other processes, such as RNA degradation and dilution due to cell division. Regardless, the mean and noise of the produced RNA numbers are directly linked to the phenotype (details in supplement) [38], so we expect our results to hold qualitatively in the presence of other processes.

2.2 Cells, plasmids, chemicals, and growth conditions

We used *E. coli* strain BW25113 ($lacI^+$ $rrnB_{\text{T14}}$ $\Delta lacZ_{\text{WJ16}}$ *hsdR514* $\Delta araBAD_{\text{AH33}}$ $\Delta rhaBAD_{\text{LD78}}$) [39], which contains the constitutive promoters $\text{P}_{\text{lacI+}}$ and P_{araC} producing, respectively, LacI repressors [40] and AraC repressors. As this strain does not contain the *tetR* gene responsible for encoding TetR repressors, any gene downstream to a P_{tetA} promoter is expressed constitutively.

We constructed five target systems on a single-copy pBELO plasmid. The first plasmid features the $\text{P}_{\text{lacO3O1}}$ promoter controlling the production of an RNA molecule coding for a red fluorescent mCherry protein followed by 48 binding sites for the MS2-GFP protein (mCherry-48BS). The other four systems are modified versions of the first, with the $\text{P}_{\text{lacO3O1}}$ promoter being replaced by the following promoters: (i) P_{BAD} promoter; (ii) $\text{P}_{\text{lacO3O1-tetA}}$ dual-tandem promoter; (iii) $\text{P}_{\text{lacO3O1-BAD}}$ dual-tandem promoter; and (iv) $\text{P}_{\text{lacO3O1-lacO3O1}}$ dual-bidirectional promoter. All strains aside from its target system also contain either a medium-copy plasmid pZA25 with the reporter gene P_{ara}-MS2-GFP or a low-copy plasmid pZS12 with the reporter gene P_{lac}-MS2-GFP. These plasmids are responsible for producing the fusion protein MS2-GFP, both producing an abundance of MS2-GFP when activated as detailed below. The reporter plasmids were generously provided by Orna Amster-Choder (Hebrew University of Jerusalem, Israel) [41], and Philippe Cluzel (Harvard University, USA) [42], respectively. The activity of the promoters $\text{P}_{\text{lacO3O1}}$, $\text{P}_{\text{lacO3O1-tetA}}$, and $\text{P}_{\text{lacO3O1-lacO3O1}}$ is regulated by the repressor LacI and the inducer isopropyl β-D-1-thiogalactopyranoside (IPTG). Meanwhile, the activity of P_{BAD} is regulated by the repressor AraC and the inducer L-arabinose. Finally, the activity of $\text{P}_{\text{lacO3O1-BAD}}$ is regulated by both repressors (LacI and AraC) and both inducers (IPTG and L-arabinose).

Cells were grown overnight in lysogeny broth (LB) medium supplemented with appropriate antibiotics (34 μg/ml of chloramphenicol, 50 μg/ml of ampicillin, and 50 μg/ml of kanamycin) with shaking at 250 rpm.

We made subcultures, by diluting the stationary-phase culture into fresh M9 medium supplemented with glycerol (0.4% final concentration) and the appropriate antibiotics. Cells were left in the incubator until reaching OD_{600} of about 0.25. For the pZA25-P_{ara}-MS2-GFP reporter plasmid activation, 0.4% of L-arabinose was added to the culture, which was then incubated at 37 °C for 60 minutes. Cells containing the pZS12-P_{lac}-MS2-GFP reporter plasmid were incubated in the same way and were activated with 1 mM IPTG. Next, for the activation of $P_{lacO3O1}$, $P_{lacO3O1\text{-}tetA}$, and $P_{lacO3O1\text{-}lacO3O1}$ target plasmids, specific concentrations of IPTG (either 5 µM or 1 mM) were added to the culture. For activating the P_{BAD} or $P_{lacO3O1\text{-}BAD}$ target plasmids, 0.1% of L-arabinose was added. For the latter, similar concentrations of IPTG (5 µM or 1 mM) were added as well. Inducer-activated cells were then left in the incubator for 90 minutes, prior to microscopy observation.

2.3 Microscopy and image analysis

Cells were visualized using a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with a 100× Apo TIRF (1.49 NA, oil) objective. Cells and fluorescent spots within were imaged by Highly Inclined and Laminated Optical sheet (HILO) microscopy, using an EMCCD camera (iXon3 897, Andor Technology), a 488 nm argon laser (Melles-Griot), and an emission filter (HQ514/30, Nikon). Phase-contrast images were acquired by a CCD camera (DS-Fi2, Nikon). The software for image acquisition was NIS-Elements (Nikon, Japan). An example of each channel is shown in Figure 1.

We performed time-lapse fluorescence and phase-contrast imaging of the cells (the latter for cell segmentation and lineage construction). For this, 8 µl of cells were placed on a microscope slide between a coverslip and a M9 glycerol agarose gel pad. During image acquisition, cells were constantly supplied with fresh media containing IPTG and L-arabinose, at the same concentration as when in liquid culture, by a micro-perfusion peristaltic pump (Bioptechs) at 0.3 ml/minute. Images were captured for 5 hours, once per minute in the case of fluorescence and once per 5 minutes in the case of phase-contrast. During image acquisition, cells were kept in a temperature-controlled chamber (FCS2, Bioptechs) at optimal temperature (37 °C).

Time series microscopy images were processed as in [43] by, first, aligning consecutive images so as to maximize the cross-correlation of fluorescence intensities. Next, we annotated manually the region occupied by each cell in the time series. Afterwards, the location, dimension, and orientation of each cell in each frame is obtained by principal component analysis, assuming that fluorescence inside the cell is uniform [44]. Cell lineages were then extracted using CellAging, based on overlapping areas in consecutive frames [44]. Next, the intensity of each cell is fit with a surface (quadratic polynomial of the distance from the cell border) in least-deviations sense [45]. This surface represents the cellular background intensity which is subtracted to obtain the foreground intensity. Next, the foreground intensity is fit with a set of Gaussian surfaces, in least-deviations sense, with decreasing heights until the heights are in the 99% confidence interval of the background noise (estimated assuming a normal distribution and using median absolute deviation) [45]. The Gaussians represent fluorescent RNA spots, and the volume under each represent the total spot intensity. Finally, as MS2-GFP-tagged RNA lifetimes are much longer than cell division times [46], the cellular foreground intensity will be an increasing curve, with each jump corresponding to the appearance of a novel tagged RNA. The moments when a jump occurs are estimated using a specialized curve fitting algorithm [27]. The intervals between jumps in individual cells correspond to time intervals between consecutive RNA production events.

3 Results and discussion

3.1 Analytical distributions of production time intervals

From the perspective of the production kinetics of X alone, the reaction system of Equation (1) is equivalent to:

$$I_2 \underset{k_2}{\overset{k_4}{\rightleftharpoons}} P_0 \xrightarrow{k_1} I_1 \xrightarrow{k_3} P_0 + X \tag{4}$$

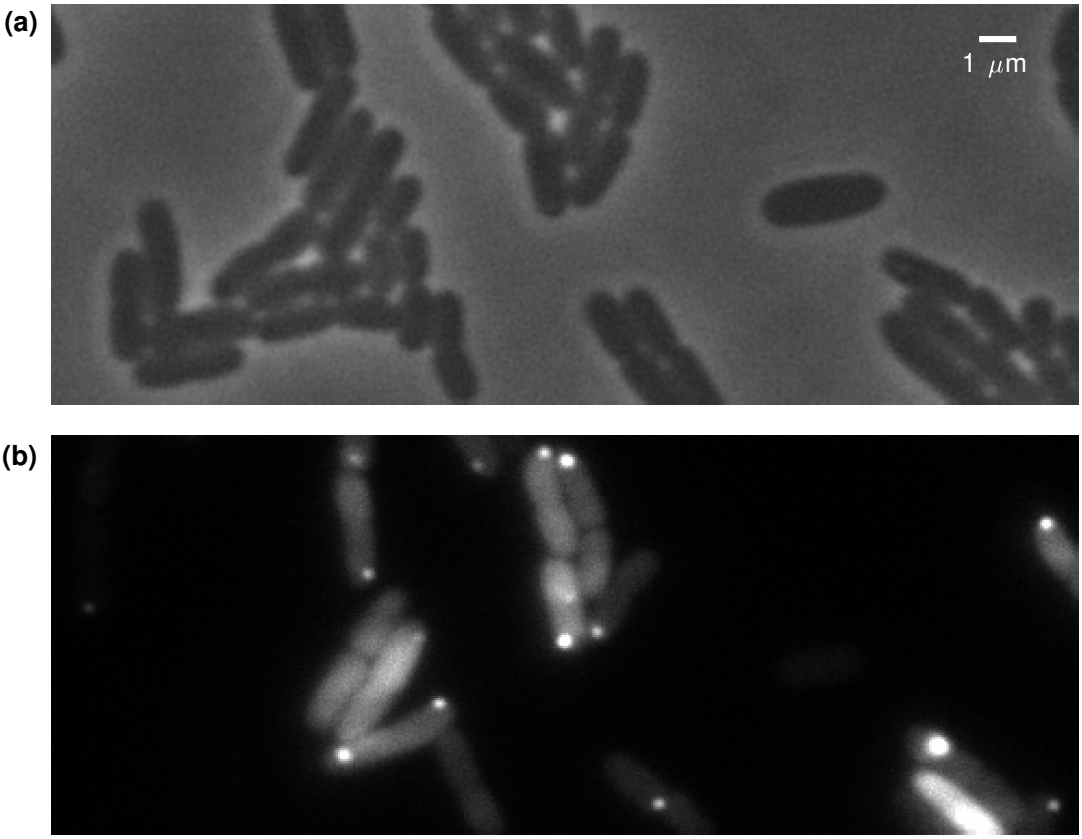

Figure 1: Example images of live *E. coli* expressing GFP-tagged RNAs. (a) Phase contrast image of the live *E. coli* with the $P_{lacO3O1\text{-}tetA}$ construct taken after 1 hour of induction with 1 mM IPTG induction at 37 °C. (b) HILO image visualizing the abundant GFP inside the same *E. coli* cells and the target RNA bound by an array of GFPs appearing as bright spots.

which is potentially a highly noisy process [47, 29]. While the expression of gene 1 might not be noisy on its own, its expression is perturbed by the transcription machinery occupying the shared promoter region for expression of gene 2, introducing (random) temporal gaps in the expression.

Let $\mathcal{G}(\cdot)$ denote the distribution of consecutive productions of X in Equation (4). The mean and variance of the time intervals between the productions of X are given by [29]:

$$\begin{aligned} \tau_X &\sim \mathcal{G}(k_4, k_2, k_1, k_3) \\ \mathbb{E}[\tau_X] &= \left(1 + \tfrac{k_2}{k_4}\right) {k_1}^{-1} + {k_3}^{-1} \\ \mathbb{V}\mathrm{ar}[\tau_X] &= \left(\left(1 + \tfrac{k_2}{k_4}\right)^2 + 2\,\tfrac{k_1}{k_4}\,\tfrac{k_2}{k_4}\right) {k_1}^{-2} + {k_3}^{-2} \end{aligned} \tag{5}$$

while, due to the symmetry of the model, the production intervals of Y are $\tau_Y \sim \mathcal{G}(k_3, k_1, k_2, k_4)$.

By comparing Equation (2) with Equation (5) we find that, regardless of the parameters, in a bidirectional configuration, the mean and variance of the time intervals between RNA productions of each gene are increased. Consequently, while ${\tau_X}^{(1)}$ is always sub-Poissonian, τ_X can exhibit either sub- or super-Poissonian behavior.

RNA production according to the model is exemplified in Figure 2 (b), and the expected interval distribution in Figure 2 (c). While the production intervals of each gene are often somewhat regular, as indicated by the bulk of the distribution, large outliers are present due to the temporal gaps, which coincide with the transcriptional activity of the other gene (see Figure 2 (b)).

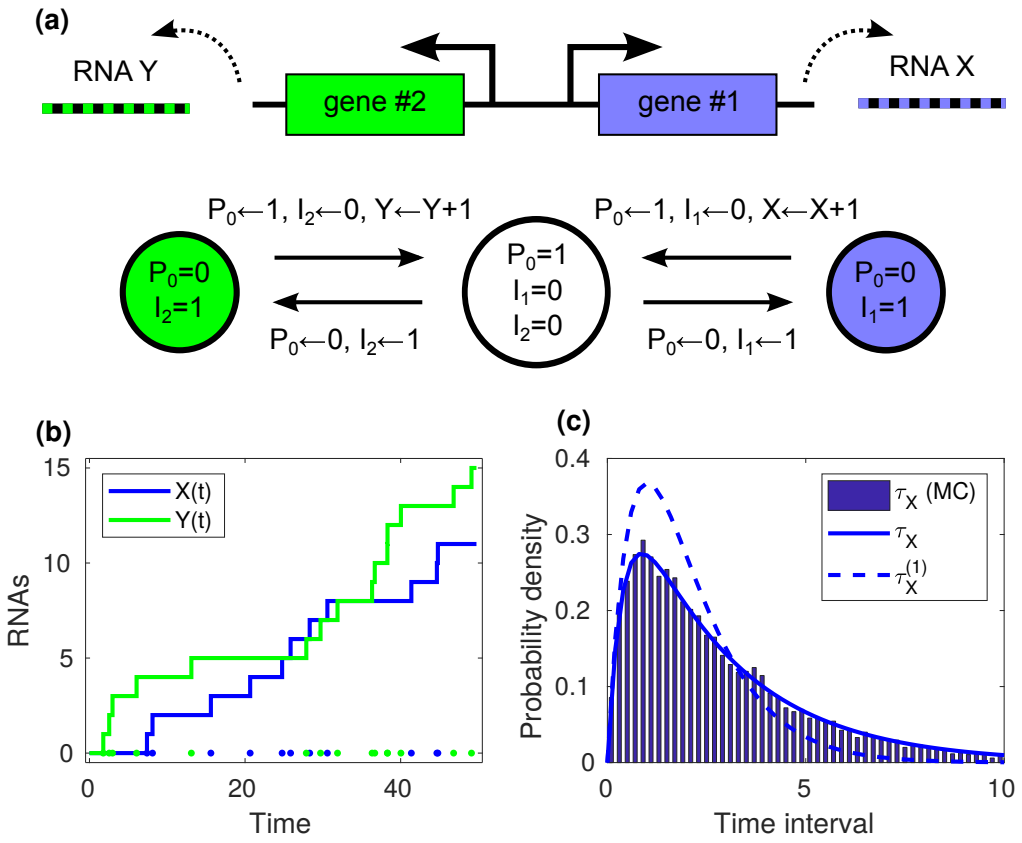

Figure 2: Model schematic and simulated examples. (a) Schematic of gene-pair in a head-to-head configuration: genes 1 and 2 produce RNAs X and Y, respectively. The shared promoter can be in three-states: free, or occupied for transcription initiation of gene 1 or 2. (b) Produced RNA numbers over time in a single Monte Carlo simulation. The dots denote the moments when RNAs were produced. (c) Distribution of intervals between consecutive productions of X in 10,000 simulations. The parameter values are $(k_1, k_2, k_3, k_4) = (1, 1, 1, 1)$. Here, τ_X and $\tau_X{}^{(1)}$ have a mean (variance) of 3 (7) and 2 (2), respectively.

As the marginals shown in Equation (5) fail to capture the co-expression of the two genes, further analysis is necessary. The time between consecutive productions by either gene, i.e. a jump in $X(t)+Y(t)$, is (detailed in the supplement):

$$\begin{aligned} \tau_{X+Y} &\sim \mathcal{E}(k_1+k_2) + \mathcal{E}^+\left(\tfrac{k_1}{k_1+k_2}, k_3, k_4\right) \\ \mathbb{E}[\,\tau_{X+Y}\,] &= \left(1+\tfrac{k_1}{k_3}+\tfrac{k_2}{k_4}\right)(k_1+k_2)^{-1} \\ \mathbb{V}\mathrm{ar}[\,\tau_{X+Y}\,] &= \left(1+\left(\tfrac{k_1}{k_3}-\tfrac{k_2}{k_4}\right)^2+2\,\tfrac{k_1}{k_3}\,\tfrac{k_2}{k_3}+2\,\tfrac{k_1}{k_4}\,\tfrac{k_2}{k_4}\right)(k_1+k_2)^{-2} \end{aligned} \tag{6}$$

where $\mathcal{E}(\lambda)$ is an exponential distribution with rate λ and $\mathcal{E}^+(p_1, \cdots, p_{n-1}, \lambda_1, \cdots, \lambda_n)$ is a hyperexponential distribution with mixing probabilities $p_1, \cdots, p_n$ and rates $\lambda_1, \cdots, \lambda_n$. Again, this distribution can feature

either sub- or super-poissonian behavior, depending on its parameter values. By combining Equation (3), (5), and (6), one can determine the asymptotic covariance and the (Pearson) correlation ρ_{XY} between the produced RNA numbers $X(t)$ and $Y(t)$ (detailed in the supplement).

3.2 Noise and correlation in the transcription kinetics of genes in a head-to-head configuration

Based on the above, we first analyzed how the noise and correlation in the transcription kinetics of a head-to-head configuration depends on the dynamics of the individual genes. For this, the parameterization λ, q_{12}, q_{13}, q_{24} was found to be insightful. Here, $\lambda \doteq \mu_{X+Y}$ is a timescale parameter (mean total production rate) and $q_{ij} \doteq k_i \,/\, k_j$ denote ratios of rates of two reactions. Further, q_{12} controls the bias, i.e. the expression ratio of each gene: for large (small) q_{12}, gene 1 (gene 2) is expressed more frequently. Finally, q_{13} and q_{24} control the relative durations of closed and open complex formation, which equal $1 \,/\, (1 + q_{13})$ and $q_{13} \,/\, (1 + q_{13})$, respectively, for gene 1. Specifically, if $q_{13} > 1$ ($q_{13} < 1$), then $k_1 > k_3$ and the gene is limited at the open (closed) complex formation.

The mean RNA numbers are controlled by the bias and the scale: $\mu_X = \lambda^{-1}\, q_{12} \,/\, (1 + q_{12})$ and $\mu_Y = \lambda^{-1} \,/\, (1 + q_{12})$. As such, the stage at which the transcription kinetics of each gene is rate-limited does not affect the mean number of produced RNAs. Meanwhile, the noise and correlation exhibit complex behavior, which can be divided into a few regions. The regions and their properties are shown in Table 1 (and Table S1). The noise of each gene and the correlation coefficient are shown in Figure 3 and Figure 4 (a), and their analytical forms in the supplement.

Table 1: Noise and correlation in RNA production kinetics in the different regions of the parameter space of head-to-head configuration. Here, $\sim 1^-$ ($\sim 1^+$) denotes weakly sub- (super-) Poissonian behavior (noise of about 1), while ~ 1 denotes that both behaviors are possible. Finally, $< 1^*$ indicates that < 1 holds at least for one of the genes, possibly for both.

Region	Condition	Noise η_X	Noise η_Y	Correlation ρ_{XY}
A	$q_{24} > 1$, $q_{24} > q_{13}$	> 1	$\sim 1^-$	> 0
A	$q_{13} > q_{24}$, $q_{13} > 1$	$\sim 1^-$	> 1	> 0
B	q_{13}, $q_{24} < 1$	~ 1	~ 1	$\sim 0^-$
C	$q_{13} \sim q_{24} > 1$	$< 1^*$	$< 1^*$	< 0
D	$q_{13} \sim 1$, $q_{24} < 1$	< 1	$\sim 1^+$	< 0
D	$q_{13} < 1$, $q_{24} \sim 1$	$\sim 1^+$	< 1	< 0
E	$q_{13} \sim q_{24} \sim 1$	$< 1^*$	$< 1^*$	< 0

Region A: For $q_{24} > 1$, $q_{24} > q_{13}$, the expression of gene 2 is most limited at the open complex formation, while that of gene 1 is more symmetric. As such, the promoter region is mostly occupied, and gene 1 must express either fast or rarely. In the former case, there is a burst of production of proteins X after each Y, so the expression of the two genes is positively correlated, and while gene 2 is Poissonian, solely controlled by its open complex formation process, gene 1 is highly noisy as the geometric burst of RNA is separated by the gaps created by the other gene. In the latter case, the expression of gene 1 is controlled by uniform random productions and the correlation vanishes. Specifically, in the latter case, the noise of gene 1 is $1{+}2\, q_{12}$ (super-Poissonian) and gene 2 is Poissonian. The correlation for large q_{12} is $\sqrt{1/2}$, which is maximal for the configuration, while for small q_{12} the correlation vanishes. The part $q_{13} > q_{24}$, $q_{13} > 1$ is symmetric. Note that the bias q_{12} controls the upper bounds for noise and correlation.

Region B: Here, both genes are limited at the closed complex formation. Thus, the promoter region is rarely occupied, as the expression is limited by an RNAP finding the gene and initiating transcription. This causes the expression of both genes to be Poissonian, as each is limited by a single step, and uncorrelated,

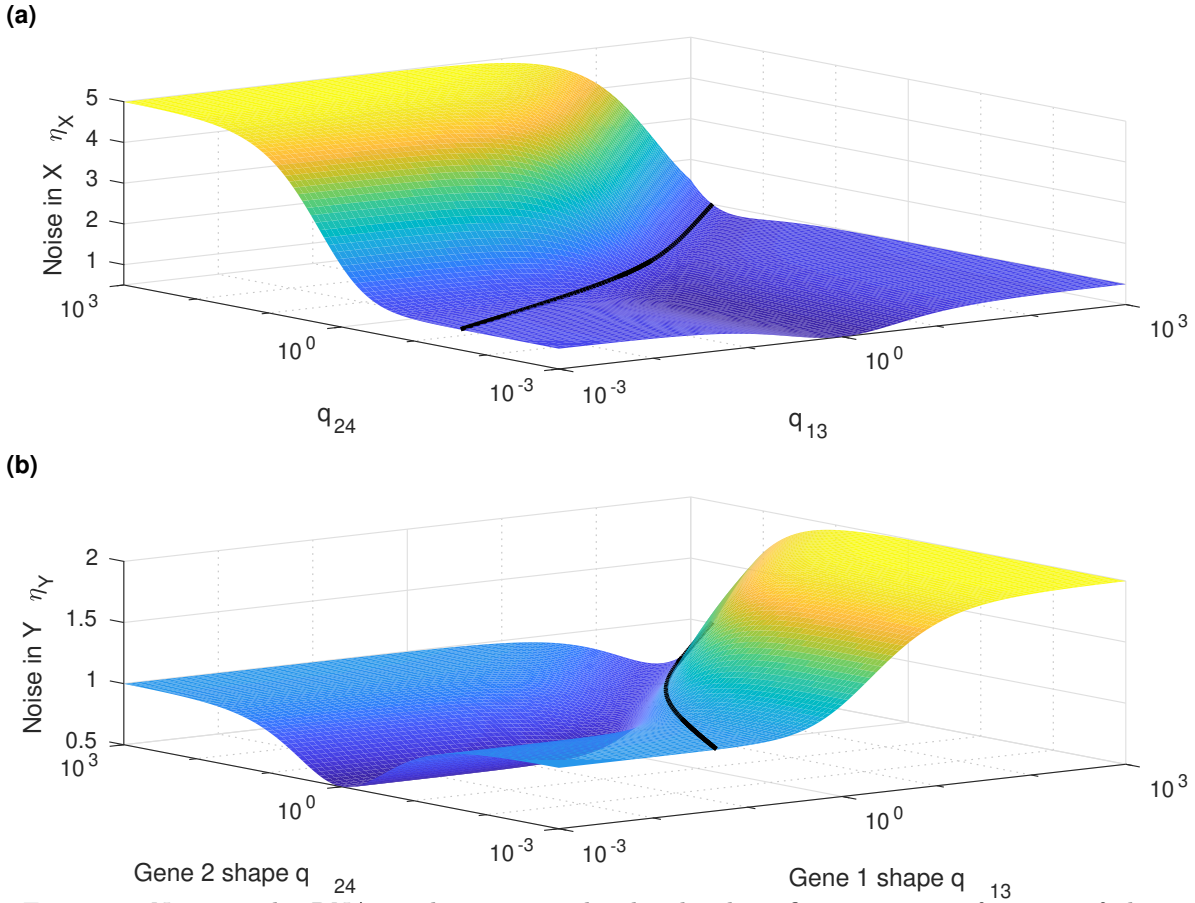

Figure 3: Noise in the RNA production in a head-to-head configuration as a function of their relative durations of closed and open complex formations. (a) gene 1 and (b) gene 2. The black curves denote unity, and $q_{12}=2$.

as their activities do not interfere at the promoter region.

Region C: Both genes are limited at the open complex formation, which makes them to alternate in occupying the shared promoter region. The noise is set by the bias q_{12}, which determines the gene more disturbed by the activity of the other. Specifically, the noise of gene 1 equals $1/2+q_{12}/2$ and the noise of gene 2 equals $1/2+{q_{12}}^{-1}/2$. As the genes inhibit each other by competing for the shared promoter region, the expression patterns are anticorrelated.

Region D: For $q_{13}\sim 1$, $q_{24}<1$, gene 2 is limited during the closed complex formation, so it does not block the shared promoter area. Meanwhile, gene 1 is limited at both stages, making its RNA production to be sub-Poissonian. The expression of gene 2, originally Poissonian, becomes affected by periods of inactivity as gene 1 employs the promoter, increasing the noise, as controlled by the bias, yielding noise of $1+{q_{12}}^{-1}/2$. The correlation is negative, as gene 1 inhibits the expression of gene 2. The part $q_{13}<1$, $q_{24}\sim 1$ is symmetric.

Region E: Both genes have similar closed and open complex formation durations, resulting in low noise in a non-bidirectional configuration. If their closed complex formation durations are similar (i.e. $q_{12}\sim 1$), both genes are of low noise ($\sim 7/9$) and their expression is anticorrelated ($\sim -2/7$), as they alternate in activity. Otherwise, one is of low noise ($\sim 5/9$), unaffected by the configuration, while the other is of high noise, with its expression being disturbed by the frequent gaps caused by the other. Specifically, the noise is $5/9+2\,q_{12}/9$ for gene 1 and $5/9+2\,{q_{12}}^{-1}/9$ for gene 2. The correlation is negative, with a maximum of

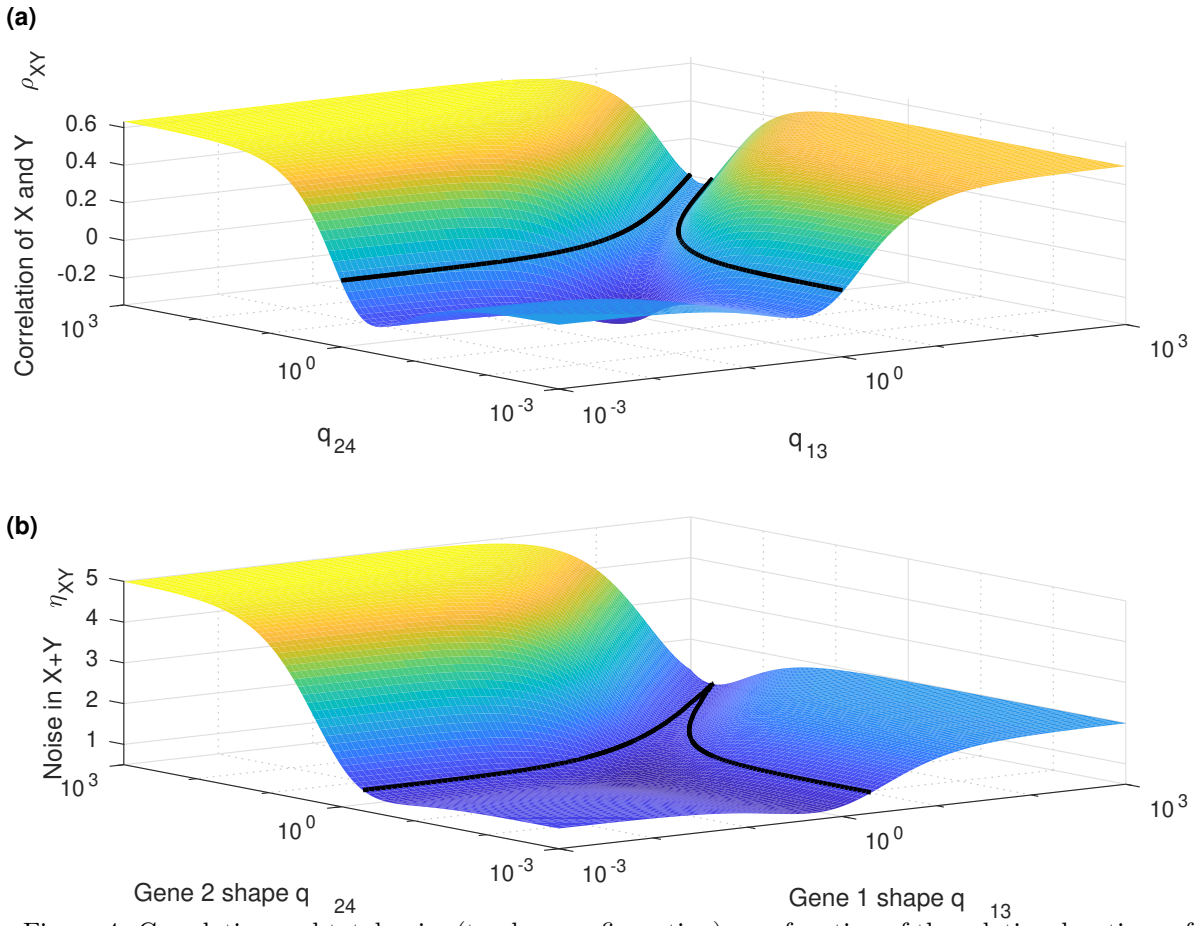

Figure 4: Correlation and total noise (tandem configuration) as a function of the relative durations of closed and open complex formation. (a) Correlation between the RNA production kinetics of two head-to-head genes. (b) Noise of RNA production of a gene with two initiation sites. The black curves denote zero or unity, and $q_{12}=2$.

$-2/7$ at $q_{12}=1$, and minima of $-1/\sqrt{10}$ at $q_{12}\to 0$ and $q_{12}\to\infty$.

In summary, for coupled gene activity, one (or both) genes must not be limited at the closed complex formation alone. When coupled, both genes are low noise only if both feature similar relative closed-to-open complex durations. In this case, their expression is likely anticorrelated, which may have further implications if the genes are involved in regulating the same or complementary processes. If the relative closed-to-open complex durations differ, one is of high noise and the other of low noise, while their expression is, surprisingly, positively correlated. While our analysis lacks processes following the transcription initiation, the presence of e.g. first-order degradation pulls the noise toward unity and the correlation toward zero, leaving the conclusions qualitative useful.

3.3 Noise in a gene with two initiation sites

Next, we consider the dynamics of a common RNA product controlled by a promoter with two TSSs (see Figure S1 (c)). This is common in *E. coli* [48] and more so in, e.g. plant mitochondria [49]. The configuration is readily accommodated by our model, by considering the dynamics of $X+Y$. As the mean and variance of $X+Y$ follow the mean and covariance of (X,Y), the results can be derived from those obtained in the

previous section.

Figure 4 (b) shows the noise for $X + Y$, representing the RNAs produced through either TSS. The noise is low only if both TSSs exhibit production dynamics with low noise, i.e. in the regions C, D, or E. Compared to individual TSSs, the RNA number fluctuations are lower, being suppressed by the negative correlation. If one TSS exhibits highly noisy production (region A), the RNA numbers become highly noisy, regardless of the dynamics of the other TSS. Finally, in region B, the production is exponential-like, as multiple TSSs only increase the RNAP to promoter binding affinity, which makes their dynamics indiscernible from that of a single TSS. This suggests that low noise tandem architectures exert selection pressure on both promoter components.

3.4 Model predictions for empirical validation

To validate our predictions, we observed transcription in live *E. coli* at the single-RNA level in various constructs. Three of the constructs feature synthetic genes whose production is controlled by a single promoter (specifically $P_{lacO3O1}$, P_{tetA}, and P_{BAD} (see Figure S1 (a)). The remaining constructs feature pairs of genes sharing promoter elements. One of these constructs is $P_{lacO3O1\text{-}lacO3O1}$, with overlapping lacO3O1 promoters in the opposite strands (see Figure S1 (b)), with the reporter being on a single side. In the other two constructs, the expression is controlled by a $P_{lacO3O1\text{-}tetA}$ or a $P_{lacO3O1\text{-}BAD}$ dual-tandem promoters (see Figure S1 (c)). In all these, the expression of the lacO3O1 promoter is modulated by the IPTG concentration, an inducer for the lac promoter[50]. Meanwhile, aTc concentration is held constant at 15 ng/ml, in order to trigger full expression of the tetA promoter. Similarly, L-arabinose concentration is held constant at 0.1%, to trigger full expression of the BAD promoter. In all cases, RNA production dynamics was measured by time-lapse microscopy imaging using MS2-GFP tagging (see methods).

Using our models, we aim to predict the behavior of the gene pairs sharing promoter elements, given knowledge of the behavior of the constituent genes not involved in gene pair interactions (i.e. operating as isolated promoters). More specifically, we test whether, from the measured dynamics of RNA production of $P_{lacO3O1}$, P_{tetA} [26], and P_{BAD}, one can predict the kinetics of $P_{lacO3O1\text{-}lacO3O1}$, $P_{lacO3O1\text{-}tetA}$, and $P_{lacO3O1\text{-}BAD}$.

For this, we first extracted the number of RNAs in each cell in the first and the last frame of the time series for all the constructs in each condition (see Figure 5 (b) and (c), respectively). These data were used to estimate the mean and standard deviation the production intervals, and the most likely (maximum likelihood fit) model of Equation (S1) for the single promoters, through the relations of Equation (S6). The estimated intervals are shown in Table 2, along with the model parameters of Equation (S1) where applicable. A Wald test testing for a specific mean and standard deviation was used to compute a p-value to confirm that the model predicts the mean and variance of the RNA distributions. The measurement data and the resulting model fits are exemplified for the $P_{lacO3O1\text{-}tetA}$ construct induced with 1000 µM of IPTG in Figure 5. We also extracted the intervals from the full time series for several of the constructs (about 120 frames, one every minute) to verify that the production time intervals can be correctly estimated from the RNA distributions (see Table S2 and Figure 5 (a)). The RNA counts and the time intervals are given in ESM file 2 and 3, respectively, and constitute the raw measurement data after the image analysis, as opposed to the model fits show in Table 2, Table 3, and Table 4.

The results in Table 2 indicate that changing IPTG concentration alters the noise of the lacO3O1 promoter in addition to changing its mean expression rate, which is expected to be due to changes in the open-to-closed complex duration ratio, and is in agreement with previous reports [28]. The p-values indicate that there is no evidence that any of the models fit the measurements poorly.

Next, using the above parameters (i.e. k_1 and k_3 in Table 2), we constructed the models for the dual promoters through Equation (5) and Equation (6). The obtained models are shown in Table 3. The predicted mean and standard deviation show an agreement with the measured behavior of the dual promoter constructs, while the noise and correlation indicate that that promoters operate at different regions of the open-to-closed complex ratio space (these values cannot be directly measured with our system, but can be inferred from the model fit). The results indicate that the model predicts the behavior of the dual-promoter measurements

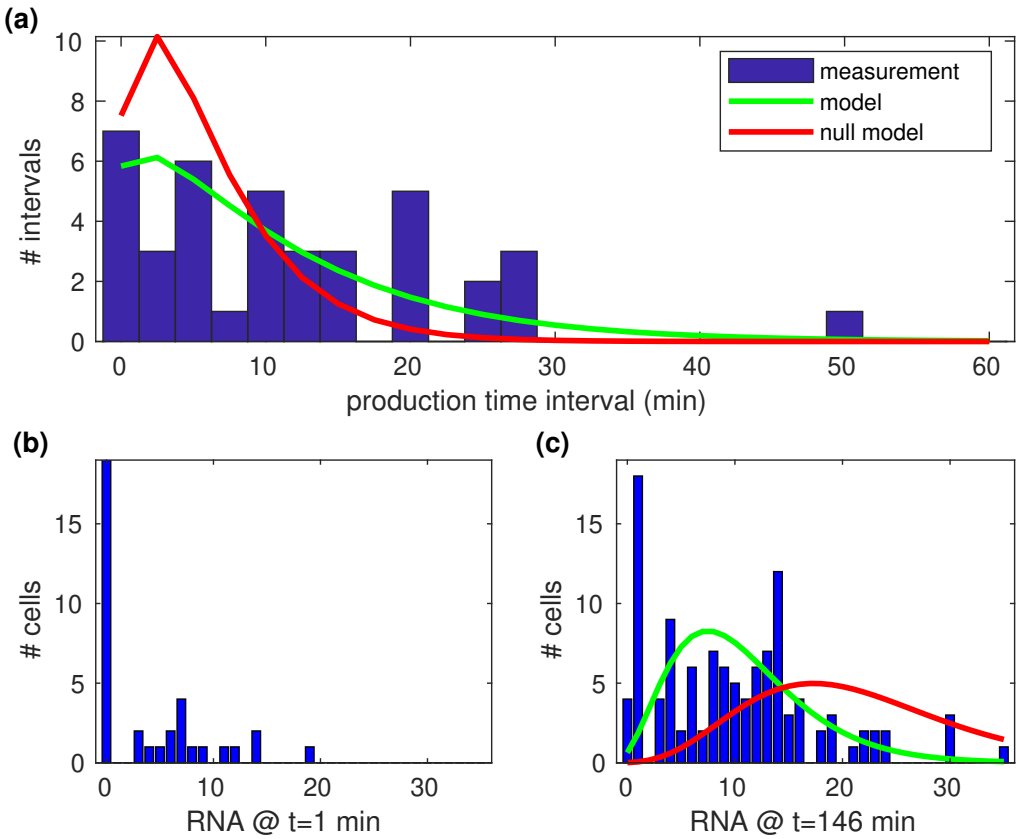

Figure 5: Measured production time intervals and RNA distributions in the first and last frame of the time series along with model predictions for the $P_{lacO3O1\text{-}tetA}$ construct induced with 1000 µM of IPTG. (a) Production time interval distribution in 39 cells and the corresponding model predictions using a model with interactions during transcription initiation (model; see Table 3) or a model with no interactions (null model; see Table 4), (b) RNA distribution in 65 cells 1 min after induction (these data are used only for predicting the subsequent RNA distribution in (c)), and (c) RNA distribution in 115 cells 146 min after induction, along with the model predictions.

well, and that the noise is modulated by the change in the coordination between the two promoters in the dual promoter construct.

As our methodology cannot identify which of the steps correspond to ${k_1}^{-1}$ and ${k_3}^{-1}$ in Table 2, we also considered the alternative step ordering. The dual-promoter model fits had a p-values $<3.964\times10^{-3}$ in for the lacO3O1-tetA construct at 5 µM IPTG, and p-values $<4.614\times10^{-3}$ for the lacO3O1-BAD construct at 1000 µM IPTG, indicating that the alternatives are not likely for lacO3O1 at 5 µM and BAD. The step order for lacO3O1 at 1000 µM IPTG and tetA cannot be resolved from these data, but the alternatives result in a qualitatively similar dual-promoter models and p-values >0.116. For the constructs containing these two promoters, we report the most likely models, all suggesting the order specified in Table 2. These findings are also supported by prior evidence using a different methodology[28].

The fact that the measurements fall into the different regions of operation (see Figure 4 and Table 1) is apparent in Figure 6, Figure 7, and Figure 8. Namely, the high IPTG condition falls into region E for the lacO3O1-lacO3O1 and lacO3O1-tetA, and into region D for the lacO3O1-BAD construct. At low IPTG, the lacO3O1-lacO3O1 transits into region C, as both promoters are modulated by the changes in the inducer

Table 2: Estimated RNA production intervals for each of the promoter constructs. The table shows the promoter, induction, estimated paramater of model Equation (S1) for the single promoters, the estimated mean, standard deviation (sd), and noise (coefficient of variation) of the RNA production intervals, and the p-value of the test of model versus data.

Promoter	IPTG (μM)	k_1^{-1} (s)	k_3^{-1} (s)	Mean (s)	Sd (s)	Noise	P-value
lacO3O1	1000	362.3	737.3	1099.6	821.5	0.558	0.446
lacO3O1	5	25.8	1236.8	1262.6	1237.0	0.960	0.273
tetA	-	287.4	385.5	672.9	480.9	0.511	0.075
BAD	-	1036.7	333.7	1370.4	1089.1	0.632	0.059
lacO3O1-tetA	1000	-	-	702.2	638.8	0.828	0.604
lacO3O1-tetA	5	-	-	1111.6	1089.1	0.960	0.164
lacO3O1-lacO3O1	1000	-	-	1659.3	1437.0	0.750	0.112
lacO3O1-lacO3O1	5	-	-	2205.9	2119.3	0.923	0.971
lacO3O1-BAD	1000	-	-	866.8	612.9	0.500	0.099
lacO3O1-BAD	5	-	-	1274.7	1248.9	0.960	0.698

Table 3: Models derived for the dual promoters from the individual promoter fits of Table 2 using the model with interactions during transcription initiation. The table shows the promoter/induction scheme, the mean and standard deviation (sd) of the RNA production intervals and the correlation between the RNA numbers assuming the derived models, and the p-value of the test of model versus data.

Promoter	IPTG (μM)	Mean (s)	Sd (s)	Noise	Correlation	P-value
lacO3O1-lacO3O1	1000	1836.9	1685.2	0.842	−0.188	0.168
lacO3O1-lacO3O1	5	2499.3	2486.5	0.990	−0.010	0.931
lacO3O1-tetA	1000	701.4	616.1	0.772	−0.166	0.603
lacO3O1-tetA	5	1190.4	1212.9	1.038	+0.172	0.123
lacO3O1-BAD	1000	901.3	731.4	0.659	−0.070	0.141
lacO3O1-BAD	5	1240.0	1230.9	0.985	+0.105	0.620

concentration, while the lacO3O1-tetA and lacO3O1-BAD transit into (opposite directions) of region A. This explains the widely different noise levels in the estimated and measured intervals (see Table 2 and Table S2), which are well predicted by our models in each case (compare with Table 3).

Finally, we verified that a model with no interactions between the two promoters would not explain the measurements. For this, we attempted to predict the mean, noise, and intervals in a dual-promoter measurement using independent expression from the constituent promoters (i.e. Equation (S4)) as predicted from the single-promoter measurements. The results in Table 4 show that the associated model fails to explain the observed dual-promoter behavior, and the apparent mismatch is exemplified in Figure 5. Note that the models are also unaffected by the (k_1, k_3) identifiability problem. While the mean and noise of the system consisting of two independent promoters trivially follow from their independent components, the time intervals of the combined production do not. In particular, the intervals are not independent. We also considered the possibility that while the promoters might have interactions, their expression levels may be altered by the other promoter utilizing the same finite pool of RNA polymerases. For this, we assumed that the number of RNA polymerases modulate the closed complex formation rate (i.e. $k_1 = R\,\tilde{k}_1$ where $\tilde{k}_1$ is the per-polymerase closed complex formation rate, and R represents an RNA polymerase), which will cause a slight reduction of the closed complex formation rate, as determined by the closed to open-to-closed complex

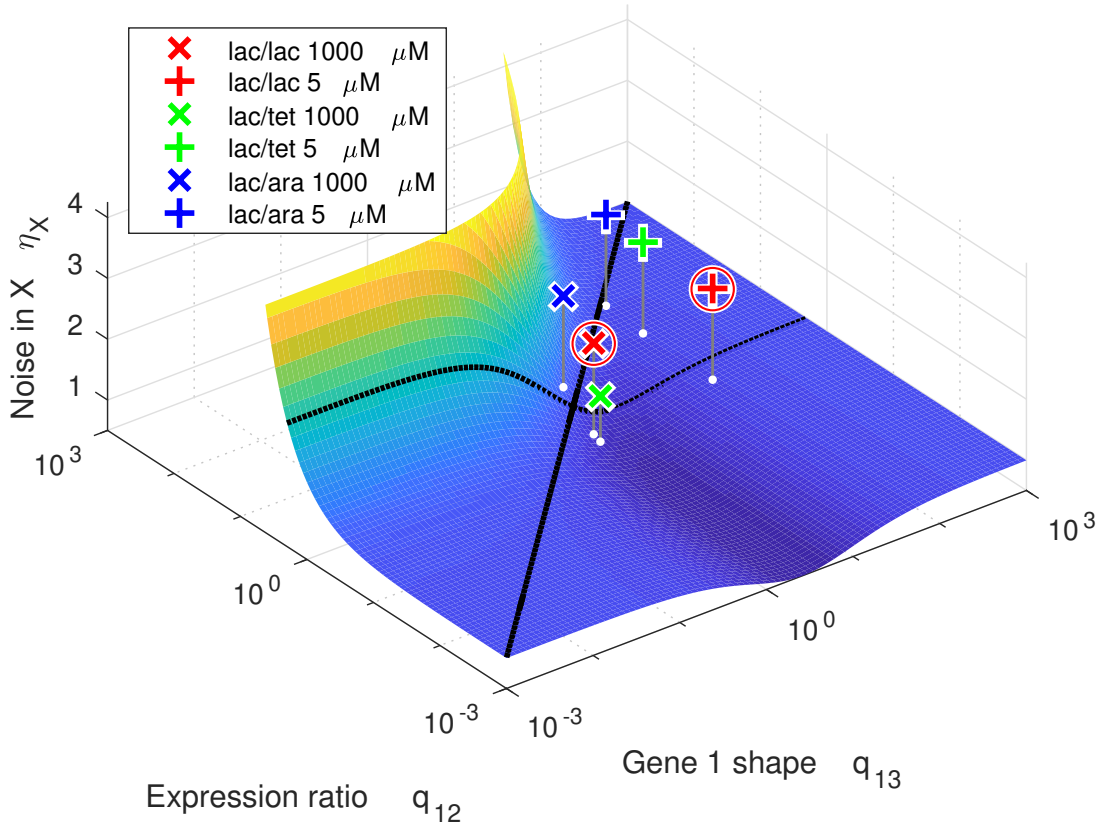

Figure 6: Noise of RNA production of one side of a dual promoter as a function of the relative durations of closed and open complex formation and the expression ratio of the promoters. The black curves denote unity and $q_{12} = 2$ (see Figure 3), and the markers the predictions for the measurements, circles representing the validated ones (see Table 3).

duration ratio of the other promoter. Any of these models (all R and all step orders) failed to explain the behavior of our dual-promoter measurements as well. The effects are most extreme for $R = 2$, but we verified that models for other R have no better fit. Our model is recovered at $R = 1$ and the independent model without polymerases is recovered at $R = \infty$.

We conclude that our model of closely-spaced promoters that assumes interactions between the promoters is the one that well predicts the measurements in each setting, for both the head-to-head and tandem constructs, while a model with no interactions cannot explain the observed measurements. Relevantly, our models reveal that the observed changes arise from changes in the coordination between the two coupled transcription start sites of our synthetic constructs.

4 Conclusion

We analyzed a stochastic model of two genes in a head-to-head configuration as a function of whether each gene is rate-limited during the closed and/or open complex formation. Compared to individual genes, in the bidirectional configuration, the transcription activity is slower and noisier in both genes, as each gene interferes with the activity of the other, allowing two genes with sub-Poissonian dynamics to exhibit super-

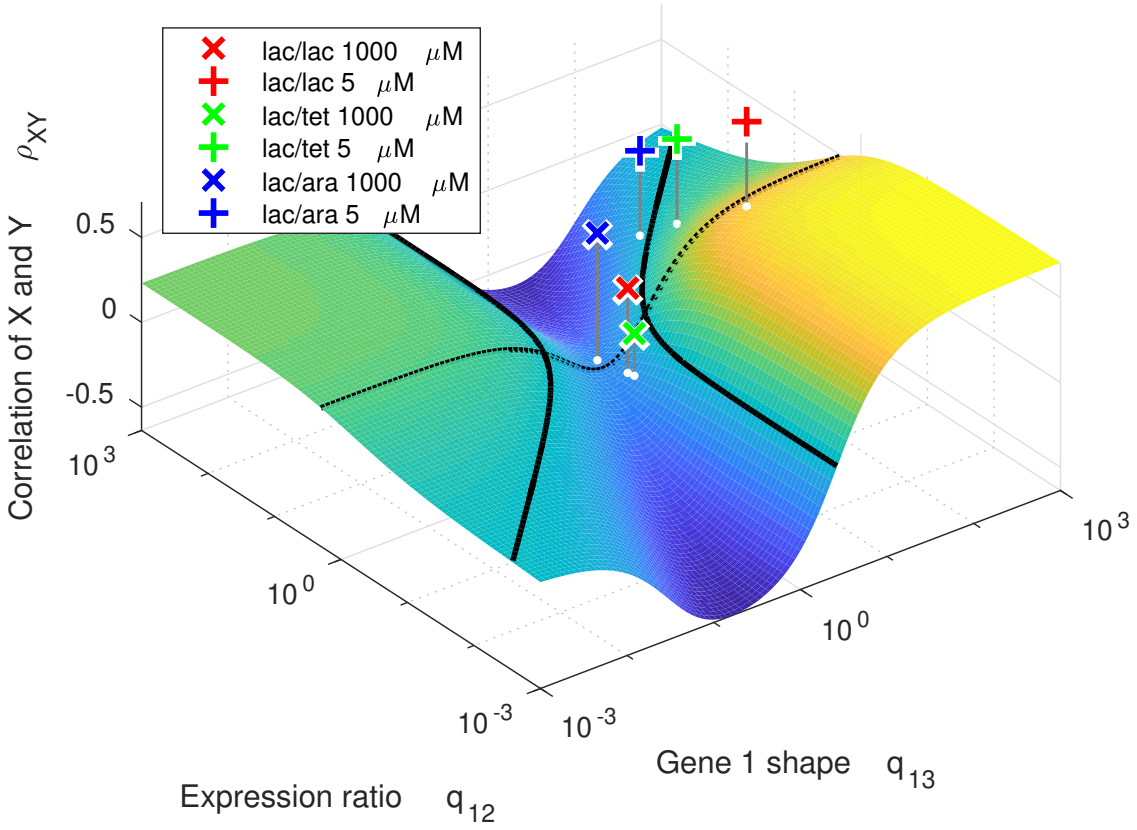

Figure 7: Predicted correlation between the RNA productions initiated by each start site of the dual promoter as a function of the relative durations of closed and open complex formation and the expression ratio of the promoters. The black curves denote zero and $q_{12} = 2$ (see Figure 4 (a)), and the markers the predictions for the measurements.

Table 4: Null models derived for the dual independent promoters from the individual promoter fits of Table 2. The table shows the promoter/induction scheme, and the mean and standard deviation (sd) of the RNA production intervals assuming the null models, and the p-value of the test of model versus data for maximally ($R = 2$) and minimally ($R = \infty$) RNA polymerase starved null models.

Promoter	IPTG (μM)	Mean (s)	Sd (s)	Noise	P-val. $R = 2$	P-val. $R = \infty$
lacO3O1-lacO3O1	1000	1099.6	821.5	0.558	4.116×10^{-4}	6.601×10^{-5}
lacO3O1-lacO3O1	5	1262.6	1237.0	0.960	9.981×10^{-3}	8.386×10^{-3}
lacO3O1-tetA	1000	417.5	303.5	0.529	2.127×10^{-3}	4.267×10^{-4}
lacO3O1-tetA	5	439.0	358.5	0.667	7.289×10^{-5}	5.049×10^{-6}
lacO3O1-BAD	1000	610.1	468.9	0.591	6.325×10^{-6}	1.456×10^{-7}
lacO3O1-BAD	5	657.1	588.7	0.803	3.464×10^{-3}	5.124×10^{-4}

Poissonian dynamics when coupled. Importantly, provided information on the kinetics of the constituent promoters when not sharing promoter elements, the models were shown to be able to predict well the behavior

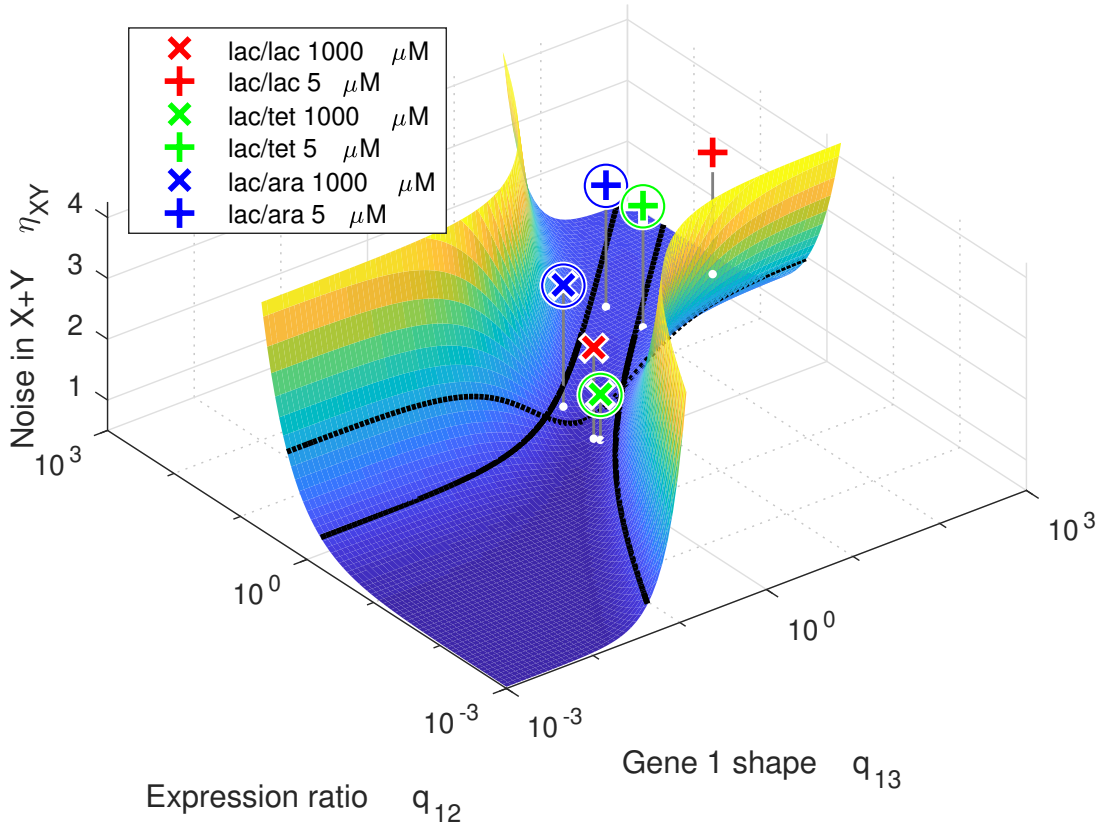

Figure 8: Noise of RNA production of the dual promoter as a function of the relative durations of closed and open complex formation and the expression ratio of the promoters. The black curves denote unity and $q_{12} = 2$ (see Figure 4 (b)), and the markers the predictions for the measurements, circles representing the validated ones (see Table 2 and Table 3).

of the pairs of the same genes when sharing promoter elements, implying that they capture accurately the effects of the complex interference caused by the shared promoter region.

We found that for such prediction to be accurate, the models have to account for the two-rate-limiting step kinetics of active transcription in *E. coli*. In particular, the time-length of such rate limiting steps, namely the closed and open complex formations, controls not only the expression rate and noise of each gene (as in isolated genes, see e.g. [24]), but also the kinetics of the temporal gaps caused by the transcription events of the opposite gene. This programs the behavior intricately: a similar rate-limiting step structure combined with a rate-limiting open complex formation is required for both genes to feature low noise; otherwise one tends to be highly noisy. Also, orderly systems tend to exhibit strong negative correlation, while the genes alternate expression, but the correlation can be lost or become positive if the open-to-closed complex formation time-lengths are incompatible. As such, not only the mean and variance of the durations of each stage, but also the mechanistic underpinnings, affect the dynamics of closely-spaced gene-pairs, implying that promoters with seemingly identical dynamics in isolation may differ widely in their dynamics in a closely-spaced configuration. Relevantly, as shown, the results generalize to the behavior of individual genes with multiple transcription initiation sites.

Overall, these results suggest that, in *E. coli*, the kinetics of the rate-limiting steps in active transcription needs to be considered for dissecting the dynamics of pairs of genes sharing promoter elements. In this

regard, we find it to be striking that pairs of closely-spaced promoters, by tuning the kinetics of their closed and open complex formation (which are sequence dependent and, thus, evolvable) tunes the orderliness of the whole gene-pair. This new knowledge provides an important route to follow in the engineering of pairs of closely-spaced promoters with desired dynamics and contributes to a better understanding of the dynamics of natural pairs of closely spaced genes and their potential role in the gene expression programs of *E. coli*.

Author contributions

AH and ASR conceived the study; AH devised the models and analysis; SO and RNV performed the experiments; AH and SO analyzed the data; AH and ASR drafted the manuscript; All authors contributed in writing the manuscript and have approved the final version.

Data accessibility

The datasets supporting this article have been uploaded as part of the supplementary material.

Competing interests

The authors declare that they have no conflict of interest.

Funding Statement

This work was supported by the Alfred Kordelin Foundation (AH); the Vilho, Yrjo and Kalle Vaisala Foundation (SMDO); the Tampere University of Technology President's Graduate Programme (RN-V); the Jane and Aatos Erkko Foundation [grant number 610536] (ASR); and Academy of Finland [grant numbers 295027 and 305342] (ASR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

[1] Adachi, N. and Lieber, M. R. (2002) Bidirectional gene organization: a common architectural feature of the human genome. *Cell,* **109**(7), 807–809.

[2] Trinklein, N. D., Force Aldred, S., Hartman, S. J., Schroeder, D. I., Otillar, R. P., and Myers, R. M. (2004) An abundance of bidirectional promoters in the human genome. *Genome Res.,* **14**(1), 62–66.

[3] Herbert, M., Kolb, A., and Buc, H. (1986) Overlapping promoters and their control in *Escherichia coli*: The *gal* case. *Proc. Natl. Acad. Sci. U.S.A,* **83**(9), 2807–2811.

[4] Moss Bendtsen, K., Erdossy, J., Csiszovszki, Z., Lo Svenningsen, S., Sneppen, K., Krishna, S., and Semsey, S. (2011) Direct and indirect effects in the regulation of overlapping promoters. *Nucl. Acids Res.,* **39**(16), 6879–6885.

[5] Korbel, J. O., Jensen, L. J., von Mering, C., and Bork, P. (2004) Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat. Biotechnol.,* **22**(7), 911–917.

[6] Prescott, E. M. and Proudfoot, N. J. (2002) Transcriptional collision between convergent genes in budding yeast. *Proc. Natl. Acad. Sci. U.S.A.,* **99**(13), 8796–8801.

[7] Callen, B. P., Shearwin, K. E., and Egan, J. B. (2004) Transcriptional interference between convergent promoters caused by elongation over the promoter. *Mol. Cell,* **14**(5), 647–656.

[8] Palmer, A. C., Ahlgren-Berg, A., Egan, J. B., Dodd, I. B., and Shearwin, K. E. (2009) Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter. *Mol. Cell,* **34**(5), 545–555.

[9] Hakkinen, A., Healy, S., Jacobs, H. T., and Ribeiro, A. S. (2011) Genome wide study of NF-Y type CCAAT boxes in unidirectional and bidirectional promoters in human and mouse. *J. Theor. Biol.,* **281**(1), 74–83.

[10] Zanotto, E., Hakkinen, A., Teku, G., Shen, B., Ribeiro, A. S., and Jacobs, H. T. (2009) NF-Y influences directionality of transcription from the bidirectional *Mrps12/Sarsm* promoter in both mouse and human cells. *BBA Gene Regul. Mech.,* **1789**(5), 432–442.

[11] Martins, L., Makela, J., Hakkinen, A., Kandhavelu, M., Yli-Harja, O., Fonseca, J. M., and Ribeiro, A. S. (2012) Dynamics of transcription of closely spaced promoters in *Escherichia coli*, one event at a time. *J. Theor. Biol.,* **301**, 83–94.

[12] Sneppen, K., Dodd, I. B., Shearwin, K. E., Palmer, A. C., Schubert, R. A., Callen, B. P., and Egan, J. B. (2005) A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli. J. Mol. Biol.,* **346**(2), 399–409.

[13] Yan, C., Wu, S., Pocetti, C., and Bai, L. (2015) Regulation of cell-to-cell variability in divergent gene expression. *Nat. Commun.,* **7**, 11099.

[14] Limi, S., Zhao, Y., Guo, P., Lopez-Jones, M., Zheng, D., Singer, R. H., Skoultchi, A. I., and Cvekl, A. (2019) Bidirectional analysis of Cryba4-Crybb1 nascent transcription and nuclear accumulation of Crybb3 mRNAs in lens fibers. *Invest. Ophthalmol. Vis. Sci.,* **60**(1), 234–244.

[15] Behjati Ardakani, F., Kattler, K., Nordstrom, K., Gasparoni, N., Gasparoni, G., Fuchs, S., Sinha, A., Barann, M., Ebert, P., Fischer, J., Hutter, B., Zipprich, G., Imbusch, C. D., Felder, B., Eils, J., Brors, B., Lengauer, T., Manke, T., Rosenstiel, P., Walter, J., and Schulz, M. H. (2018) Integrative analysis of single-cell expression data reveals distinct regulatory states in bidirectional promoters. *Epigenet. Chromatin,* **11**(1), 66.

[16] Kustatscher, G., Grabowski, P., and Rappsilber, J. (2017) Pervasive coexpression of spatially proximal genes is buffered at the protein level. *Mol. Syst. Biol.,* **13**(8), 937.

[17] Scruggs, B. S., Gilchrist, D. A., Nechaev, S., Muse, G. W., Burkholder, A., Fargo, D. C., and Adelman, K. (2015) Bidirectional transcription arises from two distinct hubs of transcription factor binding and active chromatin. *Mol. Cell,* **58**(6), 1101–1112.

[18] Wei, W., Pelechano, V., Jarvelin, A. I., and Steinmetz, L. M. (2011) Functional consequences of bidirectional promoters. *Trends Genet.,* **27**(7), 267–276.

[19] Lewin, B. (2007) Genes IX, Jones and Bartlett, Sudbury, MA.

[20] Griswold, A. (2008) Genome packaging in prokaryotes: the circular chromosome of *E. coli. Nat. Education,* **1**(1), 57.

[21] Moshkin, Y. M. (2015) Chromatin–a global buffer for eukaryotic gene control. *AIMS Biophysics,* **2**(4), 531–554.

[22] Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G. D., Giaever, G., Nislow, C., and Reinberg, D. (2012) Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife,* **1**, e00078.

[23] McClure, W. R. (1980) Rate-limiting steps in RNA chain initiation. *Proc. Natl. Acad. Sci. U.S.A.,* **77**(10), 5634–5638.

[24] Lloyd-Price, J., Startceva, S., Kandavalli, V., Chandraseelan, J., Goncalves, N., Oliveira, S. M. D., Hakkinen, A., and Ribeiro, A. S. (2016) Dissecting the stochastic transcription initiation process in live *Escherichia coli*. *DNA Res.,* **23**(3), 203–214.

[25] Revyakin, A., Ebright, R. H., and Strick, T. R. (2004) Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. *Proc. Natl. Acad. Sci. U.S.A.,* **101**(13), 4776–4780.

[26] Muthukrishnan, A.-B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O., and Ribeiro, A. S. (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucl. Acids Res.,* **40**(17), 8472–8483.

[27] Hakkinen, A. and Ribeiro, A. S. (2015) Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. *Bioinformatics,* **31**(1), 69–75.

[28] Kandavalli, V. K., Tran, H., and Ribeiro, A. S. (2016) Effects of σfactor competition are promoter initiation kinetics dependent. *BBA Gene Regul. Mech.,* **1859**(10), 1281–1288.

[29] Hakkinen, A. and Ribeiro, A. S. (2016) Characterizing rate limiting steps in transcription from RNA production times in live cells. *Bioinformatics,* **32**(9), 1346–1352.

[30] Vvedenskaya, I. O., Vahedian-Movahed, H., Zhang, Y., Taylor, D. M., Ebright, R. H., and Nickels, B. E. (2016) Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection. *Proc. Natl. Acad. Sci. U.S.A.,* **113**(21), E2899–E2905.

[31] Record, Jr., M. T., Reznikoff, W. S., Craig, M. L., McQuade, K. L., and Schlax, P. J. (1996) *Escherichia coli* RNA polymerase ($E\sigma^{70}$), promoters, and the kinetics of the steps of transcription initiation. In Neidhart, F. C., Ingraham, J. L., Low, K. B., Magasanik, B., Schaechter, M., and Umbarger, H. E., (eds.), *Escherichia coli and Salmonella typhimurium: Cellular and molecular biology*, pp. 792–820 ASM Press Washington, DC 2nd edition.

[32] Wang, F. and Greene, E. C. (2011) Single-molecule studies of transcription: From one RNA polymerase at a time to the gene expression profile of a cell. *J. Mol. Biol.,* **412**, 814–831.

[33] Patrick, M., Dennis, P. P., Ehrenberg, M., and Bremer, H. (2015) Free RNA polymerase in *Escherichia coli*. *Biochimie,* **119**, 80–91.

[34] McQuarrie, D. A. (1967) Stochastic approach to chemical kinetics. *J. Appl. Probab.,* **4**, 413–478.

[35] Gillespie, D. T. (2009) Stochastic simulation of chemical kinetics. *Annu. Rev. Phys. Chem.,* **58**, 35–55.

[36] Kaern, M., Elston, T. C., Blake, W. J., and Collins, J. J. (2005) Stochasticity in gene expression: From theories to phenotypes. *Nat. Rev. Genet.,* **6**(6), 451–464.

[37] Cox, D. R. (1962) Renewal theory, Methuen, London, UK.

[38] Pedraza, J. M. and Paulsson, J. (2008) Effects of molecular memory and bursting on fluctuations in gene expression. *Science,* **319**(5861), 339–343.

[39] Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K. A., Tomita, M., Wanner, B. L., and Mori, H. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.,* **2**(1), 1–11.

[40] Glascock, C. B. and Weickert, M. J. (1998) Using chromosomal *lacI*Q1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. *Gene,* **223**(1–2), 221–231.

[41] Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., and Amster-Choder, O. (2011) Translation-independent localization of mRNA in *E. coli*. *Science,* **331**(6020), 1081–1084.

[42] Le, T. T., Harlepp, S., Guet, C. C., Dittmar, K., Emonet, T., Pan, T., and Cluzel, P. (2005) Real-time RNA profiling within a single bacterium. *Proc. Natl. Acad. Sci. U.S.A.,* **102**(26), 9160–9164.

[43] Oliveira, S. M. D., Hakkinen, A., Lloyd-Price, J., Tran, H., Kandavalli, V., and Ribeiro, A. S. (2016) Temperature-dependent model of multi-step transcription initiation in *Escherichia coli* based on live single-cell measurements. *PLoS Comput. Biol.,* **12**(10), e1005174.

[44] Hakkinen, A., Muthukrishnan, A.-B., Mora, A., Fonseca, J. M., and Ribeiro, A. S. (2013) CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*. *Bioinformatics,* **29**(13), 1708–1709.

[45] Hakkinen, A., Kandhavelu, M., Garasto, S., and Ribeiro, A. S. (2014) Estimation of fluorescence-tagged RNA numbers from spot intensities. *Bioinformatics,* **30**(8), 1146–1153.

[46] Golding, I. and Cox, E. C. (2004) RNA dynamics in live *Escherichia coli* cells. *Proc. Natl. Acad. Sci. U.S.A.,* **101**(31), 11310–11315.

[47] Peccoud, J. and Ycart, B. (1995) Markovian modeling of gene-product synthesis. *Theor. Popul. Biol.,* **48**(2), 222–234.

[48] Mendoza-Vargas, A., Olvera, L., Olvera, M., Grande, R., Vega-Alvarado, L., Taboada, B., Jimenez-Jacinto, V., Salgado, H., Juarez, K., Contreras-Moreira, B., Huerta, A. M., Collado-Vides, J., and Morett, E. (2009) Genome-wide identification of transcription start sites, promoters and transcription factor binding sites in *E. coli*. *PLoS ONE,* **4**(10), e7526.

[49] Tracy, R. L. and Stern, D. B. (1995) Mitochondrial transcription initiation: promoter structures and RNA polymerases. *Curr. Genet.,* **28**(3), 205–216.

[50] 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. *Nucl. Acids Res.,* **29**(18), 3873–3881.