

## RESEARCH ARTICLE

Analytical kinetic model of native tandem promoters in *E. coli*

Vatsala Chauhan<sup>1</sup>, Mohamed N. M. Bahrudeen<sup>1</sup>, Cristina S. D. Palma<sup>1</sup>, Ines S. C. Baptista<sup>1</sup>, Bilena L. B. Almeida<sup>1</sup>, Suchintak Dash<sup>1</sup>, Vinodh Kandavalli<sup>2</sup>, Andre S. Ribeiro<sup>1\*</sup>

**1** Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Finland, **2** Department of Cell and Molecular Biology, Uppsala University, Uppsala, Sweden

☞ These authors contributed equally to this work.

\* [andre.sanchesribeiro@tuni.fi](mailto:andre.sanchesribeiro@tuni.fi)



## OPEN ACCESS

**Citation:** Chauhan V, Bahrudeen MNM, Palma CSD, Baptista ISC, Almeida BLB, Dash S, et al. (2022) Analytical kinetic model of native tandem promoters in *E. coli*. PLoS Comput Biol 18(1): e1009824. <https://doi.org/10.1371/journal.pcbi.1009824>

**Editor:** Eli Zunder, University of Virginia, UNITED STATES

**Received:** August 13, 2021

**Accepted:** January 11, 2022

**Published:** January 31, 2022

**Copyright:** © 2022 Chauhan et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** A data package was deposited in Dryad (Ref. [59] in main manuscript) under the DOI: [10.5061/dryad.bnzs7h4bs](https://doi.org/10.5061/dryad.bnzs7h4bs). It contains the flow-cytometry and microscopy data, along with the MATLAB, R and Python codes used. The data is accessible through this link: <https://datadryad.org/stash/share/CYS3FJMMhrs8aqPq4ILFsGunynao-Au0wTPvGYWS4oQ>. Meanwhile, RNA-seq data is deposited in NCBI GEO with the accession code GSE183139 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183139>).

## Abstract

Closely spaced promoters in tandem formation are abundant in bacteria. We investigated the evolutionary conservation, biological functions, and the RNA and single-cell protein expression of genes regulated by tandem promoters in *E. coli*. We also studied the sequence (distance between transcription start sites ' $d_{TSS}$ ', pause sequences, and distances from oriC) and potential influence of the input transcription factors of these promoters. From this, we propose an analytical model of gene expression based on measured expression dynamics, where RNAP-promoter occupancy times and  $d_{TSS}$  are the key regulators of transcription interference due to TSS occlusion by RNAP at one of the promoters (when  $d_{TSS} \leq 35$  bp) and RNAP occupancy of the downstream promoter (when  $d_{TSS} > 35$  bp). Occlusion and downstream promoter occupancy are modeled as linear functions of occupancy time, while the influence of  $d_{TSS}$  is implemented by a continuous step function, fit to *in vivo* data on mean single-cell protein numbers of 30 natural genes controlled by tandem promoters. The best-fitting step is at 35 bp, matching the length of DNA occupied by RNAP in the open complex formation. This model accurately predicts the squared coefficient of variation and skewness of the natural single-cell protein numbers as a function of  $d_{TSS}$ . Additional predictions suggest that promoters in tandem formation can cover a wide range of transcription dynamics within realistic intervals of parameter values. By accurately capturing the dynamics of these promoters, this model can be helpful to predict the dynamics of new promoters and contribute to the expansion of the repertoire of expression dynamics available to synthetic genetic constructs.

## Author summary

Tandem promoters are common in nature, but investigations on their dynamics have so far largely relied on synthetic constructs. Thus, their regulation and potentially unique dynamics remain unexplored. We first performed a comprehensive exploration of the conservation of genes regulated by these promoters in *E. coli* and the properties of their input transcription factors. We then measured protein and RNA levels expressed by 30

**Funding:** This work was supported by the Jane and Aatos Erkkö Foundation (10-10524-38) [ASR]; Finnish Cultural Foundation (50201300 to [SD] and 00200193 to [ISCB]); Suomalainen Tiedeakatemia [CSDP]; Tampere University Graduate Program [VC, MNMB, BLBA]; and EDUFI Fellowship (TM-19-11105) [SD]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

*Escherichia coli* tandem promoters, to establish an analytical model of the expression dynamics of genes controlled by such promoters. We show that start site occlusion and downstream RNAP occupancy can be realistically captured by a model with RNAP binding affinity, the time length of open complex formation, and the nucleotide distance between transcription start sites. This study contributes to a better understanding of the unique dynamics tandem promoters can bring to the dynamics of gene networks and will assist in their use in synthetic genetic circuits.

## Introduction

Closely spaced promoters exist in all branches of life in convergent, divergent, and tandem formations [1–7]. Models of tandem promoters [8–10] have largely been based on measurements of synthetic constructs [11–13] and predict that such promoter arrangements result in unique transcription dynamics due to the interference between RNAPs transcribing the promoters [9,10,14–19].

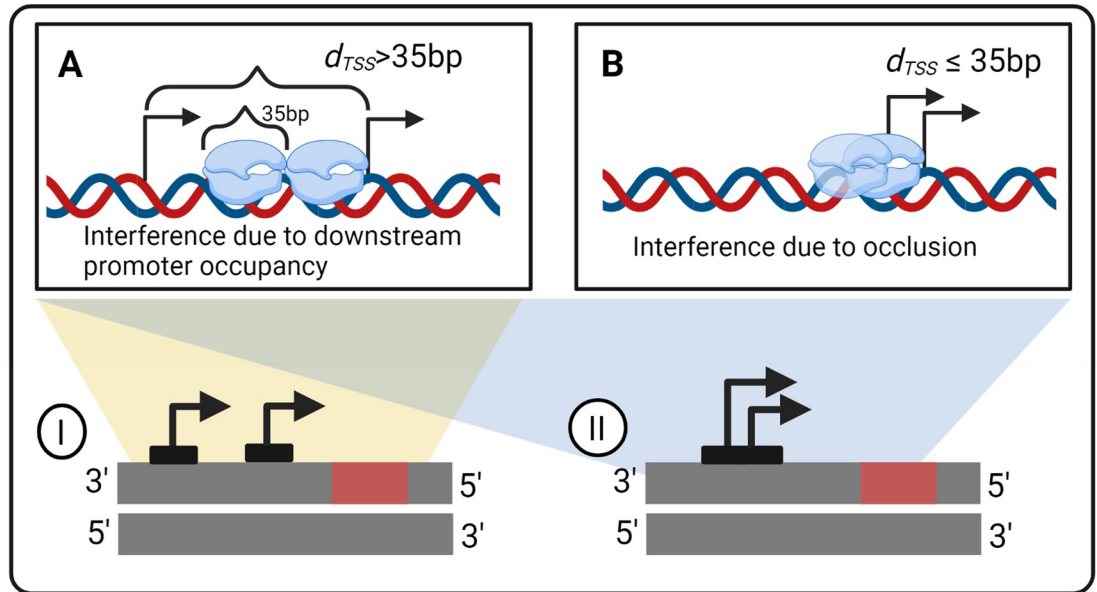
When an RNAP is committed to form the open complex (OC), a process lasting up to hundreds of seconds [20–22], it occupies approximately 35 base pairs (bp), from the transcription start site (TSS, position 0) until position -35 [23–25]. If the TSS of a neighbouring promoter is closer than 35 bp it will not be possible for both promoters to be occupied simultaneously, since an RNAP occupying one of them will ‘occlude’ the other, preventing it from being reached [9]. However, if the promoters are more than 35 bp apart, this occlusion does not occur. Instead, interference will occur when RNAPs elongating from the upstream promoter collide with an RNAP occupying the downstream promoter [14] (in either closed or open complex formation), forcing one of the RNAPs to fall-off (both scenarios are likely possible, and we expect it to differ with, e.g., the binding affinity of the RNAP to the downstream promoter). Meanwhile, models based on empirical parameter values suggest that collisions between two elongating RNAPs are rare (because events such as pausing or simultaneous initiations from both promoters are rare). Also, even if and when such collisions occur, they are unlikely to result in fall-offs since the RNAPs are moving at similar speeds and in the same direction [9,10,26].

Models suggest that both forms of interference decrease the mean RNA production rate while increasing its noise based on the distance between promoters ( $d_{TSS}$ ), their strengths [10], and the time spent between commitment of the RNAP to OC and escape from the promoter region [27]. These hypotheses have yet to be empirically validated in natural tandem promoters.

We studied how  $d_{TSS}$  and the time spent by RNAPs on the TSSs affect gene expression dynamics due to interference between the transcription processes of tandem promoters (Fig 1). We consider only the natural tandem promoters that neither overlap with nor have in between another gene (positionings I and II, which differ in if the promoter regions overlap or not) (see the other arrangements in Fig A in the S2 Appendix). The numbers of these arrangements in *E. coli* are shown in Table H in the S3 Appendix. From the measurements of these genes’ protein levels, we then establish a model that we use to explore the state space of potential dynamics under the control of tandem promoters (Fig 2 illustrates our workflow).

## Results

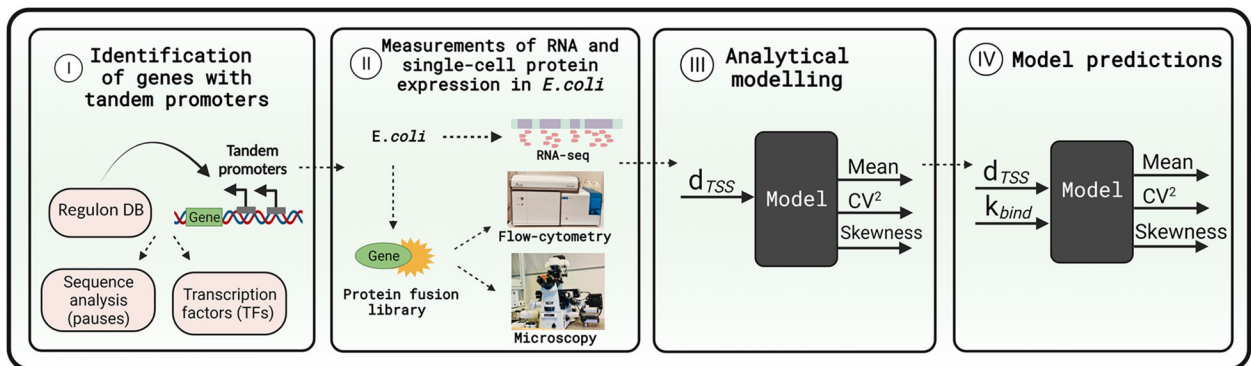
*E. coli* has 831 genes controlled by two or more promoters in tandem formation (RegulonDB and section ‘Selection of natural genes controlled by tandem promoters’



**Fig 1. Interference between tandem promoters with different arrangements relative to each other.** (A) Interference by an RNAP occupying the downstream promoter on the activity of the elongating RNAP from upstream promoter. The TSSs need to be at least 36 bp apart (the length occupied by an RNAP when in OC, [23,25]) (B) Interference by occlusion of one of the promoter's TSS by an RNAP on the TSS of the other promoter. The distance between the TSSs need to be  $\leq 35$  bp apart. Blue clouds are RNAPs. Black arrows sit on TSSs and point towards the direction of transcription elongation. Arrangements (I-II) of two promoters studied in the manuscript in tandem formation are represented. The red rectangles are the protein coding regions. We studied only the natural tandem promoters that neither overlap with nor have in between another gene (arrangements I and II, which differ based on whether the promoter regions overlap or not). Other arrangements (not considered in this study) are shown in Fig A in the S2 Appendix. Figure created with BioRender.com.

<https://doi.org/10.1371/journal.pcbi.1009824.g001>

in the S1 Appendix). However, to study the dynamics of genes controlled by tandem promoters, we focused on only 102 of them, because their activity is expected to be undisturbed by neighboring genes in the DNA (arrangements I and II in Fig 1), for reasons



**Fig 2. Workflow.** (I) We identified genes controlled by tandem promoters in Regulon DB. (II) Next, we measured the single-cell protein levels of those genes with arrangements I and II that are tagged in the YFP strain library [28]. We also measured the mean RNA fold changes of these genes over time (S1 Appendix, section 'RNA-seq measurements and data analysis'). (III) We used the single-cell data to tune the model. (IV) Finally, we used the model to explore the state space of protein expression. Figure created with BioRender.com.

<https://doi.org/10.1371/journal.pcbi.1009824.g002>

described in section ‘Selection of natural genes controlled by tandem promoters’ in the [S1 Appendix](#).

Further, these promoters do not have specific short nucleotide sequences capable of affecting RNAP elongation (section ‘Pause sequences’ in the [S4 Appendix](#)). Also, the 102 genes expressed by these promoters are not overrepresented in a particular biological process (section ‘Over-representation test’ in the [S4 Appendix](#)). From time-lapse RNA-seq data ([S1 Appendix](#), section ‘RNA-seq measurements and data analysis’), we also did not find evidence that their dynamics are affected by their input transcription factors (TFs) in our measurement conditions (section ‘Input-output transcription factor relationships’ in the [S4 Appendix](#)) nor by H-NS in a consistent manner (section ‘Regulation by H-NS’ in the [S4 Appendix](#)). Finally, they do not exhibit any particular TF network features (Table C in the [S3 Appendix](#)). As such, neither input TFs nor specific nucleotide sequences are considered in the model below. In addition to all of the above, we found no correlations between the shortest distance from the TSS of upstream promoters from the oriC region in the DNA and expression levels (section ‘Relationship with the oriC region’ in the [S4 Appendix](#)).

### Model of gene expression controlled by tandem promoters

RNAPs bind, slide along, and unbind from a promoter several times until, eventually, one of them finds the TSS [29–30], commits to OC at the TSS, and initiates transcription elongation.

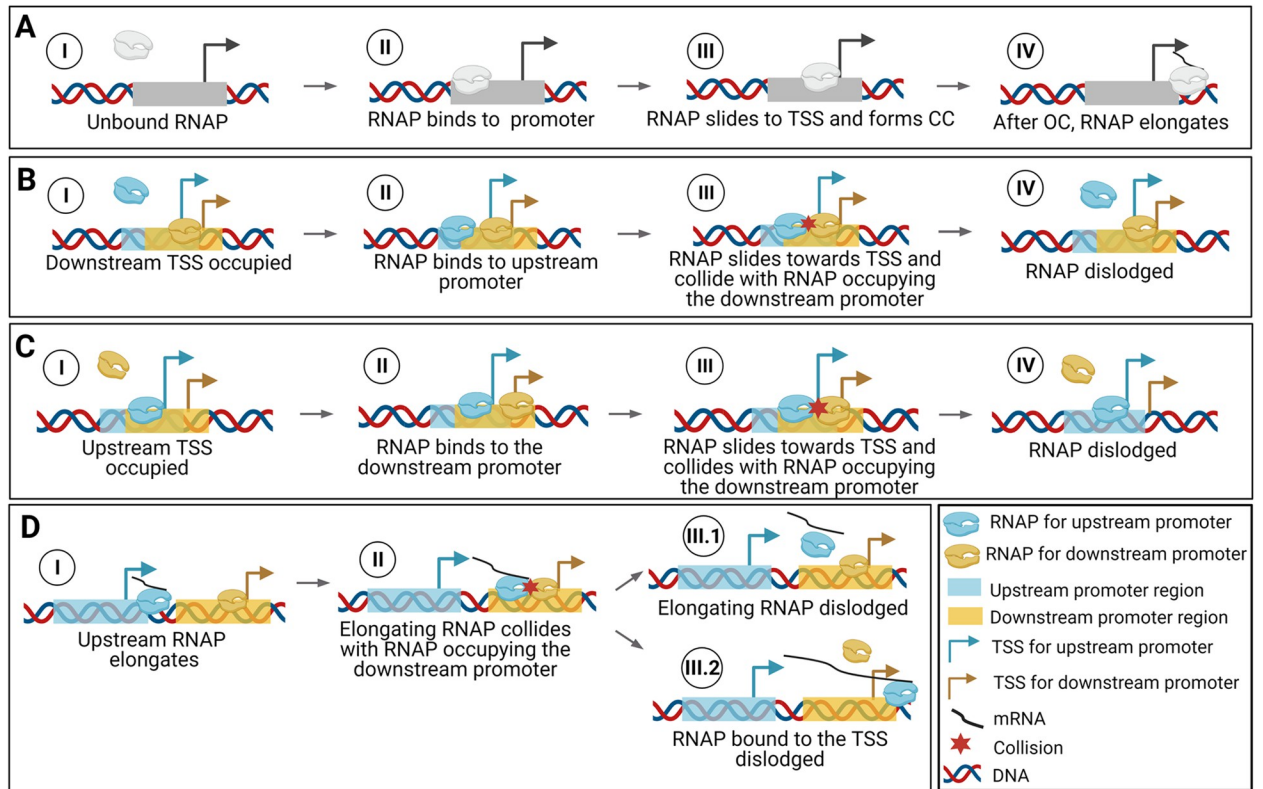
Reactions (1A1) are a 4-step (I-IV) model of transcription [20,31]. The forward reaction in step I in (1A1) models RNAP binding to a free promoter ( $P_{free}$ ), which becomes no longer free albeit the RNAP might not yet have reached the TSS. This state, pre-finding of the TSS, is here named  $P_{bound}$  and its occurrence increases with RNAP concentration,  $[R]$ . Next, as it percolates the DNA, the RNAP should find and stop at the nearest TSS and form a closed complex (CC) with the DNA (step II, Reaction 1A1). CCs are unstable, i.e. reversible [22] (reaction 1A2) but, eventually, one of them will commit to OC irreversibly [32], via step III, Reaction 1A1 [21–22]. It follows RNAP escape from the TSS, freeing the promoter (step IV, Reaction 1A1) [33–37]. Then, the RNAP elongates ( $R_{elong}$ ) until producing a complete RNA (reaction 1A3) and freeing itself.

These set of reactions usually model well stochastic transcription dynamics [20]. However, if two promoters are closely spaced in tandem formation, they can interfere [38]. [Fig 3](#) shows sequences of events that can lead to interference between tandem promoters, not accounted for by the model above.

From [Fig 3](#), if the TSSs are sufficiently close, the occupancy of one TSS by an RNAP will occlude the other TSS, blocking its kinetics [18]. This is accounted for by reaction 1A5, which competes with CC formation in reaction 1a1. Its rate constant,  $k_{occlusion}$ , is defined in the next section. In (1A5), ‘u/d’ stands for occlusion of the upstream promoter by an RNAP on the TSS of the downstream promoter.

Instead, if the TSSs are not sufficiently close, they will still interfere since the elongating RNAP ( $R_{elong}$ ) starting from the upstream promoter can collide with RNAPs on the TSS of the downstream promoter. This can dislodge either RNAP via (reaction 1A4) or (reaction 2A3), depending on the sequence-dependent binding strength of the RNAP to the TSS [9].

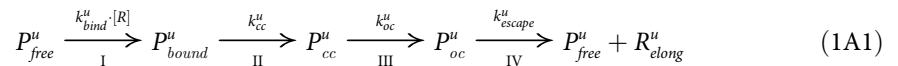
Finally, once reaction 1A1 occurs, either reaction 1A3 or 1A4 occur. To tune their competition, we introduced the terms  $\omega_d$  and  $(1 - \omega_d)$  in their rate constants, with  $\omega_d$  being the fraction of times that an elongating RNAP from an upstream promoter finds an RNAP occupying the downstream promoter. Meanwhile,  $f$  is the fraction of times that the RNAP occupying the downstream promoter falls-off due to the collision with an elongating RNAP, whereas  $1-f$  is



**Fig 3. Events leading to transcriptional interference between tandem promoters.** (A) Sequence of events in transcription in isolated promoters. A similar set of events occurs in tandem promoters, if only one RNAP interacts with them at any given time. (B / C) Interference due to the occlusion of the *downstream* / *upstream* promoter by a bound RNAP, which will impede the incoming RNAP from binding to the TSS. (D) Interference of the activity of the RNAP incoming from the upstream promoter by the RNAP occupying the downstream promoter. One of these RNAPs will be dislodged by the collision. Created with BioRender.com.

<https://doi.org/10.1371/journal.pcbi.1009824.g003>

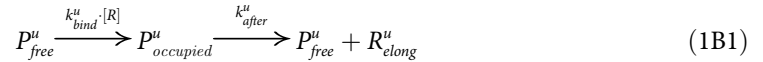
the fraction of times that it is the elongating RNAP that falls-off.



Next, we reduced the model and derived its analytical solution. First, since  $P_{cc}$  completion is expected to be faster than  $P_{bound}$  completion ([10] and references within) we merged them into a single state,  $P_{occupied}$ , which represents a promoter occupied by an RNAP prior to commitment to OC, whose time length is similar to  $P_{bound}$ .

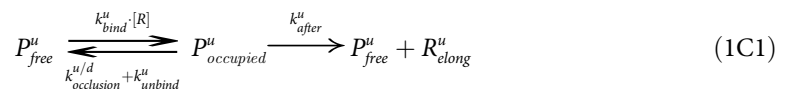


Similarly, in standard growth conditions, the occurrence of multiple failures in escaping the promoter per OC completion should only occur in promoters with the highest binding affinity to RNAP. Thus, in general promoter escape should be faster than OC [20,32]. We thus merged OC and promoter escape into one step named ‘events after commitment to OC’, with a rate constant  $k_{after}$ . The simplified model is thus:



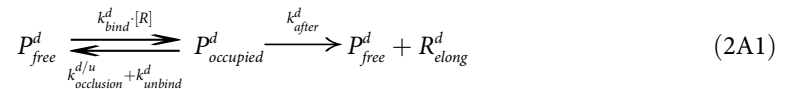
These two steps are not merged since only the first differs with RNAP concentration [20,26,39]. Further, reports [40–41] indicate that *E. coli* has ~100–1000 RNAPs free for binding at any moment but ~4000 genes, suggesting that the number of free RNAPs is a limiting factor.

Finally, we merge (1A2), (1A5) and (1B1) in one multistep without affecting the model kinetics:

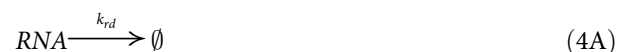


Overall, this reduced model of transcription of upstream promoters has a multistep reaction of transcription initiation (1C1), a reaction of transcription elongation (1A3) and a reaction for failed elongation due to RNAPs occupying the downstream promoter (1A4).

Regarding RNA production from the downstream promoter, it should either be affected by occlusion if  $d_{TSS} \leq 35$ , or by RNAPs elongating from the upstream promoter if  $d_{TSS} > 35$  (Fig 3). We thus use reactions (2A1), (2A2), and (2A3) to model these promoters’ kinetics:



Finally, one needs to include a reaction for translation (reaction 3), as a first order process since protein numbers follow RNA numbers linearly (Fig F in the S2 Appendix), and reactions for RNA and protein decay accounting for degradation and for dilution due to cell division (reactions 4A and 4B, respectively). TF regulation is not included as noted above (Fig C and panel A of Fig D in the S2 Appendix).



### Transcription interference by occlusion

In a pair of tandem promoters, the  $k_{occlusion}$  of one of them should increase with the fraction of time that the other one is occupied. Further, it should decrease with increasing  $d_{TSS}$  between the two promoters' TSS. We thus define  $k_{occlusion}$  for the upstream (Eq 5A) and downstream (Eq 5B) promoters, respectively as:

$$k_{occlusion}^{u/d} = k_{occl}^{max} \cdot I(d_{TSS}) \cdot \omega_d \tag{5A}$$

$$k_{occlusion}^{d/u} = k_{occl}^{max} \cdot I(d_{TSS}) \cdot \omega_u \tag{5B}$$

Here,  $k_{occl}^{max}$  is the maximum occlusion possible. It occurs when the two TSSs completely overlap each other ( $d_{TSS} = 0$ ) and the TSS of the 'other' promoter is always occupied. Meanwhile,  $I(d_{TSS})$  models distance-dependent interference.

We tested four models of interference: 'exponential 1', 'exponential 2', 'step', and 'zero order' (Table 1). The first two assume that the effects of occlusion decrease exponentially with  $d_{TSS}$  (first and second order dependency, respectively).

Meanwhile, the 'Step' model assumes that interference only occurs precisely in the region in the DNA occupied by the RNAP when in OC formation. For this, it uses a logistic equation to build a continuous step function, where  $L$  is the length of DNA (in bp) occupied by the RNAP in OC. As such,  $L$  tunes at what  $d_{TSS}$  the step occurs, while  $m$  is the steepness of that step (set to  $1 \text{ bp}^{-1}$ ).

Finally, the 'Zero order' model assumes (unrealistically) that interference by occlusion, is independent of  $d_{TSS}$ . Fig G in the S2 Appendix shows how  $k_{occlusion}$  differs with  $d_{TSS}$  in each model, for various parameter values.

Finally,  $\omega$  is the fraction of time that the 'other' promoter is occupied. It ranges from 0 (no occupancy) to 1 (always occupied). It is estimated for upstream and downstream promoters as:

$$\omega_u = \frac{k_{bind}^u \cdot [R]}{k_{unbind}^u + k_{bind}^u \cdot [R] + k_{after}^u} \tag{6A}$$

$$\omega_d = \frac{k_{bind}^d \cdot [R]}{k_{unbind}^d + k_{bind}^d \cdot [R] + k_{after}^d} \tag{6B}$$

Similarly, if  $k_{occupy}^{max}$  is the maximum possible interference due to RNAPs occupying the downstream promoter,  $k_{occupy}$  is defined as:

$$k_{occupy} = \omega_u \cdot k_{after} \cdot k_{occupy}^{max} \cdot (1 - f) \tag{7}$$

**Table 1. Potential models of transcriptional interference due to promoter occlusion considered.**

Interference by occlusion	$I(d_{TSS})$	$k_{occlusion}$
Exponential 1 ("Exp1")	$e^{-(b_1 \cdot d_{TSS})}$	$k_{occl}^{max} \cdot e^{-(b_1 \cdot d_{TSS})} \cdot \omega$
Exponential 2 ("Exp2")	$e^{-(b_1 \cdot d_{TSS} + b_2 \cdot d_{TSS}^2)}$	$k_{occl}^{max} \cdot e^{-(b_1 \cdot d_{TSS} + b_2 \cdot d_{TSS}^2)} \cdot \omega$
Step ("Step")	$1 - \frac{1}{1 + e^{-m \cdot (d_{TSS} - L)}}$	$k_{occl}^{max} \cdot \left(1 - \frac{1}{1 + e^{-m \cdot (d_{TSS} - L)}}\right) \cdot \omega$ , for $m = 1 \text{ bp}^{-1}$
Zero order ("ZeroO")	$k$	$k_{occl}^{max} \cdot \omega$

<https://doi.org/10.1371/journal.pcbi.1009824.t001>

### Analytical solution of the moments of the single-cell protein numbers

Next, we derived an analytical solution of the expected mean single-cell protein numbers at steady state,  $M_p$ , which is later tuned to fit the empirical data. For any gene, regardless of the underlying kinetics of transcription,  $k_r$  is the *effective* rate of RNA production. Based on the reactions above, the mean protein numbers in steady state will be (see sections “Analytical model of mean RNA levels controlled by a single promoter in the absence of a closely spaced promoter” and “Derivation of mean protein numbers at steady state produced by a pair of tandem promoters” in the [S1 Appendix](#)):

$$M_p = \frac{k_r \cdot k_p}{k_{rd} \cdot k_{pd}} \quad (8)$$

This equation applies to a pair of tandem promoters as well. In that case, assuming that  $k_{bind}$  of the two tandem promoters is similar, we have:

$$k_r = \left( \frac{\frac{k_{bind} \cdot [R] \times k_{after} \cdot (1 - \omega_d \cdot f)}{k_{occlusion} + k_{bind} \cdot [R] + k_{unbind} + k_{after}} + \frac{k_{bind} \cdot [R] \times k_{after}}{k_{occlusion} + k_{occupy} + k_{bind} \cdot [R] + k_{unbind} + k_{after}}} \right) \quad (9)$$

To derive the other moments, we considered that empirical single-cell protein numbers in *E. coli* are well fit by negative binomials [28]. Consequently,  $M_p$  and the squared coefficient of variation  $CV_p^2$ , should be related as (Equations S28 to S38 in the [S1 Appendix](#)):

$$\log_{10}(CV_p^2) = \log_{10}(C_1) - \log_{10}(M_p), \quad \text{with} \quad C_1 = \frac{k_p}{k_{pd} + k_{rd}} \quad (10)$$

This relationship matches empirical data at the genome wide level, except for genes with high transcription rates [42]. Additionally, we further derived a relationship (Section ‘ $CV^2$  and Skewness of single-cell protein expression of a model tandem promoters’ in the [S1 Appendix](#)) between  $M_p$  and the skewness,  $S_p$ , of the single-cell distribution of protein numbers:

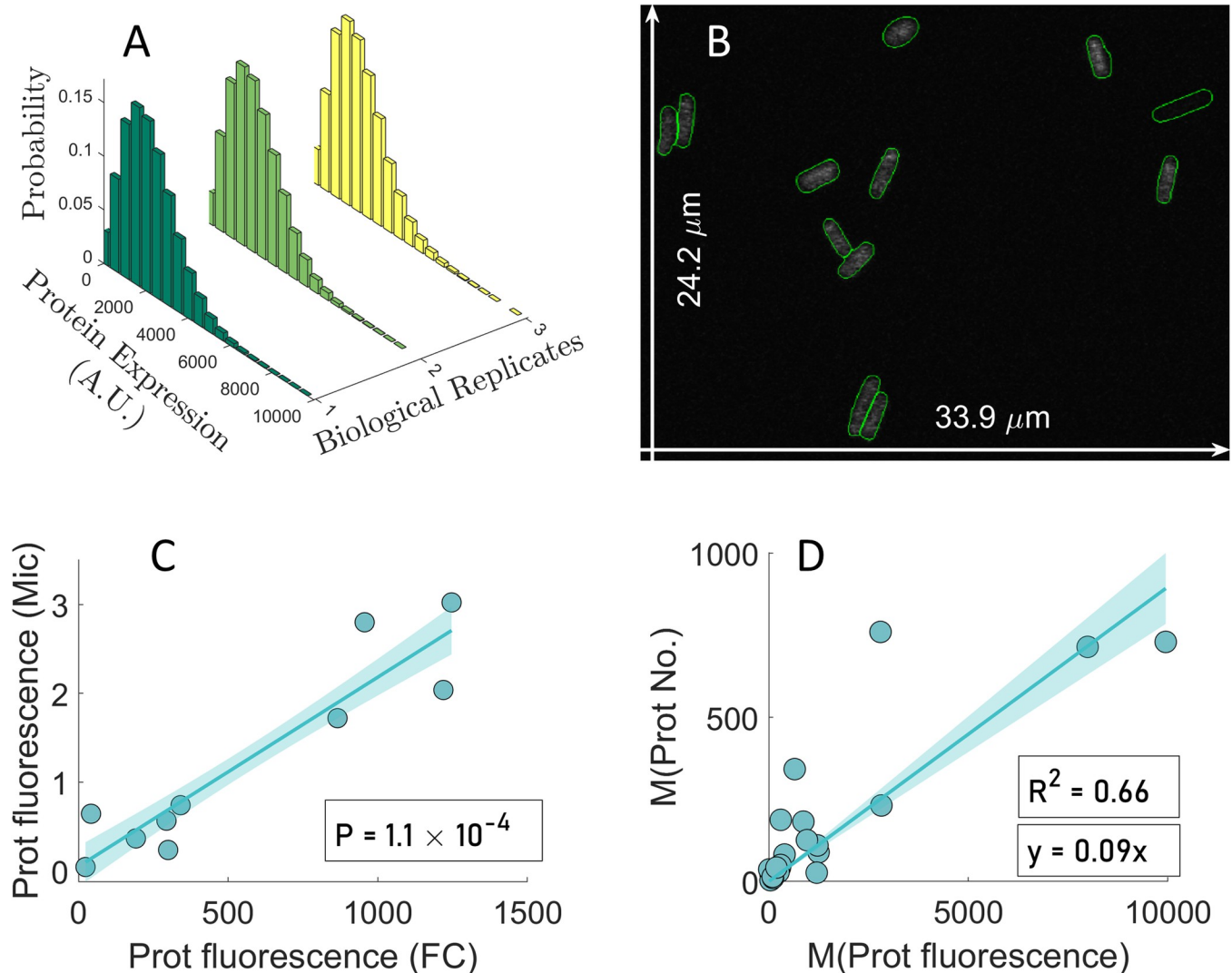
$$\log_{10}(S_p) = \log_{10}(C_2) - \frac{1}{2} \cdot \log_{10}(M_p), \quad \text{with} \quad C_2 = 2\sqrt{C_1} - \frac{1}{\sqrt{C_1}} \quad (11)$$

### Single-cell distributions of protein numbers

To validate the model, we measured by flow-cytometry the single-cell distributions of protein fluorescence of 30 out of the 102 genes known to be controlled by tandem promoters (with arrangements I and II). Measurements were made in 1X and 0.5X media (3 replicates per condition) using cells from the YFP strain library (section ‘Strains and Growth Conditions’ in the [S1 Appendix](#)). Data from past studies show that, in these 30 genes, RNA and protein numbers are well correlated (Fig F in the [S2 Appendix](#)) in standard growth conditions. Past studies also suggest that most of these genes are active during exponential growth (~95% of our 30 genes selected should be active, according to data in [43] using Send-seq technology).

Single-cell distributions of protein expression levels are shown in [Fig 4A](#) for one of these genes as an example. The raw data from all 30 genes (only one replicate) are shown in [Fig H](#) in the [S2 Appendix](#). Finally, the mean,  $CV^2$  and skewness for each gene, obtained from the triplicates, are shown in Excel sheets 1 and 2 in the [S2 Table](#). In addition, we also show this mean,  $CV^2$  and skewness after subtracting the first, second, and third moments of the single-cell distribution of the fluorescence of control cells, which do not express YFP (Sheets 3, 4 in the [S2](#)





**Fig 4. Single cell protein numbers by microscopy and flow-cytometry.** (A) Example single-cell distributions (3 biological replicates) of fluorescence (in arbitrary units) of cells with a YFP tagged gene controlled by a pair of tandem promoters obtained by flow-cytometry, 'FC'. (B) Example confocal microscopy image of cells overlapped by the results of cell segmentation from the corresponding phase contrast image. The two white arrows show the dimensions of the image, for scaling purposes. (C) Mean single-cell protein fluorescence of 10 genes (Table G in the [S3 Appendix](#)) when obtained by FC plotted against when obtained by microscopy, 'Mic'. (D) Mean single-cell protein fluorescence (own measurements) plotted against the corresponding mean single-cell protein numbers reported in [28]. From the equation of the best fitting line without y-intercept (y-intercept = 0), we obtained a scaling factor, *sf*, equal to 0.09.

<https://doi.org/10.1371/journal.pcbi.1009824.g004>

[Table](#)) (Section 'Subtraction of background fluorescence from the total protein fluorescence' in flow-cytometry in the [S1 Appendix](#)).

Based on the analysis of the data of these 30 genes, we removed from subsequent analysis those genes (5 in 1X and 14 in 0.5X) whose mean, variance, or third moment of their protein fluorescence distributions are lower than in control cells (not expressing YFP), i.e., than cellular autofluorescence (Sheets 3, 4 in [S2 Table](#)). As such, only one gene studied here (in condition 1X alone) codes for a protein that is associated to membrane-related processes, which might affect its quantification (section 'Proteins with membrane-related positionings' in [S4 Appendix](#)). As such, we do not expect this phenomenon to influence our results significantly. The data from these genes removed from further analysis is shown in Fig F in [S2 Appendix](#) alone, for illustrative purposes.

We started by testing the accuracy of the background-subtracted flow-cytometry data by confronting it with microscopy data (also after background subtraction, see section ‘Microscopy and Image Analysis’ in the [S1 Appendix](#)). We collected microscopy data on 10 out of the 30 genes (Table G in the [S3 Appendix](#)). The microscopy measurements of the mean single-cell fluorescence expressed by these genes (example image in [Fig 4B](#)), were consistent, statistically, with the corresponding data obtained by flow-cytometry ([Fig 4C](#)).

Next, we converted the fluorescence distributions from flow-cytometry (25 genes in 1X and 16 genes in 0.5X) into protein number distributions. In [Fig 4D](#) we plotted our measurements of mean protein fluorescence in 1X against the protein numbers reported in [28] for the same genes, in order to obtain a scaling factor ( $sf = 0.09$ ). Using  $sf$ , we estimated  $M_p$ ,  $CV_p^2$ , and  $S_p$  of the distribution of protein numbers expressed by the tandem promoters in (Sheets 5, 6 in [S2 Table](#)) (Section ‘Conversion of protein fluorescence to protein numbers’ in [S1 Appendix](#)).

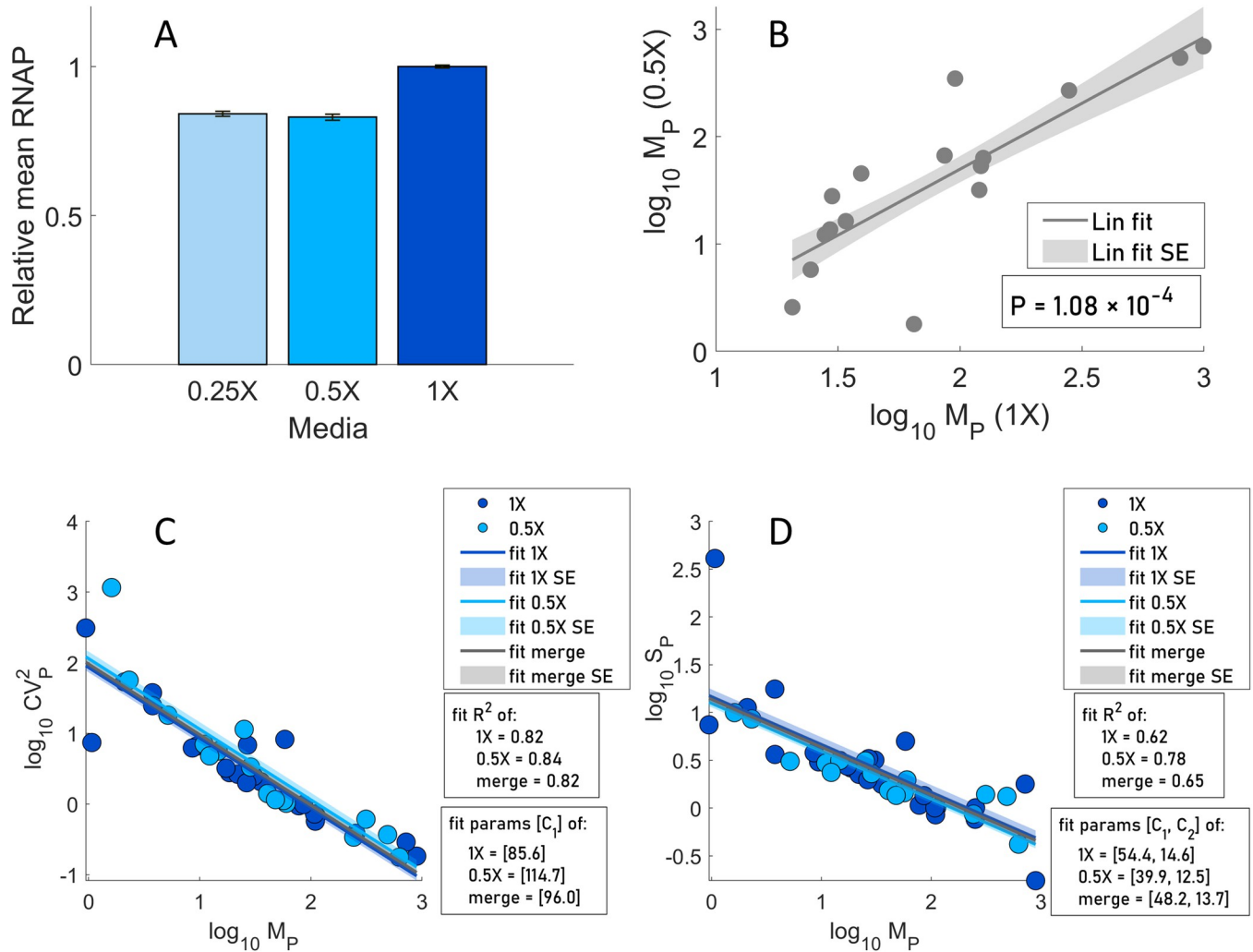
To test the robustness of the estimation of the scaling factor, we also estimated a scaling factor from 10 other genes present in the YFP strain library [28] (listed in Table B in [S3 Appendix](#)). These genes were selected as described in the section ‘Selection of natural genes controlled by single promoters’ in [S1 Appendix](#). Using the data from this new gene cohort (Panel A of Fig I in [S2 Appendix](#)) reported in [S3 Table](#), we estimated a scaling factor of 0.08, supporting the previous result. Meanwhile, since when merging the data from tandem and single promoters, the resulting scaling factor equals 0.09 (Panel B of Fig I in [S2 Appendix](#)), we opted for using 0.09 from here onwards.

We also tested how sensitive the estimated scaling factor is to the removal of data points. Specifically, for 1000 times, we discarded  $N$  randomly selected data points, and estimated the resulting scaling factor. We then compared, for each  $N$ , the mean and the median of the distribution of 1000 scaling factors (Fig J in [S2 Appendix](#)). Since the median is not sensitive to outliers, if mean and median are similar, one can conclude that the scaling factor is not biased by a few data points. Visibly, the mean and the median only start differing for  $N$  larger than 6, which corresponds to nearly 30% of the data.

### Log-log relationship between the mean single-cell protein numbers of tandem promoters and the other moments

We plotted  $M_p$  against  $CV_p^2$  and  $S_p$  in log-log plots, in search for the fitting parameters, ‘ $C_1$ ’ and ‘ $C_2$ ’, to estimate the rate of protein production per RNA (Eq 10). To increase the state space covered by our measurements, in addition to M9 media (named ‘1X’), we also used diluted M9 media (named ‘0.5X’), known to cause cells to have lower RNAP concentrations ([Fig 5A](#)) (Section ‘Strains and growth conditions’ in the [S1 Appendix](#)), without altering the division rate (Panels A and B of Fig K in the [S2 Appendix](#)). We note that 1X and 0.5X only refer to the degree of dilution of the original media and not to how much RNAP concentration and consequently, protein concentrations, were reduced by media dilution. From the same figures, we attempted stronger dilutions, but no further decreases in RNAP concentration were observed and the growth rate decreased.

Next, from [Fig 5B](#), most genes (of those expressing tangibly in both media) suffered similar reductions (well fit by a line) in protein numbers with the media dilution, as expected by the model of gene expression (Eqs 8 and 9). This linear relationship could also be interpreted as evidence that the difference in expression of these genes between the two conditions is not affected by TFs in our measurement conditions. Namely, if TF influences existed, and TF numbers changed, they would likely be diversely affected by their output genes (weakly and strongly activated, repressed, etc.) and, thus, our proteins of interest would not have changed in such similar manners (linearly).



**Fig 5. Relative RNAP concentrations along with the relationships between the moments of the single cell distributions of protein numbers.** (A) Relative RNAP levels measured by flow-cytometry (Section ‘flow-cytometry and data analysis’ in the S1 Appendix) in three media. (B) Scatter plot between  $M_P$  in M9 (1X) and diluted M9 (0.5X) media. Also shown are the best fitting line and standard error and p-value for the null hypothesis that the slope is zero. (C)  $M_P$  vs  $CV_P^2$  and (D)  $M_P$  vs  $S_P$  of single-cell protein numbers of genes with tandem promoters in M9 (1X) and M9 diluted (0.5X) media. The lines and their shades are the best fitting lines and standard errors, respectively. ‘Merge’ stands for data from both 0.5X and 1X conditions.

<https://doi.org/10.1371/journal.pcbi.1009824.g005>

Meanwhile, as in [42,44],  $CV_P^2$  decreases linearly with  $M_P$  (log-log scale), irrespective of media ( $R^2 > 0.8$  in all fitted lines), in agreement with the model (Fig 5C). Fitting Eq 10 to the data, we extracted  $C_1$  in each condition.  $S_P$  also decreases linearly with  $M_P$ , irrespective of the media (Fig 5D). Similar to above, Eq 11 was fitted to each data set and  $C_1$  and  $C_2$  were obtained ( $R^2 > 0.6$  for all lines).

Since  $C_1$  from Fig 5C and 5D differed slightly (likely due to noise), we instead obtained  $C_1$  and  $C_2$  values that maximized the mean  $R^2$  of both plots. Using ‘fminsearch’ function in MATLAB [45], we obtained  $C_1 = 72.71$  and  $C_2 = 16.94$  ( $R^2$  of 0.80 and 0.61, respectively) for Fig 5C and Fig 5D, respectively.

### Inference of parameter values and model predictions as a function of $d_{TSS}$

We next used the model, after fitting, to predict how  $d_{TSS}$  and the promoters’ occupancy regulate the moments of the single-cell distribution of protein numbers ( $M_P$ ,  $CV_P^2$ , and  $S_P$ ) under

Table 2. Parameter values imposed identically on all models.

Parameter description	Parameter	Value	References
Inverse of the mean time to complete OC	$k_{after}$	$0.005 \text{ s}^{-1}$	Differs between promoters. Since empirical data lacks, we used the data from <i>in vivo</i> single RNA measures for Lac-Ara-1 [20].
RNA and protein dilution due to division	$k_{dil} = \frac{\ln(2)}{D}$	$1.005 \times 10^{-4} \text{ s}^{-1}$	Legend of Fig H in the S2 Appendix.
RNA degradation	$k_{rddeg}$	$2.3 \times 10^{-3} \text{ s}^{-1}$	[28]
RNA decay due to dilution from cell division and due to degradation	$k_{rd} = k_{rddeg} + k_{dil}$	$2.4 \times 10^{-3} \text{ s}^{-1}$	From row 2.
Protein degradation	$k_{pdeg}$	$2.93 \times 10^{-5} \text{ s}^{-1}$	[46], estimates it to be from $\sim 6 \times 10^{-5}$ to $\sim 2 \times 10^{-5}$ . We used the value in [47], in that interval.
Protein decay due to dilution by cell division and degradation	$k_{pd} = k_{pdeg} + k_{dil}$	$1.3 \times 10^{-4} \text{ s}^{-1}$	From rows 2 and 5.
Fall-off probability of the RNAP occupying the downstream promoter	$f$	50% (0.5)	Set here (likely sequence-dependent)
Protein production rate constant	$k_p = C_I \times (k_{pd} + k_{rd})$	$0.18 \text{ s}^{-1}$	$C_I$ is estimated here.
Free RNAP per cell	$[R]$	144/cell in 1X and 120/cell in 0.5X media	See main text.

<https://doi.org/10.1371/journal.pcbi.1009824.t002>

the control of tandem promoters. We started by assuming the parameter values from the literature listed in Table 2 and tuned the remaining parameters.

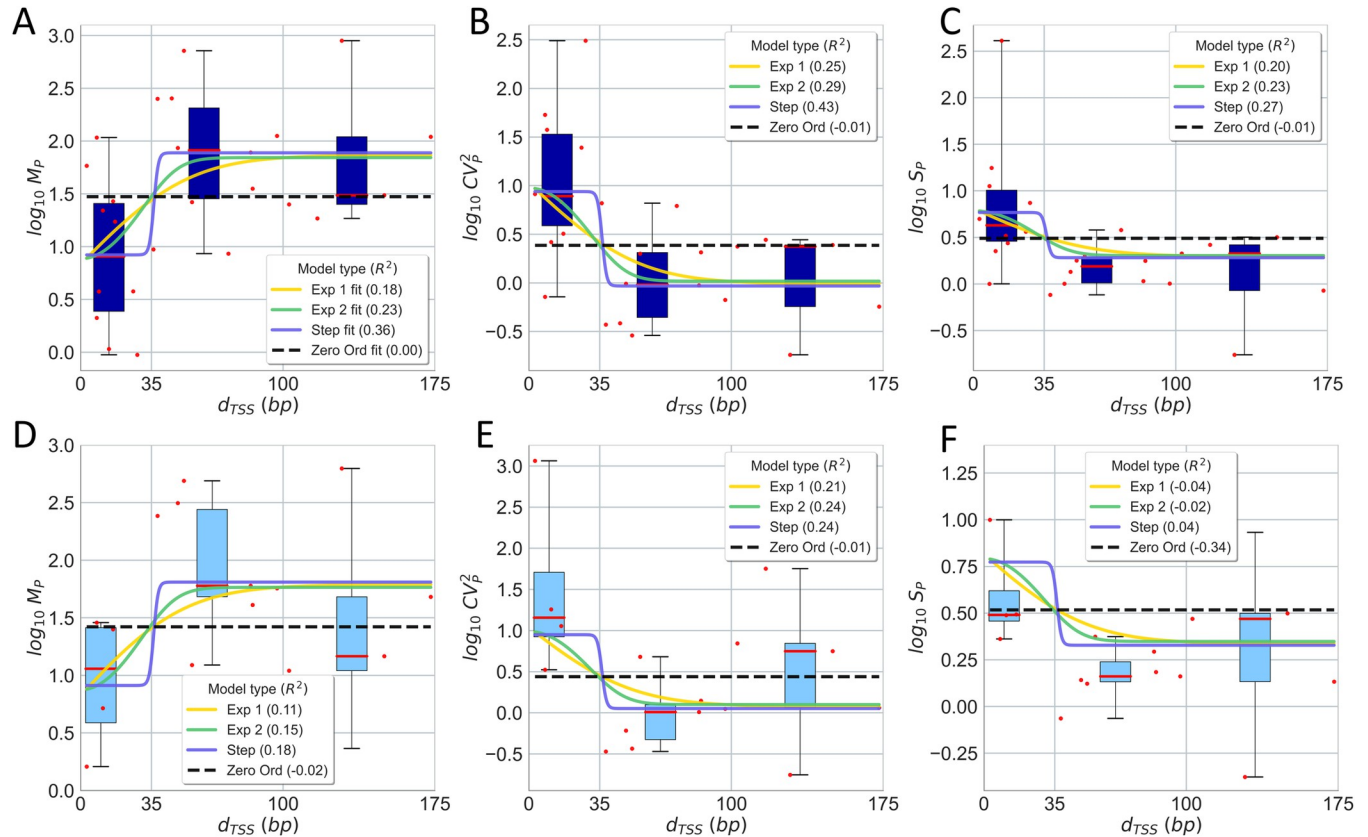
To set the RNAP numbers in Table 2, we considered that the RNAPs affecting transcription rates are the *free* RNAPs in the cell, and that, for doubling times of 30 min in rich medium, there are  $\sim 1000$  free RNAPs per cell [41]. Meanwhile, for doubling times of 60 min in minimal medium, there are  $\sim 144$  [40]. In both our media, we observed a doubling time of  $\sim 115$  mins (Fig 5B). Thus, we expect the free RNAP in 1X to also be  $\sim 144$ /cell or lower. Meanwhile, in 0.5X, we measured the RNAP concentration to be 17% lower than in 1X (Fig 5A) and no morphological changes. Thus, we assume the free RNAP in 0.5X to equal  $\sim 120$ /cell.

Next, we fitted the Eqs (8) and (9) relating  $d_{TSS}$  with  $\log_{10}(M_p)$  in all interference models (Table 1), using the data on  $M_p$  in 1X medium (Fig 6A) and the ‘fit’ function of MATLAB. For this, we set  $k^{\max} = k_{occupy}^{\max} = k_{occl}^{\max}$ , for simplicity, as well as realistic bounds for each parameter to infer. To avoid local minima, we performed 200 searches, each starting from a random initial point, and selected the one that maximized  $R^2$ . Results are shown in Table 3.

Next, we inserted all parameter values (empirical and inferred) in Eqs (10) and (11) to predict  $CV_p^2$  and  $S_p$  in 1X medium (Fig 6B and 6C). Also, we inserted the same parameter values and the estimated RNAP numbers in 0.5X medium in Eqs (8–11) to obtain the analytical solutions for  $M_p$ ,  $CV_p^2$  and  $S_p$  for 0.5X medium (Fig 6D, 6E and 6F).

From Fig 6, the data is ‘noisy’, which suggests that it is not possible to establish if the models are significantly different. As such, here we only select the one that best explains the data, based on the  $R^2$  values of the fittings. Table 3 shows the mean  $R^2$  for  $M_p$ ,  $CV_p^2$ , and  $S_p$  when confronting the model with the data. Overall, from the  $R^2$  values, the step model is the one that best fits the data. Meanwhile, the ‘ZeroO’ model is the least accurate, which supports the existence of distinct kinetics when  $d_{TSS}$  is smaller or larger than 35 nucleotides, which is the length of the RNAP when committed to OC on the TSS [23–25].

In summary, the proposed model of expression of genes under the control of a pair of tandem promoters is based on a standard model of transcription of each promoter, which are subject to interference, either due to occlusion of the TSSs or by RNAP occupying the TSS of the downstream promoter. The influence of each occurrence of these events is well modeled by linear functions of TSS occupancy times, while their dependency on  $d_{TSS}$  is modeled by a



**Fig 6. Empirical data and analytical model of how  $d_{TSS}$  influences the single-cell protein numbers of genes controlled by tandem promoters.** (A) Mean, (B)  $CV^2$ , and (C)  $S$  of single protein numbers in the 1X media as a function of  $d_{TSS}$ . (D), (E), and (F) show the same for the 0.5X media, respectively. Each red dot is the mean from 3 biological repeats for a pair of promoters (S2 Table). The dots were also grouped in 3 ‘boxes’ based on their  $d_{TSS}$ . In each box, the red line is the median and the top and bottom are the 3<sup>rd</sup> and 1<sup>st</sup> quartiles, respectively. The vertical black bars are the range between minimum and maximum of the red dots. In A, all lines are best fits. In B, C, D, E, and F, all lines are model predictions, based on the parameters used to best fit A. The insets show the  $R^2$  for each model fit and prediction.

<https://doi.org/10.1371/journal.pcbi.1009824.g006>

continuous step function. If  $d_{TSS}$  is larger than 35 bp, effects from the RNAP occupying the downstream promoter can occur, else occlusion can occur.

We then confronted the analytical solutions of the step model with stochastic simulations (Section ‘Stochastic simulations for the step inference model’ in the S1 Appendix). We first assumed various  $d_{TSS}$ , but fixed  $k_{bind}$  for simplicity. Visibly,  $M_P$ ,  $CV_p^2$ , and  $S_P$  of the stochastic simulations are well-fitted by the analytical solution, supporting the initial assumption that  $CV_p^2$ , and  $S_P$  follow a negative binomial (Fig M in the S2 Appendix).

However, natural promoters are expected to differ in  $k_{bind}$  as they differ in sequence [48,49]. Thus, we introduced this variability and studied whether the analytical model holds. To change the variability, we obtained each  $k_{bind}$  from gamma distributions (means shown in Table 3 and CVs in Table I in the S3 Appendix). We chose a gamma distribution since its values are non-negative and non-integer (such as rate constants). Meanwhile, all parameters of the step model, aside from  $k_{bind}$ , are obtained from Tables 2 and 3. For  $d_{TSS} \leq 35$  and  $d_{TSS} > 35$ , and each CV considered, we sampled 10000 pairs of values of  $k_{bind}$  [R], and calculated  $M$ ,  $CV^2$  and  $S$  for each of them. Next, we estimated the average and standard deviation of each statistics. From Fig N in the S2 Appendix, if  $CV(k_{bind}) < 1$ , the analytical solution is robust. In that the standard error of the mean is smaller than  $M_P/3$ . Notably, for such CV, the strength of the



Table 3. Parameter values inferred for each model.

Interference model	Inferred parameter values	Average $R^2$	Average $R^2$
		( $M, CV^2, S$ ) 1X medium	( $M, CV^2, S$ ) 0.5X medium
Exponential 1	$k_{bind}[R] = 1.09 \times 10^{-2} \text{ s}^{-1} \times (\text{cell vol})^{-1}$ $k_{bind} = 7.53 \times 10^{-5} \text{ s}^{-1}$ $k_{unbind} = 0.84 \text{ s}^{-1}$ $k^{max} = 677.7 \text{ s}^{-1}$ $b_1 = 5.08 \times 10^{-2} \text{ bp}^{-1}$	0.21 (Fig 6A–6C)	0.09 (Fig 6D–6F)
Exponential 2	$k_{bind}[R] = 9.71 \times 10^{-3} \text{ s}^{-1} \times (\text{cell vol})^{-1}$ $k_{bind} = 6.74 \times 10^{-5} \text{ s}^{-1}$ $k_{unbind} = 0.80 \text{ s}^{-1}$ $k^{max} = 554.8 \text{ s}^{-1}$ $b_1 = 7.92 \times 10^{-8} \text{ bp}^{-1}$ $b_2 = 1.47 \times 10^{-3} \text{ bp}^{-2}$	0.25 (Fig 6A–6C)	0.12 (Fig 6D–6F)
Step	$k_{bind}[R] = 6.62 \times 10^{-3} \text{ s}^{-1} \times (\text{cell vol})^{-1}$ $k_{bind} = 4.60 \times 10^{-5} \text{ s}^{-1}$ $k_{unbind} = 0.49 \text{ s}^{-1}$ $k^{max} = 313.4 \text{ s}^{-1}$ $L = 35.11 \text{ bp}$ (by best fitting, which corresponds to 35 bp)	0.35 (Fig 6A–6C)	0.15 (Fig 6D–6F)
zero order	$k_{bind}[R] = 4.63 \times 10^{-3} \text{ s}^{-1} \times (\text{cell vol})^{-1}$ $k_{bind} = 3.22 \times 10^{-5} \text{ s}^{-1}$ $k_{unbind} = 0.57 \text{ s}^{-1}$ $k^{max} = 6.48 \text{ s}^{-1}$	-0.007 (Fig 6A–6C)	-0.12 (Fig 6D–6F)

<https://doi.org/10.1371/journal.pcbi.1009824.t003>

two paired promoters would have to differ unrealistically by more than 2000%, on average (Table I in the S3 Appendix). Thus, we find the analytical solution to be reliable.

From our estimation of  $k_p$ , we further estimated a protein-to-RNA ratio,  $\frac{M_p}{M_{RNA}} = \frac{k_p}{k_{pd}}$ . From Eq 8 and Table 2, we find that  $\frac{k_p}{k_{pd}} \sim 1418$  in both media, which agrees with previous estimations (~1832 in [27]).

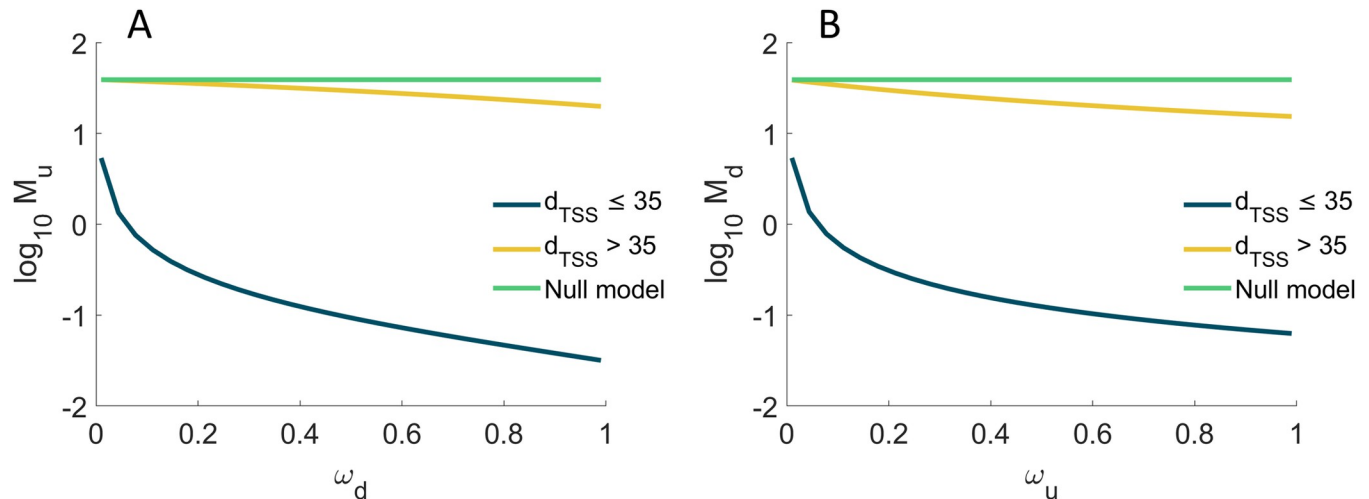
Next, we used the fitted model to predict (using Eqs 8 to 11) the influence of promoter occupancy ( $\omega$ ) on the  $M_p$ ,  $CV_p^2$  and  $S_p$  of upstream and downstream promoters. We set  $d_{TSS}$  to 20 bp to represent promoters where  $\leq 35$ , and to 100 bp to represent promoters with  $d_{TSS} > 35$ . Then, for each cohort, we changed  $\omega$  from 0.01 to 0.99 (i.e., nearly all possible values). In addition, we estimated these moments when  $k_{occlusion}$ ,  $k_{occupy}$ , and  $\omega$  are all set to zero (i.e., the two promoters do not interfere), for comparison.

From Fig 7, a pair of tandem promoters can produce less proteins than a single promoter with the same parameter values, if  $d_{TSS} \leq 35$ , which makes occlusion possible. Meanwhile, if  $d_{TSS} > 35$ , tandem promoters can only produce protein numbers in between the numbers produced by one isolated promoter and the numbers produced by two isolated promoters. In no case can two interfering tandem promoters produce more than two isolated promoters with equivalent parameter values. I.e., according to the model, the interference between tandem promoters cannot enhance production.

Meanwhile, the kinetics of the upstream (Fig 7A and panel A of Fig O in the S2 Appendix) and downstream promoters (Fig 7B and panel B of Fig O in the S2 Appendix) only differ in that the downstream promoter is more responsive to  $\omega$ .

Finally, consider that the model predicts that transcription interference should occur in tandem promoters, either due to occlusion if  $d_{TSS} \leq 35$  occupancy or due to occupancy of the downstream promoter if  $d_{TSS} > 35$ . Meanwhile, in single promoters, neither of these phenomena occurs. Thus, on average, two single promoters should produce more RNA and proteins than a pair of tandem promoters of similar strength. Using the genome wide data from [28] on





**Fig 7. Mean protein numbers produced as a function of other promoter's occupancy.**  $M_p$  of the single-cell distribution of the number of proteins produced (A) by the upstream promoter alone, and (B) by the downstream promoter alone. Results are shown as a function of the fraction of times that the upstream ( $0.01 \leq \omega_u \leq 0.99$ ) and the downstream ( $0.01 \leq \omega_d \leq 0.99$ ) promoter are occupied by RNAP. The null model is estimated by setting  $k_{occlusion}$ ,  $k_{occupy}$ , and  $\omega$  to zero.

<https://doi.org/10.1371/journal.pcbi.1009824.g007>

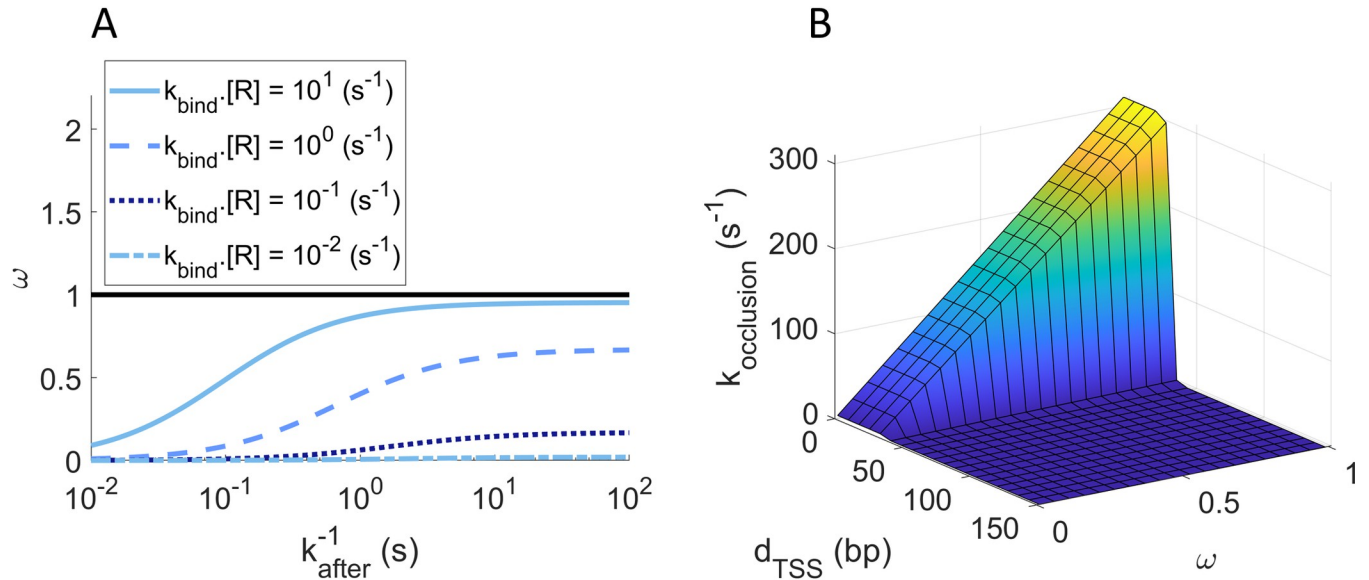
protein expression levels during exponential growth we estimated the double of the mean expression level (it equals 183.8) of genes controlled by single promoters (section ‘Selection of natural genes controlled by single promoters’ in the [S1 Appendix](#)). Meanwhile, also using data from [28], the mean expression level of genes controlled by tandem promoters equals 148 (estimated from the 26 that they have reported on), in agreement with the hypothesis. Nevertheless, this data is subject to external variables (e.g., TF interference). A definitive test would require the use of synthetic constructs, lesser affected by external influences.

### Regulatory parameters of promoter occupancy and occlusion

Since the occupancy,  $\omega$ , of each of the tandem promoters is responsible for transcriptional interference by occlusion and by RNAPs occupying the downstream promoter, we next explored the biophysical limits of  $\omega$ . Eqs 6A and 6B define the occupancies of the upstream and downstream promoters,  $\omega_u$  and  $\omega_d$ , respectively. For simplicity, here we refer to both of them as  $\omega$ . Fig 8A shows that  $\omega$  increases with the rate of RNAP binding ( $k_{bind} \cdot [R]$ ), but only within a certain range of (high) values of the time from binding to elongating ( $k_{after}^{-1}$ ). I.e., RNAPs need to spend a significant time in OC, if they are to cause interference, which is expected. Similarly,  $\omega$  changes with  $k_{after}^{-1}$ , but only for high values of  $k_{bind} \cdot [R]$ . I.e., if it's rare for RNAPs to bind, the occupancy will necessarily be weak.

In detail, from Fig 8A,  $\omega$  can change significantly within  $10^{-2} < k_{bind} \cdot [R] < 10 \text{ s}^{-1}$  and  $10^{-2} < k_{after}^{-1} < 10^2 \text{ s}$ . For these ranges, we expect RNA production rates ( $k_p$ , Eqs 5A, 5B, 6B, 7 and 9) to vary from  $\sim 10^{-5}$  (if  $d_{TSS} \leq 35$ ) and  $\sim 10^{-4}$  (if  $d_{TSS} > 35$ ) until  $10 \text{ s}^{-1}$ . In agreement, in *E. coli*, promoters have RNA production rates from  $\sim 10^{-3}$  to  $10^{-1} \text{ s}^{-1}$  when induced [20–21,39,50–51] and  $\sim 10^{-4}$  to  $10^{-6} \text{ s}^{-1}$  when non-fully active [28]. Thus,  $\omega$  can differ within realistic intervals of parameter values.

Next, we estimated  $k_{occlusion}$ , the rate at which a promoter occludes the other as a function of  $d_{TSS}$  and  $\omega$  using Eqs 6A and 6B.  $k^{max}$  is shown in Table 3. To model  $I(d_{TSS})$  we used the step function in Table 1. Overall,  $k_{occlusion}$  changes linearly with  $\omega$ , when and only when  $d_{TSS} \leq 35$  (Fig 8B).



**Fig 8. Promoter occupancy  $\omega$  estimated for the step model.** (A)  $\omega$  as a function of the rate constant for a free RNAP to bind to the unoccupied promoter ( $k_{bind} \cdot [R]$ ) and of the time for that RNAP to start elongation after commitment to OC,  $k_{after}^{-1}$ . The horizontal black line at  $\omega = 1$ , is the maximum fraction of time that the promoter can be occupied (i.e., the maximum promoter occupancy). (B)  $k_{occlusion}$  plotted as a function of  $\omega$  and  $d_{TSS}$ . Since  $k_{occlusion}$  increases with  $\omega$  if and only if  $d_{TSS} \leq 35$ , it renders the simultaneous occupation of both TSS's impossible.

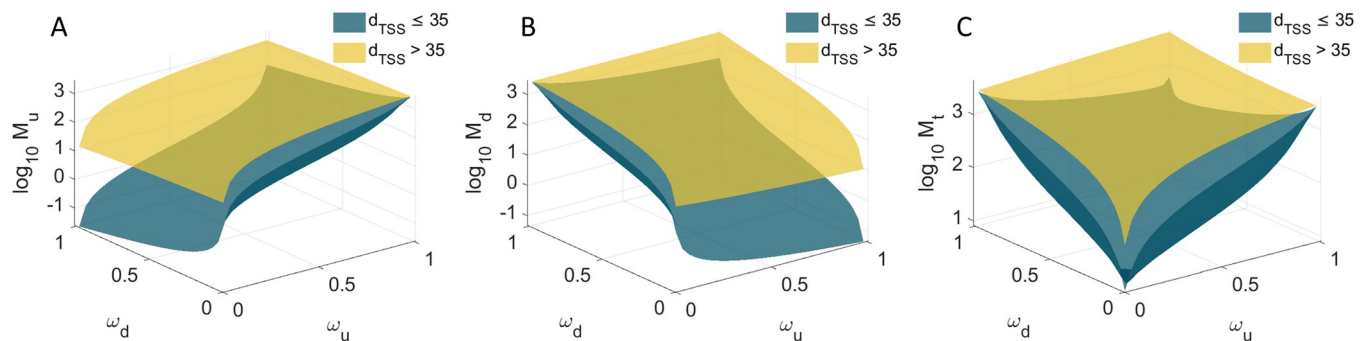
<https://doi.org/10.1371/journal.pcbi.1009824.g008>

### State space of the single cell statistics of protein numbers of tandem promoters

We next studied how much the single-cell statistics of protein numbers ( $M_p$ ,  $CV_p^2$ , and  $S_p$ ) of the upstream, 'u', and downstream, 'd', promoters changes with  $\omega_u$ ,  $\omega_d$ , and  $d_{TSS}$ . Here,  $\omega_u$  and  $\omega_d$  are increased from 0 to 1 by increasing the respective  $k_{bind}$  (Eqs 6A and 6B).

From Fig 9A, if  $d_{TSS} \leq 35$  bp, reducing  $\omega_d$  while also increasing  $\omega_u$  is the most effective way to increase  $M_u$ , since this increases the number of RNAPs transcribing from the upstream promoter that are not hindered by RNAPs occupying the downstream promoter. If  $d_{TSS} > 35$  bp, the occupancy the downstream promoter,  $\omega_d$ , becomes ineffective.

Oppositely, from Fig 9B, if  $d_{TSS} \leq 35$  bp, increasing  $\omega_d$  while also decreasing  $\omega_u$  is the most effective way to increase  $M_d$  since this increases the number of RNAPs transcribing from the



**Fig 9. Mean protein expression as a function of both promoters' occupancy.** Expected mean protein numbers due to the activity of: (A) the upstream promoter alone, (B) the downstream promoter alone, and (C) both promoters.  $M_p$  is shown as a function of the fraction of times that the upstream ( $0 \leq \omega_u \leq 1$ ) and the downstream ( $0 \leq \omega_d \leq 1$ ) promoters are occupied by RNAP, when  $d_{TSS} > 35$  (yellow) and  $d_{TSS} \leq 35$  (dark green) bp.

<https://doi.org/10.1371/journal.pcbi.1009824.g009>

downstream promoter does not interfere by RNAPs elongating from the upstream promoter. If  $d_{TSS} > 35$  bp, the occupancy the upstream promoter,  $\omega_u$ , becomes ineffective.

Finally, from Fig 9C, regardless of  $d_{TSS}$ , for small  $\omega_d$  and  $\omega_u$ , as the occupancies increase,  $M_t$  increases quickly and in a non-linear fashion. However, as both  $\omega_d$  and  $\omega_u$  reach high values,  $M_t$  decreases for further increases, if  $d_{TSS} \leq 35$  bp. Instead, if  $d_{TSS} > 35$  bp,  $M_t$  appears to saturate.

From Fig P in the S2 Appendix,  $CV_p^2$  and  $S_p$  behave inversely to  $M_p$ .

Relevantly, in all cases, the range of predicted protein numbers (Fig 9C) are in line with the empirical values ( $\sim 10^{-1}$  to  $10^3$  proteins per cell) (Fig 4D).

## Discussion

*E. coli* genes controlled by tandem promoters have a relatively high mean conservation level (0.2, while the average gene has 0.15, with a p-value of 0.009), suggesting that they play particularly relevant biological roles (section ‘Gene Conservation’ in the S1 Appendix). From empirical data on single-cell protein numbers of 30 *E. coli* genes controlled by tandem promoters, we found evidence that their dynamics is subject to RNAP interference between the two promoters. This interference reduces the mean single-cell protein numbers, while increasing its  $CV^2$  and skewness, and can be tuned by  $\omega$ , the promoters’ occupancy by RNAP, and by  $d_{TSS}$ . Since both of these parameters are sequence dependent [21,31] the interference should be evolvable. Further, since  $\omega$  of at least some of these genes should be under the influence of their several input TFs, the interference has the potential to be adaptive.

We proposed models of the dynamics of these genes as a function of  $\omega$  and  $d_{TSS}$ , using empirically validated parameter values. In our best fitting model, transcription interference is modelled by a step function of  $d_{TSS}$  (instead of gradually changing with  $d_{TSS}$ ), since the only detectable differences in dynamics with changing  $d_{TSS}$  were between tandem promoters with  $d_{TSS} \leq 35$  and  $d_{TSS} > 35$  nucleotides (the latter cohort of genes having higher mean expression and lower variability). We expect that causes this difference tangible is the existence of the OC formation. In detail, the OC is a long-lasting DNA-RNAP formation that occupies that strict region of DNA at the promoter region [24,31]. As such, occlusion should share these physical features. Because of that, when  $d_{TSS} \leq 35$ , an RNAP bound to TSS always occludes the other TSS, significantly reducing RNA production. Meanwhile, if  $d_{TSS} > 35$ , interference occurs when an RNAP elongating from the upstream promoter is obstructed by an RNAP occupying the downstream promoter.

Meanwhile, contrary to  $d_{TSS}$ , if one considers realistic ranges of the other model parameters, it is possible to predict a very broad range of accessible dynamics for tandem promoter arrangements. This could explain the observed diversity of single-cell protein numbers as a function of  $d_{TSS}$  (Fig 6). At the evolutionary level, such potentially high range of dynamics may provide high evolutionary adaptability and thus, it may be one reason why genes controlled by these promoters are relatively more conserved.

One potentially confounding effect which was not accounted for in this model is the accumulation of supercoiling. Closely spaced promoters may be more sensitive to supercoiling buildup than single promoters [52–54]. If so, it will be useful to extend the model to include these effects [26]. Using such model and measurements of expression by tandem promoters when subject to, e.g. Novobiocin [55], may be of use to infer kinetic parameters of promoter locking due to positive supercoiling build-up.

Other potential improvements could be expanding the model to tandem arrangements other than I and II (Fig 1), to include a third form of interference (transcription elongation of a nearby gene).

One open question is whether placing promoters in tandem formation increases the robustness of downstream gene expression to perturbations (e.g., fluctuations in the concentrations of

RNAP or TF regulators). A tandem arrangement likely increases the robustness to perturbations which only influence one of the promoters. Another open question is why several of the 102 tandem promoters with arrangements I and II appeared to behave independently from their input TFs (according to the RNA-seq data), albeit having more input TFs (1.62 on average) than expected by chance (the average *E. coli* gene only has 0.95). As noted above, we hypothesize that these input TFs may become influential in conditions other than the ones studied here.

Here, we also did not consider any influence from the phenomenon of “RNAP cooperation” [56]. This is based on this being an occurrence in elongation, and we expect interactions between two *elongating* RNAPs to rarely affect the interference between tandem promoters [9]. However, potentially, it could be of relevance in the strongest tandem promoters.

Finally, a valuable future study on tandem promoters will require the use of synthetic tandem promoters (integrated in a specific chromosome location) that systematically differ in promoter strengths and nucleotide distances. This would allow extracting parameter values associated to promoter interference to create a more precise model than the one based on the natural promoters (which is influenced by TFs, etc). Similarly, measuring the strength of individual natural promoters would contribute to this effort.

Overall, our model, based on a significant number of natural tandem promoters whose genes have a wide range of expression levels, should be applicable to the natural tandem promoters not observed here (at least of arrangements I and II), including of other bacteria, and to be accurate in predicting the dynamics of synthetic promoters in these arrangements.

Currently, predicting how gene expression kinetics change with the promoter sequence remains challenging. Even single- or double-point mutants of known promoters behave unpredictably, likely because the individual sequence elements influence the OC and CC in a combinatorial fashion. Consequently, the present design of synthetic circuits is usually limited to the use of a few promoters whose dynamics have been extensively characterized (Lac, Tet, etc.). This severely limits present synthetic engineering.

We suggest that a promising methodology to create new synthetic genes with a wide range of predictable dynamics is to assemble well-characterized promoters in a tandem formation, and to tune their target dynamics using our model. Specifically, for a given dynamics, it is possible to invert the model and find a suitable pair of promoters with known occupancies and corresponding  $d_{TSS}$  (smaller or larger than 35), which achieve these dynamics. A similar strategy was recently proposed in order to achieve strong expression levels [57]. Our results agree and further expand on this by showing that the mean expression level can also be reduced and expression variability can further be fine-tuned.

Importantly, this can already be executed, e.g., using a library of individual genes whose expression can be measured [28]. From this library, we can select any two promoters of interest and arrange them as presented here, in order to obtain a kinetics of expression as close as possible to a given target. Note that these dynamics have a wide range, from weaker to stronger than that of either promoter (albeit no stronger than their sum, Fig 9C). Given the number of natural genes whose expression is already known and given the present accuracy in assembling specific nucleotide sequences, we expect this method to allow the rapid engineering of genes with desired dynamics with an enormous range of possible behaviours. As such, these constructs could represent a recipe book for the components of gene circuits with predictable complex kinetics.

## Materials and methods

Using information from RegulonDB v10.5 as of 30<sup>th</sup> of January 2020 [58], we started by searching natural genes controlled by two promoters (Section ‘Selection of natural genes

controlled by tandem promoters' in the [S1 Appendix](#)). Next, we studied their evolutionary conservation and ontology (Sections 'Gene conservation' and 'Gene Ontology' in the [S1 Appendix](#)) and analysed their local topological features within the TFN of *E. coli* (Section 'Network topological properties' in the [S1 Appendix](#)).

RNA-seq measurements were conducted in two points in time (Section 'RNA-seq measurements and data analysis' in the [S1 Appendix](#)), to obtain fold changes in RNA numbers of genes controlled by tandem promoters with arrangements I and II, their input TFs, and their output genes ([Fig 1](#)). We used this data to search for relationships between input and output genes.

Next, a model of gene expression was proposed, and reduced to obtain an analytical solution of the single-cell protein expression statistics of tandem promoters (Sections 'Derivation of mean protein numbers at steady state produced by a pair of tandem promoters' and ' $CV^2$  and skewness of the distribution of single-cell protein numbers of model tandem promoters' in the [S1 Appendix](#)). This analytical solution was compared to stochastic simulations conducted using the simulator SGNS2. (Section 'Stochastic simulations for the step inference model' in the [S1 Appendix](#)).

We collected single-cell flow-cytometry measurements of 30 natural genes controlled by tandem promoters (Section 'Flow-cytometry and data analysis' in the [S1 Appendix](#)) to validate the model. For this, first, from the original data, we subtracted the cellular background fluorescence (Section 'Subtraction of background fluorescence from the total protein fluorescence' in the [S1 Appendix](#)). Then, we converted the fluorescence intensity into protein numbers (Section 'Conversion of protein fluorescence to protein numbers in the [S1 Appendix](#)'). From this we obtained empirical data on  $M$ ,  $CV^2$ , and  $S$  of the single-cell distributions of protein numbers in two media (Sections 'Media and chemicals' and 'Strains and growth conditions' in the [S1 Appendix](#)). Flow-cytometry measurements were also compared to microscopy data, supported by image analysis (Section 'Microscopy and Image analysis' in the [S1 Appendix](#)), for validation.

Comparing the data from RegulonDB (30.01.2020) used here, with the most recent (21.07.2021), we found that the numbers of genes controlled by tandem promoters of arrangements I and II differed by ~4% (from 102 to 98). Regarding those whose activity was measured by flow-cytometry, this difference is ~3% (30 to 31). Globally, 163 TF-gene interactions differed (~3.4%) while for the 98 genes controlled by tandem promoters of arrangements I and II, only 10 TF-gene interactions differ (~2.7%). Finally, globally the numbers of TUs differed by ~1%, promoters by ~0.6%, genes by ~1%, and terminators by ~15% (which did not affect the genes studied, as they changed by ~4% only). These small differences should not affect our conclusions.

Finally, a data package is provided in Dryad [[59](#)] with flow-cytometry and microscopy data and codes used. The RNAseq data has been deposited in NCBI's Gene Expression Omnibus [[60](#)] and are accessible through GEO Series accession number GSE183139 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE183139>).

## Dryad DOI

[10.5061/dryad.bnzs7h4bs](https://doi.org/10.5061/dryad.bnzs7h4bs).

## Supporting information

**S1 Appendix. Extended Materials and Methods.**  
(DOCX)

**S2 Appendix. Supporting Figures.**

(DOCX)

**S3 Appendix. Supporting Tables.**

(DOCX)

**S4 Appendix. Supporting Results.**

(DOCX)

**S1 Table. Gene Ontology.** Overrepresentation tests using the PANTHER Classification System. List of biological processes which are overrepresented using Fisher's exact tests are shown. (Excel)

(XLSX)

**S2 Table. Protein statistics.** Statistics of single-cell distributions of protein fluorescence of genes controlled by tandem promoters as measured by flow-cytometry in 1X and 0.5X diluted M9 media conditions. (Excel)

(XLSX)

**S3 Table. Protein statistics.** Statistics of single-cell distributions of protein fluorescence of genes controlled by single promoter as measured by flow-cytometry in 1X M9 media condition. (Excel)

(XLSX)

**Acknowledgments**

The authors thank Jason Lloyd-Price for proof-reading and editing the text.

**Author Contributions**

**Conceptualization:** Vatsala Chauhan, Mohamed N. M. Bahrudeen, Cristina S. D. Palma, Andre S. Ribeiro.

**Formal analysis:** Vatsala Chauhan, Mohamed N. M. Bahrudeen.

**Funding acquisition:** Andre S. Ribeiro.

**Investigation:** Vatsala Chauhan, Mohamed N. M. Bahrudeen, Cristina S. D. Palma, Ines S. C. Baptista, Bilena L. B. Almeida, Suchintak Dash, Vinodh Kandavalli, Andre S. Ribeiro.

**Methodology:** Vatsala Chauhan, Mohamed N. M. Bahrudeen, Ines S. C. Baptista.

**Project administration:** Andre S. Ribeiro.

**Software:** Mohamed N. M. Bahrudeen, Ines S. C. Baptista.

**Supervision:** Andre S. Ribeiro.

**Writing – original draft:** Vatsala Chauhan, Mohamed N. M. Bahrudeen, Ines S. C. Baptista, Andre S. Ribeiro.

**Writing – review & editing:** Vatsala Chauhan, Mohamed N. M. Bahrudeen, Cristina S. D. Palma, Ines S. C. Baptista, Andre S. Ribeiro.

**References**

1. Herbert M, Kolb A, Buc H. Overlapping promoters and their control in *Escherichia coli*: the gal case. *Proc Natl Acad Sci U S A*. 1986; 83: 2807–2811. <https://doi.org/10.1073/pnas.83.9.2807> PMID: 3010319



2. Beck CF, Warren RA. Divergent promoters, a common form of gene organization. *Microbiol Rev.* 1988; 52: 318–326. <https://doi.org/10.1128/mr.52.3.318-326.1988> PMID: 3054465
3. Adachi N, Lieber MR. Bidirectional gene organization: a common architectural feature of the human genome. *Cell.* 2002; 109: 807–809. [https://doi.org/10.1016/s0092-8674\(02\)00758-4](https://doi.org/10.1016/s0092-8674(02)00758-4) PMID: 12110178
4. Trinklein ND, Aldred SF, Hartman SJ, Schroeder DI, Otillar RP, Myers RM. An abundance of bidirectional promoters in the human genome. *Genome Res.* 2004; 14: 62–66. <https://doi.org/10.1101/gr.1982804> PMID: 14707170
5. Shearwin KE, Callen BP, Egan JB. Transcriptional interference—a crash course. *Trends Genet.* 2005; 21: 339–345. <https://doi.org/10.1016/j.tig.2005.04.009> PMID: 15922833
6. Prescott EM, Proudfoot NJ. Transcriptional collision between convergent genes in budding yeast. *Proc Natl Acad Sci U S A.* 2002; 99: 8796–8801. <https://doi.org/10.1073/pnas.132270899> PMID: 12077310
7. Korbelt JO, Jensen LJ, von Mering C, Bork P. Analysis of genomic context: prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nat Biotechnol.* 2004; 22: 911–917. <https://doi.org/10.1038/nbt988> PMID: 15229555
8. Wei W, Xiang H, Tan H. Two tandem promoters to increase gene expression in *Lactococcus lactis*. *Bio-technol Lett.* 2002; 24: 1669–1672. <https://doi.org/10.1023/A:1020653417455>
9. Sneppen K, Dodd IB, Shearwin KE, Palmer AC, Schubert RA, Callen BP, et al. A mathematical model for transcriptional interference by RNA polymerase traffic in *Escherichia coli*. *J Mol Biol.* 2005; 346: 399–409. <https://doi.org/10.1016/j.jmb.2004.11.075> PMID: 15670592
10. Martins L, Mäkelä J, Häkkinen A, Kandhavelu M, Yi-Harja O, Fonseca JM, et al. Dynamics of transcription of closely spaced promoters in *Escherichia coli*, one event at a time. *J Theor Biol.* 2012; 301: 83–94. <https://doi.org/10.1016/j.jtbi.2012.02.015> PMID: 22370562
11. Horowitz H, Platt T. Regulation of transcription from tandem and convergent promoters. *Nucleic Acids Res.* 1982; 10: 5447–5465. <https://doi.org/10.1093/nar/10.18.5447> PMID: 6755394
12. Bordoy AE, Varanasi US, Courtney CM, Chatterjee A. Transcriptional Interference in Convergent Promoters as a Means for Tunable Gene Expression. *ACS Synth Biol.* 2016; 5: 1331–1341. <https://doi.org/10.1021/acssynbio.5b00223> PMID: 27346626
13. Palmer AC, Ahlgren-Berg A, Egan JB, Dodd IB, Shearwin KE. Potent transcriptional interference by pausing of RNA polymerases over a downstream promoter. *Mol Cell.* 2009; 34: 545–555. <https://doi.org/10.1016/j.molcel.2009.04.018> PMID: 19524535
14. Callen BP, Shearwin KE, Egan JB. Transcriptional Interference between Convergent Promoters Caused by Elongation over the Promoter. *Mol Cell.* 2004; 14: 647–656. <https://doi.org/10.1016/j.molcel.2004.05.010> PMID: 15175159
15. Hoffmann SA, Hao N, Shearwin KE, Arndt KM. Characterizing Transcriptional Interference between Converging Genes in Bacteria. *ACS Synth Biol.* 2019; 8: 466–473. <https://doi.org/10.1021/acssynbio.8b00477> PMID: 30717589
16. Masulis IS, Babaeva ZS, Chernyshov SV, Ozoline ON. Visualizing the activity of *Escherichia coli* divergent promoters and probing their dependence on superhelical density using dual-colour fluorescent reporter vector. *Sci Rep.* 2015; 5: 1–10. <https://doi.org/10.1038/srep11449> PMID: 26081797
17. Vogl T, Kickenweiz T, Pitzer J, Sturmberger L, Weninger A, Biggs BW, et al. Engineered bidirectional promoters enable rapid multi-gene co-expression optimization. *Nat Commun.* 2018; 9: 1–13. <https://doi.org/10.1038/s41467-017-02088-w> PMID: 29317637
18. Adhya S, Gottesman M. Promoter occlusion: Transcription through a promoter may inhibit its activity. *Cell.* 1982; 29: 939–944. [https://doi.org/10.1016/0092-8674\(82\)90456-1](https://doi.org/10.1016/0092-8674(82)90456-1) PMID: 6217898
19. Eszterhas SK, Bouhassira EE, Martin DIK, Fiering S. Transcriptional interference by independently regulated genes occurs in any relative arrangement of the genes and is influenced by chromosomal integration position. *Mol Cell Biol.* 2002; 22: 469–479. <https://doi.org/10.1128/MCB.22.2.469-479.2002> PMID: 11756543
20. Lloyd-Price J, Startceva S, Kandavalli V, Chandraseelan JG, Goncalves N, Oliveira SMD, et al. Dissecting the stochastic transcription initiation process in live *Escherichia coli*. *DNA Res.* 2016; 23: 203–214. <https://doi.org/10.1093/dnares/dsw009> PMID: 27026687
21. Lutz R, Lozinski T, Ellinger T, Bujard H. Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res.* 2001; 29: 3873–3881. <https://doi.org/10.1093/nar/29.18.3873> PMID: 11557820
22. McClure WR. Rate-limiting steps in RNA chain initiation. *Proc Natl Acad Sci U S A.* 1980; 77: 5634–5638. <https://doi.org/10.1073/pnas.77.10.5634> PMID: 6160577
23. Krummel B, Chamberlin MJ. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J Mol Biol.* 1992; 225: 239–250. [https://doi.org/10.1016/0022-2836\(92\)90918-a](https://doi.org/10.1016/0022-2836(92)90918-a) PMID: 1593619

24. deHaseth Pieter L., Zupancic Margaret L., Record M. Thomas. RNA Polymerase-Promoter Interactions: the Comings and Goings of RNA Polymerase. *J Bacteriol.* 1998; 180: 3019–3025. <https://doi.org/10.1128/JB.180.12.3019-3025.1998> PMID: 9620948
25. Greive SJ, von Hippel PH. Thinking quantitatively about transcriptional regulation. *Nat Rev Mol Cell Biol.* 2005; 6: 221–232. <https://doi.org/10.1038/nrm1588> PMID: 15714199
26. Palma CSD, Kandavalli V, Bahrudeen MNM, Minoia M, Chauhan V, Dash S, et al. Dissecting the in vivo dynamics of transcription locking due to positive supercoiling buildup. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms.* 2020; 1863: 194515. <https://doi.org/10.1016/j.bbagr.2020.194515> PMID: 32113983
27. Häkkinen A, Oliveira SMD, Neeli-Venkata R, Ribeiro AS. Transcription closed and open complex formation coordinate expression of genes with a shared promoter region. *J R Soc Interface.* 2019; 16: 20190507. <https://doi.org/10.1098/rsif.2019.0507> PMID: 31822223
28. Taniguchi Y, Choi PJ, Li G-W, Chen H, Babu M, Hearn J, et al. Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science.* 2010; 329: 533–538. <https://doi.org/10.1126/science.1188308> PMID: 20671182
29. Friedman LJ, Mumm JP, Gelles J. RNA polymerase approaches its promoter without long-range sliding along DNA. *Proc Natl Acad Sci U S A.* 2013; 110: 9740–9745. <https://doi.org/10.1073/pnas.1300221110> PMID: 23720315
30. Skinner GM, Baumann CG, Quinn DM, Molloy JE, Hoggett JG. Promoter Binding, Initiation, and Elongation by Bacteriophage T7 RNA Polymerase: A SINGLE-MOLECULE VIEW OF THE TRANSCRIPTION CYCLE\*. *J Biol Chem.* 2004; 279: 3239–3244. <https://doi.org/10.1074/jbc.M310471200> PMID: 14597619
31. McClure WR. Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem.* 1985; 54: 171–204. <https://doi.org/10.1146/annurev.bi.54.070185.001131> PMID: 3896120
32. Saecker RM, Record MT Jr, Dehaseth PL. Mechanism of bacterial transcription initiation: RNA polymerase—promoter binding, isomerization to initiation-competent open complexes, and initiation of RNA synthesis. *J Mol Biol.* 2011; 412: 754–771. <https://doi.org/10.1016/j.jmb.2011.01.018> PMID: 21371479
33. Mekler V, Kortkhonja E, Mukhopadhyay J, Knight J, Revyakin A, Kapanidis AN, et al. Structural Organization of Bacterial RNA Polymerase Holoenzyme and the RNA Polymerase-Promoter Open Complex. *Cell.* 2002; 108: 599–614. [https://doi.org/10.1016/s0092-8674\(02\)00667-0](https://doi.org/10.1016/s0092-8674(02)00667-0) PMID: 11893332
34. Margeat E, Kapanidis AN, Tinnefeld P, Wang Y, Mukhopadhyay J, Ebricht RH, et al. Direct Observation of Abortive Initiation and Promoter Escape within Single Immobilized Transcription Complexes. *Biophys J.* 2006; 90: 1419–1431. <https://doi.org/10.1529/biophysj.105.069252> PMID: 16299085
35. Hsu LM. Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta.* 2002; 1577: 191–207. [https://doi.org/10.1016/s0167-4781\(02\)00452-9](https://doi.org/10.1016/s0167-4781(02)00452-9) PMID: 12213652
36. Hsu LM. Promoter Escape by *Escherichia coli* RNA Polymerase. *EcoSal Plus.* 2008;3. <https://doi.org/10.1128/ecosalplus.4.5.2.2> PMID: 26443745
37. Henderson KL, Felth LC, Molzahn CM, Shkel I, Wang S, Chhabra M, et al. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proc Natl Acad Sci U S A.* 2017; 114: E3032–E3040. <https://doi.org/10.1073/pnas.1618675114> PMID: 28348246
38. Ponnambalam S, Busby S. RNA polymerase molecules initiating transcription at tandem promoters can collide and cause premature transcription termination. *FEBS Lett.* 1987; 212: 21–27. [https://doi.org/10.1016/0014-5793\(87\)81549-1](https://doi.org/10.1016/0014-5793(87)81549-1) PMID: 3542569
39. Kandavalli VK, Tran H, Ribeiro AS. Effects of  $\sigma$  factor competition are promoter initiation kinetics dependent. *Biochim Biophys Acta.* 2016; 1859: 1281–1288. <https://doi.org/10.1016/j.bbagr.2016.07.011> PMID: 27452766
40. Bremer H, Dennis P, Ehrenberg M. Free RNA polymerase and modeling global transcription in *Escherichia coli*. *Biochimie.* 2003; 85: 597–609. [https://doi.org/10.1016/s0300-9084\(03\)00105-6](https://doi.org/10.1016/s0300-9084(03)00105-6) PMID: 12829377
41. Patrick M, Dennis PP, Ehrenberg M, Bremer H. Free RNA polymerase in *Escherichia coli*. *Biochimie.* 2015; 119: 80–91. <https://doi.org/10.1016/j.biochi.2015.10.015> PMID: 26482806
42. Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y, et al. Noise in protein expression scales with natural protein abundance. *Nat Genet.* 2006; 38: 636–643. <https://doi.org/10.1038/ng1807> PMID: 16715097
43. Ju X, Li D, Liu S. Full-length RNA profiling reveals pervasive bidirectional transcription terminators in bacteria. *Nat Microbiol.* 2019; 4: 1907–1918. <https://doi.org/10.1038/s41564-019-0500-z> PMID: 31308523

44. Hausser J, Mayo A, Keren L, Alon U. Central dogma rates and the trade-off between precision and economy in gene expression. *Nat Commun.* 2019; 10: 1–15. <https://doi.org/10.1038/s41467-018-07882-8> PMID: 30602773
45. Lagarias JC, Reeds JA, Wright MH, Wright PE. Convergence Properties of the Nelder—Mead Simplex Method in Low Dimensions. *SIAM J Optim.* 1998; 9: 112–147. <https://doi.org/10.1137/S1052623496303470>
46. Maurizi MR. Proteases and protein degradation in *Escherichia coli*. *Experientia.* 1992; 48: 178–201. <https://doi.org/10.1007/BF01923511> PMID: 1740190
47. Koch AL, Levy HR. Protein turnover in growing cultures of *Escherichia coli*. *J Biol Chem.* 1955; 217: 947–957. Available: <https://www.ncbi.nlm.nih.gov/pubmed/13271454> PMID: 13271454
48. Rydenfelt M, Garcia HG, Cox RS 3rd, Phillips R. The influence of promoter architectures and regulatory motifs on gene expression in *Escherichia coli*. *PLoS One.* 2014; 9: e114347. <https://doi.org/10.1371/journal.pone.0114347> PMID: 25549361
49. Buchler NE, Gerland U, Hwa T. On schemes of combinatorial transcription logic. *Proc Natl Acad Sci U S A.* 2003; 100: 5136–5141. <https://doi.org/10.1073/pnas.0930314100> PMID: 12702751
50. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell.* 2005; 123: 1025–1036. <https://doi.org/10.1016/j.cell.2005.09.031> PMID: 16360033
51. Startceva S, Kandavalli VK, Visa A, Ribeiro AS. Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression. *Biochimica et Biophysica Acta (BBA)—Gene Regulatory Mechanisms.* 2019; 1862: 119–128. <https://doi.org/10.1016/j.bbagr.2018.12.005> PMID: 30557610
52. Rhee KY, Opel M, Ito E, Hung S p., Arfin SM, Hatfield GW. Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the *ilvYC* operon of *Escherichia coli*. *Proc Natl Acad Sci U S A.* 1999; 96: 14294–14299. <https://doi.org/10.1073/pnas.96.25.14294> PMID: 10588699
53. Jia J, King JE, Goldrick MC, Aldawood E, Roberts IS. Three tandem promoters, together with IHF, regulate growth phase dependent expression of the *Escherichia coli* *kps* capsule gene cluster. *Sci Rep.* 2017; 7: 1–11. <https://doi.org/10.1038/s41598-016-0028-x> PMID: 28127051
54. Yeung E, Dy AJ, Martin KB, Ng AH, Del Vecchio D, Beck JL, et al. Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Syst.* 2017; 5: 11–24.e12. <https://doi.org/10.1016/j.cels.2017.06.001> PMID: 28734826
55. Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell.* 2014; 158: 314–326. <https://doi.org/10.1016/j.cell.2014.05.038> PMID: 25036631
56. Epshtein V, Nudler E. Cooperation between RNA polymerase molecules in transcription elongation. *Science.* 2003; 300: 801–805. <https://doi.org/10.1126/science.1083219> PMID: 12730602
57. Li M, Wang J, Geng Y, Li Y, Wang Q, Liang Q, et al. A strategy of gene overexpression based on tandem repetitive promoters in *Escherichia coli*. *Microb Cell Fact.* 2012; 11: 19. <https://doi.org/10.1186/1475-2859-11-19> PMID: 22305426
58. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeda D, et al. RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Res.* 2019; 47: D212–D220. <https://doi.org/10.1093/nar/gky1077> PMID: 30395280
59. Chauhan V, Bahrudeen MNM, Palma CSD, Ines SCB, Almeida BLB, Dash S, et al. Analytical kinetic model of native tandem promoters in *E. coli*. *Dryad, Dataset.* <https://doi.org/10.5061/dryad.bnzs7h4b>
60. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.* 2002; 30: 207. <https://doi.org/10.1093/nar/30.1.207> PMID: 11752295