Tampere University

CRISTINA PALMA

Repression Mechanisms in Bacterial Transcription

Tampere University Dissertations 693

Tampere University Dissertations 693

CRISTINA PALMA

Repression Mechanisms in Bacterial Transcription

ACADEMIC DISSERTATION
To be presented, with the permission of
the Faculty of Medicine and Health Technology
of Tampere University,
for public discussion in the auditorium F115
of the Arvo building, Arvo Ylpön katu 34, Tampere,
on 8 November 2022, at 12 o'clock.

ACADEMIC DISSERTATION
Tampere University, Faculty of Medicine and Health Technology
Finland

Responsible supervisor and Custos	Professor Andre S. Ribeiro Tampere University Finland	
Pre-examiners	Professor Suckjoon Jun University of California San Diego USA	Associate Professor Rodrigo Reyes-Lamothe McGill University Canada
Opponent	Professor David Grainger University of Birmingham UK	

The originality of this thesis has been checked using the Turnitin OriginalityCheck service.

Cover design: Roihu Inc.

ISBN 978-952-03-2618-0 (print)
ISBN 978-952-03-2619-7 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
http://urn.fi/URN:ISBN:978-952-03-2619-7

Carbon dioxide emissions from printing Tampere University dissertations have been compensated.

PunaMusta Oy – Yliopistopaino
Joensuu 2022

ACKNOWLEDGMENTS

The work presented in this thesis is the result of the collaborative efforts of several people, and I would wholeheartedly like to express my gratitude to all of them.

First, I would like to thank Professor Andre Ribeiro for his constant support and guidance. Andre recruited me as a Master's student. Over the past five years, he has shared his expertise, resources, and advice with me. As a result, I have acquired the necessary skills to complete my doctorate and pursue new challenges in the future.

Second, I am deeply grateful for the financial support provided by the Finnish Academy of Science and Letters, by the Finnish Cultural Foundation, and by the City of Tampere. Furthermore, I would also like to thank the Finnish Foundation for Technology Promotion for awarding me an encouragement grant.

Noteworthy, to obtain this funding, I counted on the support of other mentors, including Professor José Fonseca, Professor Jonathan Massera, Samuel Oliveira, Ph.D., Vinodh Kandavalli, Ph.D., and Huy Tran, Ph.D. I thank them for their constant promptness in supporting my grant applications and for offering guidance for my future. I also want to thank Professor Alice Pereira and Professor Carla Quintão of the NOVA University of Lisbon, Portugal, whose impact on my academic journey was determinant to pursue a doctoral degree.

Third, I would like to thank the past and current members of the Laboratory of Biosystem Dynamics (LBD). I thank Sofia Startceva, Ramakanth Neeli-Venkata, Nádia Gonçalves, Marco Minoia, Leonardo Martins, Vatsala Chauhan, Rahul Jagadeesan, Suchintak Dash, Bilena Almeida, Inês Baptista, and Mohamed Bahrudeen. The LBD members and alumni are not only exceptional coworkers, but also dear friends, I will always admire. I hope to continue working in an environment where I am surrounded by equally positive and encouraging coworkers. Further, I would like to thank Henna Mattila, Ph.D., Toni Montonen, M.Sc., and Carita Malkamäki, for being always available to answer my questions and doubts.

In addition, I would like to thank the pre-examiners, Professor Suckjoon Jun of the University of California San Diego, USA, and Professor Rodrigo Reyes-Lamothe of McGill University, Canada, for their insightful comments. Also, I would like to thank Professor David Grainger from the University of Birmingham, UK, for agreeing to act as the opponent during the public examination of this thesis.

Next, I would like to thank my friends who stood the test of time and distance. Ana debated the ups and downs of this journey with me. Laura G. and Jéssica are examples of self-determination. Mariana and Margarida showed me that one should always aim higher. Laura R., Inês, and Janne were a home away from home. I must note that this is neither an ordered nor a comprehensive list, as I carry several other dear friends in my heart. I am sorry I cannot include you all.

Also, I would like to thank my biggest supporters: my Parents. I will never be able to express enough gratitude. They provided the means for me to study abroad and come to Finland. Further, they contributed to the emotional support that encouraged me. They are my biggest enthusiasts, always eager to hear my next plans. It is hard to not live physically near them, but I do get to live with them always looking out for me.

(In PT: Também gostaria de agradecer ao meu maior suporte ao longo da vida: os meus Pais. Não tenho palavras para descrever o quão grata sou por tudo o que eles fizeram por mim. Foi o seu suporte que me permitiu estudar no estrangeiro e vir para a Finlândia. Além disso, foi também o seu apoio emocional que me motivou. Os meus Pais são os meus maiores entusiastas, sempre ansiosos por ouvir os meus planos futuros. É muito difícil não viver perto dos meus Pais, mas torna-se mais fácil sabendo que, independentemente de onde estou, eles estarão sempre prontos para me apoiar e a torcer por mim.)

Finally, I would like to thank the family I found in Finland. Anne and Osmo, thank you for always welcoming me in your home and making me a part of every family event. Jussi, I cannot express how grateful I am for you. The end of this journey is your success too. I look forward to pursuing our future, which will certainly be filled with exciting adventures and brisket. You, too, are the light of my life.

Tampere, 29th of July, 2022

Cristina

ABSTRACT

Gene expression in bacteria is subject to strict regulation. This allows cells to timely tune their behavior to tackle challenging environmental conditions. For example, when subject to heat shock, *E. coli* induce the expression of σ^{32}, which triggers the synthesis of several proteins that counteract protein denaturation. Simultaneously, their membrane lipid composition is readjusted, to beneficially alter its fluidity. Such complex coordinated behaviors are made possible by a global regulatory network of gene expression.

Transcription factors (TFs) are one mode of gene regulation. These are specialized proteins expressed by genes to directly (via TF-DNA interactions) or indirectly (via TF-TF interactions) activate or repress RNA production from a specific gene(s).

Another influential dynamic regulator of gene expression is DNA supercoiling. In normal conditions, the DNA is kept negatively supercoiled. When the RNA polymerase (RNAP) elongates along the DNA it unwinds it. This causes positive supercoiling downstream the RNAP and negative supercoiling upstream it. When the supercoiling levels of the DNA are significantly disturbed, elongation rates are affected and, eventually, transcription initiation can be halted, until specialized proteins intervene. Since the DNA is organized in topologically constrained segments, these effects are not homogenously spread throughout the DNA. Namely, DNA regions with high rates of transcription will have significantly more positive supercoiling buildup (PSB).

Finally, specific spatial arrangements of regulatory elements in the DNA can also be influential, by allowing genes to co-express. These arrangements include operons and transcription units, terminator sequences, and closely spaced promoters. On the latter, while most pairs of neighboring promoters in the DNA are separated by distances long enough to make them dynamically independent, some pairs are sufficiently close to have interdependent dynamics.

In this project, from empirical, genome-wide data, as well as from single-cell, single-gene data, we studied regulatory mechanisms of gene expression, in optimal

and in non-optimal conditions. In detail, we delve into the kinetics of transcription when subject to repression by regulatory molecules as well as by changes in supercoiling levels, and when controlled by promoter pairs in tandem formation.

We first developed two methods to characterize the dynamics of transcription repression by TFs and PSB. In the former, we dissected the kinetic rates constants controlling TF regulation, while in the latter we showed that the dynamics of transcription locking due to PSB is promoter-strength dependent.

Next, we identified all native genes controlled by tandem promoters and studied how the distance between two promoters makes their transcription kinetics interdependent. We found evidence that, when the distance between the promoters is less than the length of DNA occupied by an RNAP in open complex formation, a phenomenon of promoter occlusion becomes influential. Otherwise, the only form of transcription interference is due to collision between RNAPs, causing a mild decrease in rate, along with a mild increase in noise in RNA production.

Finally, we studied the kinetics of natural genes, when repressed, following shifts in temperature during cold shock. To identify the nature of the mechanism responsible for their repression, we also subjected the same genes to the effects of Gyrase inhibition. Our study revealed that high sensitivity to supercoiling is one of the natural triggers of short-term responsiveness to cold shock.

In each of the aforementioned studies, we proposed new analytical and/or stochastic models. These models were designed to mimic how each mechanism operates and influences the single-cell statistics of RNA and/or protein expression. Subsequently, they were confronted to empirical data, to test hypotheses of which steps in transcription are most affected by the changing conditions.

We found that all the mechanisms of regulation studied were of similar nature, in that they all made possible an ON-OFF kinetics of RNA production. And while they were made possible by distinct physical processes, in all cases the reduction in expression rate occurred at the expense of increased noise.

Overall, our findings and models contribute to the knowledge on the mechanisms by which gene networks self-regulate and, thus, about how cells adapt to changing environments. Further, they provide insight on how to engineer synthetic gene circuits that make use of promoter-promoter interactions and/or PSB.

CONTENTS

List of Figures

List of Tables

ABBREVIATIONS

bp	Base pairs
CFP	Cyan Fluorescent Protein
CME	Chemical master equation
CSR	Cold shock repressed
CTD	C-terminal domain
CV^2	Squared coefficient of variation
DAPI	4',6-diamidino-2-phenylindole
DM	Direct method
DNA	Deoxyribonucleic acid
dTSS	Distance between transcription starting sites
E. coli	Escherichia coli
FISH	Single-molecule fluorescence *in situ* hybridization
GFP	Green fluorescent protein
HILO	Highly inclined and laminated optical sheet
IPTG	Isopropyl ß-D-1-thiogalactopyranoside
KDE	Kernel Density Estimation
MS2-GFP	MS2 coat protein fused to green fluorescent protein
nt	Nucleotides
NTP	Nucleoside triphosphate
ODE	Ordinary differential equation
P_{BAD}	Arabinose inducible *araBAD* promoter
P_{Lac}	Native Lac promoter (IPTG inducible)
$P_{Lac/ara-1}$	Synthetically engineered promoter containing elements from the native *Lac* promoter and from the native *BAD* promoter (IPTG and Arabinose inducible)
$P_{LacO3O1}$	Synthetic engineered promoter from the native *Lac* promoter by removing the O_2 operator site (IPTG inducible)
PSB	Positive supercoiling buildup
qPCR	Quantitative polymerase chain reaction
RBS	Ribosome binding site
RFP	Red fluorescent protein

RNA	Ribonucleic acid
RNAP	RNA polymerase
RNA-seq	RNA sequencing
RP_C	RNA polymerase in closed complex form
RP_O	RNA polymerase in open complex form
rRNA	Ribosomal RNA
SSA	Stochastic simulation algorithm
TEC	Transcription elongation complex
TF	Transcription factor
TI	Transcriptional interference
TopoI	Topoisomerase I
tRNA	Transfer RNA
TSS	Transcription starting site
YFP	Yellow fluorescent protein
[RNAP]	RNA polymerase concentration

ORIGINAL PUBLICATIONS

This thesis is a collection of four publications. Throughout this book, these are referred to as **Publication I**, **Publication II**, **Publication III**, and **Publication IV**. The publications are reproduced with the publishers' consent.

Publication I **C.S.D. Palma**, S. Startceva, R. Neeli-Venkata, M. Zare, N.S.M. Goncalves, J.M. Fonseca, S.M.D. Oliveira, and A.S. Ribeiro (2017) A strategy for dissecting the kinetics of transcription repression mechanisms. *EMBEC & NBC. IFMBE Proceedings*, 65. Springer, Singapore. DOI:10.1007/978-981-10-5122-7_274.

Publication II **C.S.D. Palma**, V. Kandavalli, M.N.M. Bahrudeen, M. Minoia, V. Chauhan, S. Dash, and A.S. Ribeiro (2020) Dissecting the in vivo dynamics of transcription locking due to positive supercoiling buildup. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1863(5), 194515. DOI:10.1016/j.bbagrm.2020.194515.

Publication III V. Chauhan*, M.N.M. Bahrudeen*, **C.S.D. Palma**, I. Baptista, B.L.B. Almeida, S. Dash, V. Kandavalli, and A.S. Ribeiro (2022) Analytical kinetic model of native tandem promoters in *E. coli*. *PLOS Computational Biology*, 18(1): e1009824. DOI:10.1371/journal.pcbi.1009824. *Equal contributions.

Publication IV S. Dash*, **C.S.D. Palma***, I.S.C. Baptista, M.N.M. Bahrudeen, B.L.B. Almeida, V. Chauhan, R. Jagadeesan, and A.S. Ribeiro (2021) Positive supercoiling buildup is a trigger of E. coli's short-term response to cold shock. *bioRxiv*. DOI:10.1101/2021.12.22.473827. *Equal contributions.[1]

[1] A revised version of this pre-print has later been published. The updated reference reads: S. Dash*, **C.S.D. Palma***, I.S.C. Baptista, B.L.B. Almeida, M.N.M. Bahrudeen, V. Chauhan, R. Jagadeesan and A.S. Ribeiro (2022) Alteration of DNA supercoiling serves as a trigger of short-term cold shock repressed genes of *E. coli*, *Nucleic Acids Research*, 50(15), 8512-8528. DOI: 10.1093/nar/gkac643. *Equal contributions.

Below is a description of the author's contribution to each publication.

In **Publication I**, the author performed all image and data analysis. The author also developed the model and its analytical solution. Finally, the author collaborated with A.S. Ribeiro and S. Startceva on the writing of the manuscript, which was then revised by all co-authors.

In **Publication II**, the author conceived the study with A.S. Ribeiro and designed the experiments with V. Kandavalli and M. Minoia. The author performed the image and data analysis. Further, the author conceived, implemented, and simulated the models with M.N.M. Bahrudeen. The author wrote the manuscript with A.S. Ribeiro, supported by all other co-authors.

In **Publication III**, the author contributed to the conception of the study, the author performed data analysis and conceived the model with M.N.M. Bahrudeen. The author contributed to the writing of the manuscript, supported by all other co-authors.

In **Publication IV**, the author conceived the study with A.S. Ribeiro and S. Dash. The author contributed to the design of the experiments and carried out the microscopy experiments with S. Dash. The author developed the data analysis pipeline and analyzed the empirical data. Further, the author conceived, implemented, and simulated the stochastic model. Finally, the author wrote the manuscript with S. Dash, which was then revised by all co-authors.

Publication II and III have been included by M.N.M. Bahrudeen in his Ph.D. dissertation.

1 INTRODUCTION

1.1 Background and Motivation

Bacterial infections have been linked to numerous intestinal infections and deaths (Beuchat, 2002; Rangel et al., 2005). Because of their minimal infective dose (~10 cells) and high virulence, some bacteria are extremely dangerous (Tilden et al., 1996). Therefore, it is important to understand their survival mechanisms, as this is a promising avenue for preventing diseases.

Adaptation to life-threatening conditions, such as cold shock, starvation, and antibiotics is essential for bacterial survival. A myriad of mechanisms has evolved to detect and respond to the changes (Cheung et al., 2003; Golding, 2011; López-Maury et al., 2008; Sleator & Hill, 2002; Yamanaka, 1999). These act at multiple cellular levels, from transcription and translation kinetics (Charlebois et al., 2018; Giuliodori et al., 2004; Oliveira et al., 2016), to membrane composition (Mansilla et al., 2004), enzymatic capacity (Zhang et al., 2019), changes in protein structure that generate phenotypic variants (Sokurenko et al., 1998, 1995), and activation of specific genetic 'motifs' (Milo et al., 2002).

The earliest stage at which gene expression can be controlled is transcription initiation (Dorman & Dorman, 2016). Moreover, several studies have shown that transcription initiation is the stage at which the majority of the regulatory processes of gene expression occur (Browning & Busby, 2004, 2016; Chamberlin, 1974; Chong et al., 2014; McClure, 1985; McLeod & Johnson, 2001; Ruff et al., 2015).

Transcription initiation is a complex, multi-step process that begins with the binding of RNAP to the promoter region (McClure, 1985; Saecker et al., 2011). At first, the structure between RNAP and the DNA is in a 'closed complex form' (Alberts et al., 2002; McClure, 1985). Next, a chain of isomerization steps leads to an open complex structure (Rogozina et al., 2009; Saecker et al., 2011). After the open complex formation, the RNAP breaks contact with the promoter elements,

entering the elongation phase (Alberts et al., 2002). Controlling the kinetics of these steps should allow influencing the speed of production of RNAs.

Strategies have been proposed to dissect the kinetics of the steps of transcription initiation from *in vitro* (McClure, 1985) and, more recently, *in vivo* measurements of transcription rates for various RNAP concentrations (Lloyd-Price et al., 2016). These allowed estimating the mean time-length of open and closed complex formation and showed that these steps are the most rate-limiting in RNA production.

Several repression mechanisms studied over the past five decades are based on the binding of TFs and other regulatory molecules to the DNA (Oppenheim et al., 2005; Razo-Mejia et al., 2018; Riggs et al., 1970; Sanchez et al., 2011; Almeida et al., 2022). Also, the regulatory molecules that can alter the kinetics of the rate limiting steps of transcription initiation (Lloyd-Price et al., 2016; Lutz et al., 2001), affect not only mean RNA production rate, but also the noise in this process (Mäkelä et al., 2017).

Other physical processes can also act as mechanisms of gene regulation. For example, DNA topological changes have been identified to have a major regulatory role on bacterial gene expression (Dorman & Dorman, 2016). As an example, a recent study (Chong et al., 2014) showed that one contributor to phenotypic variability in prokaryotic gene expression is the generation of positive supercoiling buildup (PSB) during transcription (Liu & Wang, 1987; Samul & Leng, 2007; Tsao et al., 1989; Wu et al., 1988). Specifically, PSB is of significance since it leads to promoter locking, directly enhancing noise in transcription (Chong et al., 2014).

The fact that stochastic promoter locking enhances noise in gene expression (Chong et al., 2014; Golding et al., 2005), which is selectively advantageous in certain conditions (Levin, 2003; McAdams & Arkin, 1999; Ozbudak et al., 2002; Rao et al., 2002; Thattai & Van Oudenaarden, 2004), supports the hypothesis that DNA supercoiling acts as a gene regulatory mechanism. Thus, it is important to characterize the dynamics of promoter locked states.

Other studies (Rhee et al., 1999; Shearwin et al., 2005; Yeung et al., 2017) have shown that transcriptional interference between closely spaced promoters results in distinctive transcriptional dynamics. Given the widespread presence of closely

spaced promoters in the DNA of organisms (Adachi & Lieber, 2002; Trinklein et al., 2004), we proposed that such specific DNA spatial organization is not random and serves as means to produce complex transcription kinetics, by locking or inhibiting one of the transcription start sites (TSSs).

In summary, we studied gene regulatory repression mechanisms that allow bacteria to respond and adapt to internal and external changes. For this, we used experimental and theoretical methods. We expect that our models of these phenomena will be broadly applicable to a wide range of stresses causing genome wide repression phenomena.

1.2 Thesis Objectives

In this thesis, we examine how gene networks regulate themselves under optimal and non-optimal settings. We focus on the dynamics of natural repression mechanisms of gene expression, and their outcome on the single-cell statistics of RNA and/or protein numbers. Specifically, this thesis is based on three objectives:

I: Dissect the effects of transcription factors with repressing effects on the kinetics of transcription. The quantification method proposed must be of broad applicability to allow surveying the time spent by several promoters in a repressed state.

II: Dissect the effects of supercoiling buildup on the kinetics of transcription. For this, we dissected the time spent in locked state due to supercoiling accumulation and estimated the rate constants to implement a realistic model. Further, based on a recent work that suggested that cold shock responsive genes may have atypical sensitivity to supercoiling (Oliveira et al., 2019) we investigated which and by how much are genes repressed by supercoiling during cold shock.

III: Dissect the effects of the distance between tandem promoters on their coordinated dynamics of transcription. For this, we quantified to what extent gene regulation due to close proximity causes repression and, thus, generates coordinated activities.

To achieve these objectives, we performed measurements using RNA-seq, flow-cytometry and/or microscopy and, from the data, we studied the kinetics of negatively responsive genes of *E. coli* under specific conditions, and developed stochastic models to mimic their behavior, quantitatively, at single-cell level.

Objective I was completed, and the findings were presented in **Publication I**. From completing Objective II resulted **Publication II** and **Publication IV**. Finally, **Publication III** resulted from the completion of Objective III.

1.3 Thesis Outline

The present chapter 1 of this thesis introduces the research area, providing the reader with the context and rationale for this study. Chapter 2 provides a broader biological context, with an emphasis on current knowledge on transcription and its regulatory mechanisms using *E. coli* as a model organism; Chapter 3 provides theoretical background on the simulation algorithms and modeling methodologies employed in the publications; Chapter 4 contains a brief description of the methodology used for collecting the empirical data and for processing data; Chapter 5 includes a summary of the results and subsequent conclusions; Finally, Chapter 6 is an analysis of the limitations of these studies. Finally, it is discussed the impact of the results in the fields of synthetic biology and microbiology.

2 BIOLOGICAL BACKGROUND

This chapter serves as a general introduction to the fundamental biological concepts underpinning this thesis. First, the model organism used in this study is introduced. This is followed by a description of the central dogma of molecular biology and its players. Next, we provide a detailed description of the multi-step processes unique to transcription and to translation, respectively, along with some of their regulatory mechanisms. Finally, we briefly discuss the effects of temperature on these processes and the concept of noise in gene expression.

2.1 *Escherichia coli* as a Model Organism

In 1884, Theodor Escherich, discovered *Bacterium coli* while studying the gut microbes of infants (Escherich, 1884, 1988). *Bacterium coli* is now known as *Escherichia coli* (*E. coli*) (Figure 1A). Due to its versatile strains and fast growth in several media conditions, it became the organism of choice in many fundamental studies, such as of the genetic code (Crick et al., 1961), transcription (Stevens, 1960), and gene regulation (Jacob et al., 1960). The acquired knowledge, along with the development of tools and techniques to study *E. coli,* have allowed for the study of this organism in-depth, providing insight into microbial phenomena. Given this, *E. coli* has been commonly used for scientific studies and pharmaceutical production (Baeshen, 2015), and it is therefore considered the most important model organism in biology (Zimmer, 2008). For these reasons, *E. coli* was the organism selected for further study in **Publications I-IV.**

2.2 DNA, RNA and Proteins

The discovery of the DNA (deoxyribonucleic acid) in 1869 by Johannes Miescher (Dahm, 2005) was a benchmark in Molecular Biology. From that moment onward, several key concepts were established. In 1944, it was shown that DNA is the hereditary material (Avery et al., 1944) and later, in 1953, Watson and Crick characterized its structure (Watson & Crick, 1953).

The DNA structure consists of two anti-parallel chains, where each strain is composed of nucleotides. A single nucleotide is composed of three components; these include a phosphate, a sugar, and a nitrogen-containing base. The bases can be Adenine (A), Guanine (G), Cytosine (C) or Thymine (T). The nucleotides are linked to each other through the phosphate groups and the DNA base pairs (bp) are connected by hydrogen bonds. These bonds can only pair Adenine to Thymine (A-T) and Cytosine to Guanine (C-G) (Chargaff et al., 1951). It is the linear sequence of bases that encode the genetic information of cells. Given the weak stability of hydrogen bonds, other molecules can interact with the DNA, performing tasks including reading the genetic information stored in the genome.

The central dogma of Molecular Biology (Crick, 1970) frames the transfer of the genetic information (Figure 1B). Typically, information flows from DNA through RNA to proteins. Nevertheless, in special cases (e.g., viruses), information is also transferred from RNA to DNA, RNA to RNA (also known as RNA replication), or directly from DNA to proteins. So far, there is no known process capable of directly synthesizing DNA from proteins.

The RNA molecule has similarities with the DNA molecule. In the RNA, the bases are the same as in the DNA, except uracil (U), which replaces thymine. Also, RNAs are single-stranded and serve different functions, such as mRNA (messenger RNA), tRNA (transfer RNAs), rRNA (ribosomal RNA), siRNA (small interfering RNA), etc.

Proteins are chains of amino acid molecules (also known as polypeptides). A protein is formed by a series of up to 20 different amino acids. It is the sequence of amino acids that determines a protein three-dimensional structure and function. Proteins are usually known as the 'building blocks' of cells, as they participate in every cellular process including, but not limited to, cellular repair, division, and metabolic reactions.

In **Publication I** we focused on specific mRNAs. Then, in **Publications II-IV**, we studied specific mRNAs and proteins. Throughout this thesis, for simplicity, we referred to mRNAs as 'RNAs'.

2.3 Gene Expression in *Escherichia coli*

The DNA in *E. coli* is a single circular chromosome containing over 4000 genes (Blattner et al., 1997). In general, the chromosome contains the genes necessary for cell survival and division. Extra DNA, containing additional genes, can exist. These are termed 'plasmids'. Plasmids are usually circular and contain genes that can code for evolutionary advantages like antibiotic resistance (Eliasson et al., 1992).

Gene expression consists of two sequential processes: transcription and translation. Transcription is the process where a gene is transcribed into a corresponding RNA (McClure, 1985). In translation, the RNA is the template to synthesize a corresponding chain of amino acids, that will form a protein (Ramakrishnan, 2002).

In prokaryotes, there is no nuclear membrane, therefore transcription and translation are coupled (i.e., translation can be initiated before transcription is completed) (Miller et al., 1970). Nevertheless, for many genes, the RNA is the final product of gene expression. These non-coding RNAs, such as microRNAs, usually have a regulatory role (Alberts et al., 2002).

In prokaryotes, most genes are organized in operons (Figure 1C). The structure of an operon consists mainly of three elements: a promoter(s), an operator site (s), and structural genes (Osbourn & Field, 2009). The promoter sequence is the region of DNA recognized by RNA polymerase (RNAP) to start transcription. The operator is the site to which most regulatory molecules bind in order to exert control of the expression rate.

In 1960, Francois Jacob and Jacques Monod studied the gene expression mechanisms of the *lactose* (*lac*) operon (Jacob & Monod, 1961). This was pioneering research that contributed to a deeper understanding of the regulation of protein production. The lac operon includes three genes: *lacZ*, *lacY*, and *lacA* and three operator sites: O_1, O_2, and O_3. The operator sites, along with TFs, regulate the expression of the operon.

In **Publications I and II,** we used synthetic constructs, engineered from the *E. coli's* native *lac* promoter by removing the O_2 repressor binding site.

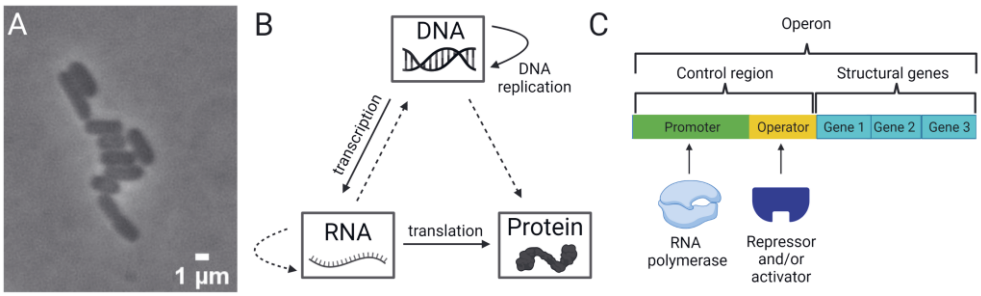

Figure 1. Gene expression in *E. coli*. **(A)** Phase-contrast image of *E. coli* cells. **(B)** The central dogma of Molecular Biology. DNA replication is the process where both DNA strands can work as templates and generate a daughter DNA molecule. In transcription, an RNA is formed from the information stored in the DNA. Translation is the process in which proteins are formed from the RNAs. Dashed arrows show rare cases of genetic information flow observed in e.g., viruses or specific laboratory conditions. **(C)** Operon structure. An operon has three components: a promoter, an operator site, and structural genes. The RNAP binds to the promoter whereas TFs (activators and/or repressors) bind to the operator sites. Generated using Biorender.

2.4 Mechanisms of Transcription

The transcription process has three steps (Alberts et al., 2002): initiation, elongation, and termination (Figure 2A). Initiation starts when the RNAP recognizes the promoter. RNAP is a key multi-unit enzyme, with five subunits (Figure 2B). Four of these subunits form the core enzyme: two alpha (α), a beta (β), a beta prime (β'), and an omega (ω) subunit (Haugen et al., 2008). The α subunits are involved in RNAP assembly and transcriptional regulation; they engage with the promoter site using both specific- and non-specific-sequence interactions. The α-subunits also interact with TFs (Finney et al., 2002; Ross et al., 1993). The β and β' are the largest subunits and contain part of the active center responsible for the synthesis of RNAs (Haugen et al., 2008; Mekler et al., 2002). The ω is the smallest subunit and is also involved in the RNAP assembly (Gunnelius et al., 2014).

The RNAP core enzyme synthesizes RNA from a DNA template, but it is unable to initiate transcription. For the RNAP core enzyme to bind specifically to the promoter and initiate transcription, it needs to bind to a fifth σ-subunit (Murakami et al., 2002; Young et al., 2002). An RNAP with five subunits is in the holoenzyme form. *E. coli* has seven different σ factors (Maeda et al., 2000). The primary σ factor is σ^{70}, which is specific to genes responsible for regulating basic cell functions in

optimal conditions (Feklístov et al., 2014). Other σ factors are expressed under specific stresses (Helmann & Chamberlin, 1988) and can activate a specific gene cohort whose task is to lessen the effects of adverse conditions.

After promoter recognition, the RNAP starts unwinding the DNA. Following the DNA unwinding and promoter escape, the transcription elongation begins. It is during transcription elongation that the RNA is synthesized. During this process, the RNAP slides along the DNA strand (3' to 5' direction), and nucleotides are added to the growing polynucleotide chain (Alberts et al., 2002). When reaching the termination site, both the RNAP and RNA are released.

2.4.1 Transcription Initiation

In prokaryotes, transcription initiation includes three main steps: Promoter recognition and binding, isomerization, and promoter escape (McClure, 1985; Saecker et al., 2011). To initiate transcription, the RNAP holoenzyme must first find and bind to the promoter. The consensus sequence between -10 and -35 positions upstream of the transcription start site (TSS) defines a promoter region. The TSS is the DNA site in which the first RNA nucleotide is transcribed (alternatively referred to as the '+1 site').

First, the holoenzyme slides rapidly across the DNA (Dangkulwanich et al., 2014; Hammar et al., 2012). As it does this, the holoenzyme adheres only weakly to the non-specific sequences of the DNA until it dissociates from it or a TSS is found. After finding the TSS, the RNAP makes specific contacts with the bases at the -35 and the -10 region. At this step, the structure formed by the RNAP and DNA is a 'closed complex (RP_C) form' since the DNA remains as a double stranded helix structure (Alberts et al., 2002; McClure, 1985).

Next, a σ factor triggers the destabilization of the DNA double helix, exposing a stretch of nucleotides on each DNA strand (~12bp), while forming a transcription bubble (Figure 2C) (Borukhov & Nudler, 2008). When this is conducted by e.g., σ^{70} and σ^{38} , it does not require ATP since it is more energetically favorable than the previous state. This is not the case for e.g., σ^{54} RNAP holoenzymes. For these, additional protein factors and energy in the form of ATP or GTP is required for formation of transcriptionally competent promoter complexes (Borukhov & Severinov, 2002).

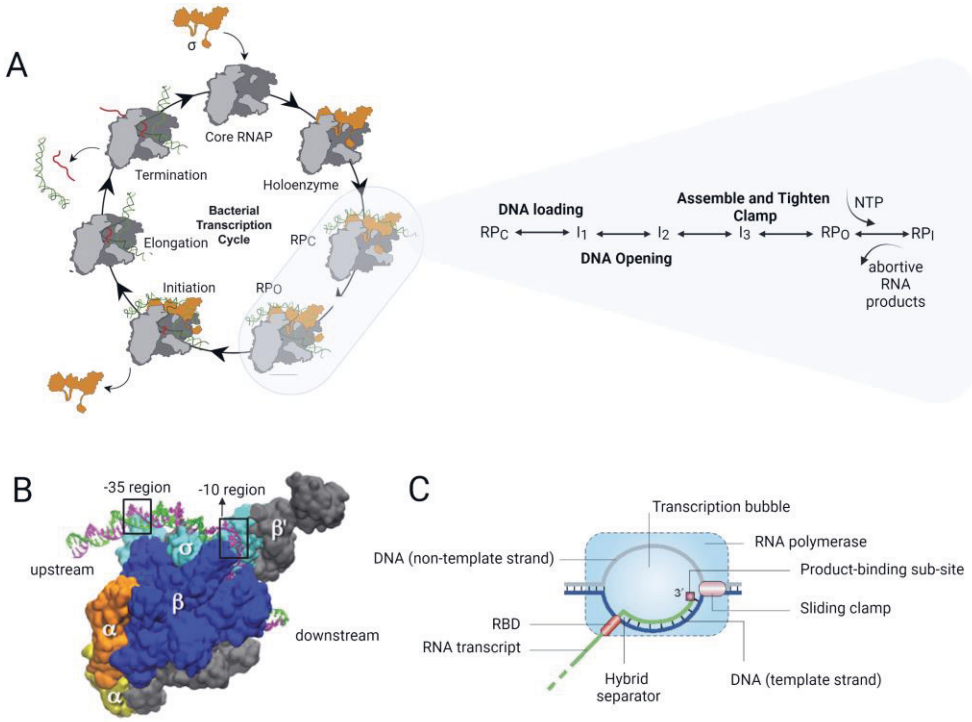

Figure 2. Transcription. **(A)** Schematic representation of the transcription cycle in *E. coli*. This cycle involves mainly 3 steps. First, the σ factor (orange) binds to the core RNAP subunits (in grey) and form the RNAP holoenzyme. Next, the holoenzyme finds and binds to the promoter forming the closed complex (RPc) complex. The RP_C goes through three intermediate isomerization steps (DNA loading, DNA opening/unwinding and assembly of the polymerase clamp) that lead to open complex (RPo) formation. Once the DNA is unwound NTPs can bind and form the initiating (RP_I) complex. Finally, the σ factor is released and elongation proceeds until reaching the termination sequence where the new RNA and the RNAP are released from the DNA (Greive & von Hippel, 2005). Picture adapted from (Tabib-Salazar et al., 2019) and (Saecker et al., 2011). **(B)** RNAP holoenzyme structure interacting with the core promoter sites (-35 and -10 positions upstream the TSS) during the RPo. Picture adapted from (Karpen & DeHaseth, 2015) **(C)** Illustration of the elongation complex. The transcription elongation complex protects ~35 bp of the DNA, whereas the transcription bubble comprises 12-14 unpaired nucleotides. This complex is maintained stable due to the formation of the RNA-DNA hybrid and other protein interactions within the RNAP. Picture adapted from (Greive & von Hippel, 2005). The final image was composed using Biorender.

This isomerization pathway from the promoter recognition complex RP_C to 'open complex (RP_O) form' has been studied for several years. Methods including chemical and enzymatic DNA footprinting have been foundational in providing knowledge about the structure of the intermediate complexes leading to RP_O (Gries et al., 2010; Rogozina et al., 2009). Evidence to date suggests that this is found to have at least three intermediate steps (Saecker et al., 2011): DNA loading, DNA unwinding, and assembly of the RNA polymerase clamp (Figure 2A). However, the isomerization intermediates are unstable and short lived (1 ms to 1 s) (Saecker et al., 2011). Therefore, this multi-step process is often represented as one rate-limiting step (Lloyd-Price et al., 2016; DeHaseth et al., 1998).

The last stage of transcription initiation is promoter escape. Until it can fully escape, RNAP needs to percolate 10-15 bp downstream from the promoter site (Goldman et al., 2009; Hsu, 2002). During this stage, the RNAP enters a cycle of abortive initiations where it synthesizes and releases a specific set of abortive short RNAs (2-17 nt in length) (Hsu et al., 1995; Marr & Roberts, 2000). The use of single-molecule-nanomanipulation with ~1 bp resolution and ~1 s temporal resolution suggested that the cycle of abortive initiations occurs through a DNA 'scrunching' mechanism (Revyakin et al., 2006). The scrunching model postulates that the RNAP, while stationary, pulls the downstream DNA into its active site (Kapanidis et al., 2006; Revyakin et al., 2006). Studies using single-molecule-fluorescence-resonance-energy-transfer validated this by showing that the distance between the DNA segments upstream and downstream of the unwound DNA decreases during the abortive initiation cycle (Kapanidis et al., 2006).

After abortive initiation, the RNAP breaks contact with the core promoter elements and enters elongation. The release of the σ factor has been suggested to happen during the promoter escape (Hsu, 2002). However, it may also be retained by RNAP during elongation (Harden et al., 2016).

In **Publication I and II,** we measured with single-molecule resolution the time-intervals between consecutive RNA production events to study the transcription initiation process. Also, in **Publication I, II and III,** we have developed models where the rate limiting steps of RP_C, RP_O, and promoter escape were included. Here, the isomerization states, which separate the RP_C and RP_O, were not included in the models as these are unstable and have negligible lifetimes (1 ms to 1 s) (Saecker et al., 2011).

2.4.2 Transcription Elongation and Termination

Transcription elongation begins when the RNAP clears the core promoter elements and continues with the addition of the nucleotides required to synthesize the RNA transcript.

The transcription elongation complex (TEC) is formed by RNAP, by a DNA subset, and by the growing polynucleotide chain. At each position in the template, the TEC usually occupies ~35 bp of the DNA. The central ones (12-14 bp) are unwound to form the transcription bubble (Figure 2C)(Korzheva et al., 2000; Zaychikov et al., 1995). The TEC slides along the template strand (3' to 5' direction) while nucleotides are added to the 3' end of the growing polynucleotide chain. The 5' end is free to create secondary structures or interact with other components (Greive & von Hippel, 2005).

The movement of TEC occurs discontinuously. As it percolates the DNA, there are several alternative reaction pathways, such as pausing (Greive & von Hippel, 2005), arrest (Fujita et al., 2016; Greive & von Hippel, 2005), editing (Greive & von Hippel, 2005), premature termination (Lewin, 2008), and pyrophosphorolysis (Erie et al., 1993). The probability that the TEC choses any of these pathways depends on sequence-specific interactions between the TEC and the template DNA, the nascent RNA transcript, and other regulatory molecules (Greive & von Hippel, 2005). The duration of pausing and arresting events ranges from less than a second to a few minutes (Herbert et al., 2006, 2010), impacting not only the mean transcription rate, but also the variability in gene expression (Rajala et al., 2010).

Further, pauses and pause escapes can also occur due to collisions between RNAPs (Epshtein & Nudler, 2003). Further, misincorporation can occur at the end of the transcription process (Greive & von Hippel, 2005). Provided no misincorporation or premature termination, elongation continues until reaching a termination site. Termination sites can be divided in two classes: intrinsic and Rho-dependent.

When occurring in intrinsic termination sites, the termination does not require the intervention of other regulatory molecules since the sequence of nucleotides codes for a stem-loop structure, followed by a sequence of nucleotides that forms weak interactions between the RNA and DNA molecules (Martin & Tinoco, 1980; Wilson & von Hippel, 1994). In Rho-dependent transcription termination, the Rho protein binds to the nascent RNA and moves towards the RNAP (in an ATP-

driven process). After reaching the RNAP, it activates its helicase activity to unwind the RNA-DNA within the transcription bubble. *In vitro* experiments suggest that the efficiency of this class of terminators is inversely proportionally to the RNAP elongation rate (Jin et al., 1992).

Both terminator classes include TEC destabilization, leading to the 'bubble' collapse and complex dissociation, releasing the RNA transcript and RNAP (Greive & von Hippel, 2005). Recent single-molecule studies have proposed an expanded bacterial transcription termination mechanism (Harden et al., 2020; Song et al., 2022). In this, after the transcript release at the terminator site, the RNAP stays associated with the DNA and can continue percolating it, eventually reinitiating transcription at another TSS.

Regarding the total RNA production time (i.e., from transcription initiation to termination), in **Publication II**, we report that, on average, $P_{LacO3O1}$ produces one RNA every 1400 s (~23 min). For comparison, in (Taniguchi et al., 2010), the RNA numbers per cell, as measured by FISH for 137 genes, equaled 0.40 (ranging from 0.02 to 3.3 RNAs per cell). Given that the RNA degradation rate is 0.004 s^{-1} (Bernstein et al., 2002) and the RNA dilution rate is 7.7×10^{-5} s^{-1} (cell doubling times are 150 min, as measured in (Taniguchi et al., 2010)) one can estimate the expected average, minimum, and maximum time interval between consecutive RNA production events ($\Delta t_{Average}$, Δt_{Min} and Δt_{Max}, respectively):

$$\Delta t_{Average} = \frac{1}{M_{RNA}} \cdot \frac{1}{(k_{\deg} + k_{dil})} = \frac{1}{0.4} \cdot \frac{1}{0.004 + (7.7\times10^{-5})} = 613 \text{ s} \approx 10 \text{ min} \quad (2.1)$$

$$\Delta t_{Min} = \frac{1}{3.3} \cdot \frac{1}{0.004 + (7.7\times10^{-5})} = 74 \text{ s} \approx 1 \text{ min} \quad (2.2)$$

$$\Delta t_{Max} = \frac{1}{0.02} \cdot \frac{1}{0.004 + (7.7\times10^{-5})} \approx 204 \text{ min} \quad (2.3)$$

The variables k_{deg} and k_{dil} correspond to the RNA degradation rate and to the RNA dilution rate, respectively. Given the results in equations (2.1-2.3), we find the RNA production rate reported in **Publication II** to be in within the expected range.

Finally, in **Publication IV**, we used a stochastic model of transcription with stepwise elongation at the nucleotide level, coupled with the dynamic accumulation of positive supercoils. This model was developed to test if arrests during elongation, caused by supercoiling, disturb the mean RNA production rate.

2.5 Mechanisms of Translation

Translation, similarly, to transcription, can also be divided into three phases: initiation, elongation, and termination. In prokaryotes, translation initiates as soon as the ribosome binding site (RBS) is synthesized (Miller et al., 1970; Yarchuk et al., 1992). It is during translation that ribosomes synthesize proteins. Ribosomes are complex molecules discovered in 1955 by George Palade (Palade, 1955) that are ribonucleoprotein complexes since they are formed by specialized RNA molecules (rRNAs) and proteins. *E. coli* has 70S ribosomes which are formed by two subunits: a large (50S) subunit and a small 30S subunit (Ramakrishnan, 2002). Each subunit has three tRNA binding sites. The tRNA molecules work as a physical link between the RNA and the polypeptide chain and they are also responsible for matching the amino acids with the corresponding codons (sequences of three nucleotides).

For translation initiation to start, the start codon needs to be recognized by a tRNA. In *E. coli*, 83 % of the start codons have the sequence AUG (3542/4284), 14% (612/4284) have the sequence GUG, 3% (103/4284) have the sequence UUG (Blattner et al., 1997), and a few have the sequence AUU (Missiakas et al., 1993; Sacerdot et al., 1982). During translation initiation, the 30S subunit of the ribosome interacts with the consensus sequence of the RBS ("5'-AGGAGG-3'"), also known as Shine-Dalgarno (Saito et al., 2020), forming a 30S-RNA complex.

This complex also binds to the anticodon stem-loops of the tRNA (usually fMet-tRNA if the start codon is AUG) (Ramakrishnan, 2002). Next, with the intervention of initiation factors, the 50S ribosomal subunit is recruited, forming an elongation-competent ribosome (Antoun et al., 2006). During translation elongation, the ribosome moves along the RNA, three nucleotides (codon) at a time. In each codon, a new tRNA binds, and the 50S subunit catalyzes the peptide bond formation between the growing chain of amino acids and the incoming amino acid on the newly bound tRNA (Ramakrishnan, 2002).

Finally, when the ribosome reaches a stop codon, it releases the newly synthesized protein from the ribosome. This process is conducted with the intervention of proteins called 'release factors' (Kisselev & Buckingham, 2000).

Since the rate of translation initiation is influenced by the start codon and by the RBS sequence (Ringquist et al., 1992), to investigate the translation efficiency profiles of different gene cohorts, we compared these sequences with their consensus sequences in **Publication IV**.

2.6 Bacterial Regulation of Gene Expression

The expression of a particular gene in bacterium is subject to strict regulation. In *E. coli,* transcription initiation is when most regulatory mechanisms act (Browning & Busby, 2004, 2016; Chamberlin, 1974; McClure, 1985). The central component of these mechanisms is the RNAP (Browning & Busby, 2004). The regulatory mechanisms can act at the formation of RNAP holoenzyme, promoter recognition by RNAP, or RNAP activity modulation (Browning & Busby, 2016).

Numerous mechanisms have been identified as regulators of gene expression; these include promoter sequence, σ factors, small ligands, TFs, and the bacterial chromosome topology (Browning & Busby, 2004, 2016). The latter two are the focus of this thesis. Our research in **Publication I** focuses on TF regulation. In **Publication II, III, and IV,** we investigated the topology (supercoiling levels) and promoter spatial arrangements of the chromosome of *E. coli* and how they influence gene expression.

2.6.1 Transcription Factors

In *E. coli*, more than 300 genes encode TFs (Pérez-Rueda & Collado-Vides, 2000) that participate in about 4700 interactions with approximately 4600 genes (Santos-Zavaleta et al., 2019).

Depending on the method of regulation, TFs can promote or block transcription initiation (Browning & Busby, 2004). Most TFs bind to the DNA, ensuring that they are promoter specific. TFs differ in the number of genes that they can control. Some TFs can control up to 50 % of all regulated genes (e.g., CRP, FNR, IHF,

ArcA, NarL, and Lrp) while others may only control one specific gene (Martínez-Antonio & Collado-Vides, 2003).

TFs that work as activators improve the binding affinity of RNAP to the DNA. The activation mechanisms by TFs are divided in three classes (Figure 3A) (Browning & Busby, 2004; Lee et al., 2012).

In class I, an activator binds to a specific DNA site located upstream of the -35 promoter region, recruiting RNAP by interacting with its C-terminal domain (α-CTD) (Ebright & Busby, 1995; Jeon et al., 1997). The α-CTD is one of the two domains of the α-subunit of RNAP (the other being an N-terminal domain) (Blattner et al., 1997; Gourse et al., 2000).

In class II, the -35 promoter region overlaps with the activator binding site. Once the activator binds to the DNA, it interacts with one of the domains of the RNAP σ-subunit, recruiting it to the promoter site (Browning & Busby, 2004; Nickels et al., 2002).

In the class III mechanism, the activator does not interact with the RNAP; instead, it binds near the promoter -35 and -10 regions and generates conformational changes that align the -35 and -10 regions. These conformational changes improve the binding affinity of RNAP (Sheridan et al., 2001, 1998).

Contrastingly, a repressor TF reduces the transcription rate at the target promoters. The specific process by which transcription is repressed can differ amongst promoters (Hawley et al., 1985; Schlax et al., 1995). The mechanisms of repression have been classified into three different classes: steric hindrance, looping, and modulation of an activator (Figure 3B) (Browning & Busby, 2004).

Repression by steric hindrance is the simplest. Here, the binding site for the repressor is near the core promoter elements, and when the binding occurs, it prevents recognition of the promoter site by RNAP (Browning & Busby, 2004). In some cases, the repressors can bind to distal operators from the core promoter elements and form a DNA loop that hampers RNAP binding (Browning & Busby, 2016). Finally, for genes that require transcription activators, repression can occur when an activator is modulated (i.e., the repressor binds to the activator molecule and inhibits its function).

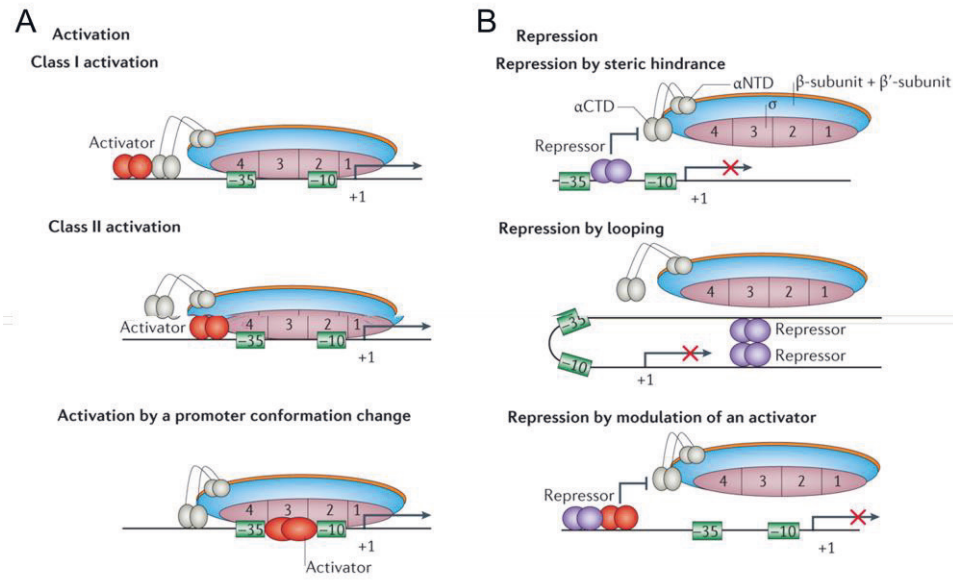

Figure 3. Gene expression regulation by TFs. **(A)** Activation. **(B)** Repression. Picture adapted with permission from (Browning & Busby, 2016).

On the other hand, the activity of TFs is also regulated, usually by controlling their activity or their expression levels. Like all proteins, TFs are also the product of gene expression, and their synthesis is also controlled by mechanisms such as small ligand binding. Small ligand binding is usually responsive to nutrient deficient and stress-induced conditions (e.g., the molecule allolactose reduces the binding affinity of the LacI repressor in the presence of lactose) (Müller-Hill, 1996). Also, σ factors and feedback loops (i.e., when the TF acts as its own repressor) (Pan et al., 2006) can regulate the expression of TFs.

In addition, cells have evolved other mechanisms: Phosphorylation is used to tune TF activity, as it can affect cellular localization, protein stability, and DNA binding affinity (Whitmarsh & Davis, 2000). In addition, the ability of several TFs to bind to the DNA is also dependent on their ability to dimerize. As an example, in the presence of oxygen, the ability of FNR to dimerize decreases, affecting its DNA binding affinity (Moore & Kiley, 2001)

In laboratory conditions, the use of inducer molecules is the most common way of controlling the binding of TFs to DNA (Garcia et al., 2010). In **Publications I** and

II, we use Isopropyl ß-D-1-thiogalactopyranoside (IPTG) and L-arabinose inducers to regulate the activity of the promoters P_{Lac}, $P_{LacO3O1}$, and P_{BAD}. In **Publications III** and **IV**, we studied genome-wide effects of perturbations on TF-gene regulations.

2.6.2 DNA Supercoiling

Given the right-handed double helical structure of DNA, the untwisting of the DNA strands during transcription generates under- and over-winding of the DNA upstream and downstream of the RNAP, respectively (Liu & Wang, 1987). The under-wound DNA is designated as negative supercoiling, while the over-wound DNA is designated as positive supercoiling (Figure 4A).

Global DNA supercoiling arises from the average superhelical density of all topological domains of the DNA, which are usually negatively supercoiled under optimal conditions (Boles et al., 1990; Vinograd et al., 1965). On average, *E. coli* has 400 topological domains with and average size of 10kb (Postow et al., 2004; Stein et al., 2005) per domain. These are dynamic domains that are formed by nucleoid-associated proteins (such as H-NS and Fis) (Wang & Greene, 2011).

In bacterium, such as *E. coli*, these topological domains act as barriers to supercoiling diffusion, leading to supercoiling buildup that, if not resolved, can hinder transcription. *E. coli* has evolved two types of topoisomerases to maintain and regulate the topological state of the DNA (Chen et al., 2013; Vos et al., 2011). Based on their structure and action mechanism, they are either grouped into type I or type II. Type I isomerases make transient single DNA breaks while type II introduce transient double stranded breaks in the DNA (Liu et al., 1980)

Gyrase and Topoisomerase I (TopoI) are the two major topoisomerases that control the global DNA supercoiling level (Chong et al., 2014; Drlica, 1992; Pruss & Drlica, 1986), and have counteracting enzymatic activities. TopoI is a type I topoisomerase and was the first topoisomerase discovered (Wang, 1971). TopoI is found in both prokaryotes and eukaryotes and is responsible for solving negative supercoiled DNA by removing one negative supercoil at a time (Zechiedrich et al., 2000). The removal of each negative supercoil is accomplished by cleaving one strand of the DNA and passing the complementary DNA strand through the open gap of the cleaved strand. After the passage of the complementary strand, the

cleaved strand is resealed. This process is known as the enzyme-bridging model of DNA relaxation by TopoI (Figure 4C) (Lima et al., 1994). Removal of negative supercoiling is important, as its accumulation can lead to the formation of detrimental R loops (Drolet, 2006), when the nascent RNA associates with the template strand, leaving the non-template strand unpaired (Drolet, 2006). These loops have been associated with inhibition of transcription elongation (Hraiky et al., 2000; Huertas et al., 2003). However, within certain boundaries, negative supercoiling facilitates the processes that require DNA unwinding, such as open complex formation (Pruss & Drlica, 1989)

Gyrase, a type II topoisomerase, in the presence of ATP, solves positive supercoils by introducing transient double-stranded breaks and passing a second double-stranded segment trough the gap (Figure 4B) (Gellert et al., 1976; Nöllmann et al., 2007). Besides solving positive supercoils, Gyrase has also been reported to be involved in an ATP-independent mechanism to relax negative supercoils (Gellert et al., 1977; Williams & Maxwell, 1999). However, Gyrase is not efficient in removing positive supercoils, causing positive supercoiling buildup (PSB) to be a common phenomenon downstream of the RNAP (Chong et al., 2014; Guptasarma, 1996). Several single-molecule fluorescence imaging studies in live *E. coli* cells have contributed to the understanding of the chromosomal distribution of Gyrase. On average there are ~12 Gyrase molecules per replisome and ~300 Gyrase molecules dispersed throughout the chromosome (Manley et al., 2008; Planck et al., 2009; Reyes-Lamothe et al., 2010). Additionally, dwell periods for Gyrase range between 2 and 8 seconds, depending on whether the molecules are proximal to the replisome or not. (Stracy et al., 2019).

Studies show that the accumulation of supercoiling can regulate gene expression (Dorman & Dorman, 2016; Drlica, 1992; Drolet, 2006; Travers & Muskhelishvili, 2005). It was reported that a promoter subject to a certain degree of negative supercoiling can increase transcription initiation rates (Burns & Minchin, 1994; Chong et al., 2014). Also, during transcription elongation, under high levels of PSB, an elongating RNAP can slow down, stall, or dissociate from the DNA (Chong et al., 2014; Kim et al., 2019; Ma et al., 2013). Recently, using single-molecule fluorescence *in situ* hybridization (FISH), Chong and colleagues (Chong et al., 2014) confirmed that RNAP has positive supercoiling-sensitive initiation rates and that PSB is one of the mechanisms responsible for transcriptional bursting in *E. coli*. These studies led to other studies to quantify the effect of supercoiling on

transcription (Ancona et al., 2019; Geng et al., 2021; Houdaigui et al., 2019; Sevier & Levine, 2017).

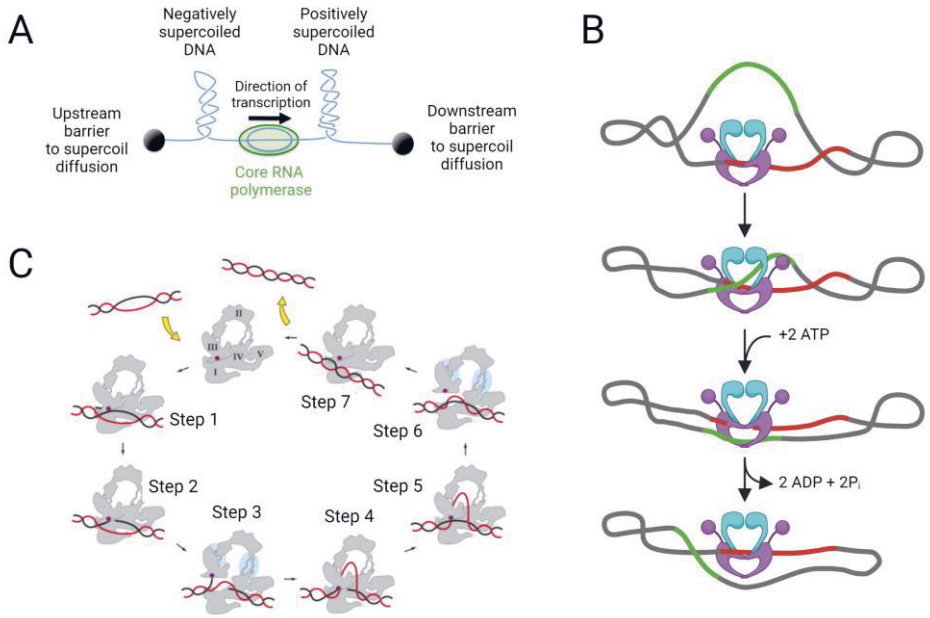

Figure 4. DNA supercoiling accumulation and relaxation by Topo I and Gyrase molecules. **(A)** Twin supercoiling domain model proposed in (Liu & Wang, 1987). As the RNAP percolates the DNA template positive supercoils accumulate downstream the RNAP while negative supercoils accumulate upstream the RNAP. Topological barriers stop supercoiling from diffusing. Nascent RNA and translation machinery are omitted for clarity. Picture adapted with permission from (Dorman, 2019). **(B)** Positive supercoiling relaxation by Gyrase. Gyrase subunits form a heterotetramer. The heterotetramer is formed by two GyrA (purple) and two GyrB (blue) molecules. Relaxation of positive supercoils can be summarized in 4 steps. The first is the binding of the DNA gate segment (red) to the Gyrase active site. Second, a transfer segment (green) makes a positive node over the gate segment. Next, two molecules of ATP bind to each GyrB subunit and trigger a conformational change that breaks the gate segment and allows the passage of the transfer segment through to a bottom chamber. Finally, the hydrolysis of ATP and release of the transfer segment introduce two negative supercoils (Higgins, 2007b). **(C)** Negative supercoiling relaxation by Topo I. Shown are the two DNA strands (red and black). Topo I is represented in grey. In steps 1,2,4,5 and 7 the molecule is in closed complex conformation. In steps 3 and 6 the molecule is in open complex conformation. In step 1, one DNA strand interacts with the topoisomerase domain I and IV . In step 2, cleavage of one DNA strand occurs. Next, the superhelical tension drives the other DNA strand (auto-transported strand) into the cavity (step 3) followed by the return of domain III to its original configuration (step 4). In step 5, the cleaved DNA strand is religated completing the cycle of removal of a negative supercoil (step 6). After step 6, Topo I can either initiate a new relaxation cycle or dissociate from the DNA. Picture adapted with permission from (Viard & de la Tour, 2007). Generated using Biorender.

Also recently, a study on supercoiled DNA minicircles using electron cryotomography (Fogg et al., 2021) suggests that supercoiling may be used by cells for adopting a multiplicity of genome conformations (including non-energetically favorable ones) that favor the binding of other regulatory molecules. The fact that the topological state of DNA affects the affinity of DNA binding proteins may be a mechanism for adaptation to extracellular stresses. In fact, previous studies have shown that DNA supercoiling is sensitive to extracellular stresses. For example, oxidative stress decreases DNA supercoiling levels (Weinstein-Fischer et al., 2000) while anaerobiosis increases it (Hsieh et al., 1991).

Nevertheless, it is important to highlight that other factors also affect genome conformation (Dorman et al., 2013), these include: DNA binding proteins (HU, H-NS etc.)(Dillon et al., 2010; Malik et al., 1996), the DNA (bio)polymeric nature (Pelletier et al, 2012), and the indirect effects caused by high concentration of molecules (also referred to as 'macromolecular crowding') (De Vries, 2010).

Finally, it is also worth noting that besides transcription, other processes such as DNA replication, recombination, and segregation also change DNA supercoiling (Alberts, 2003; Higgins, 2007a). For example, as DNA replication proceeds, if topoisomerases do not intervene, positive supercoils can accumulate ahead of the replication fork leading to the entanglement of daughter chromosomes (Postow et al., 2001). However, this thesis focuses on the effects of PSB during transcription. In detail, in **Publications II and IV**, we study the effects of this phenomenon on the kinetics of transcription and show that the fraction of time in locked states due to PSB differs with a gene's basal transcription rate. Finally, we also provide evidence that changes in DNA supercoiling act as a regulator of genes' responsiveness during cold shock.

2.6.3 Closely Spaced Promoters

Several pairs of genes in *E. coli* are co-regulated by bidirectional promoters (Santos-Zavaleta et al., 2019). These promoter pairs are characterized for having closely spaced transcription start sites (TSSs), specifically less than 1 kb apart (Trinklein et al., 2004). The geometries of closely spaced promoters can be tandem, divergent or convergent (Beck & Warren, 1988; Korbel et al., 2004)(Figure 5). In the tandem configuration, there is one promoter upstream of another, both transcribing in the same direction. In the divergent configuration, transcription happens in opposite

directions and the two RNAP binding sites may overlap (Shearwin et al., 2005). In the convergent configuration, the sequence coding for the RNA transcripts can overlap for a certain number of nucleotides (Shearwin et al., 2005). Apart from their structure, closely spaced promoters might differ in terms of the distance between their two TSSs and the position of their transcription factor binding sites. (Gama-Castro et al., 2011).

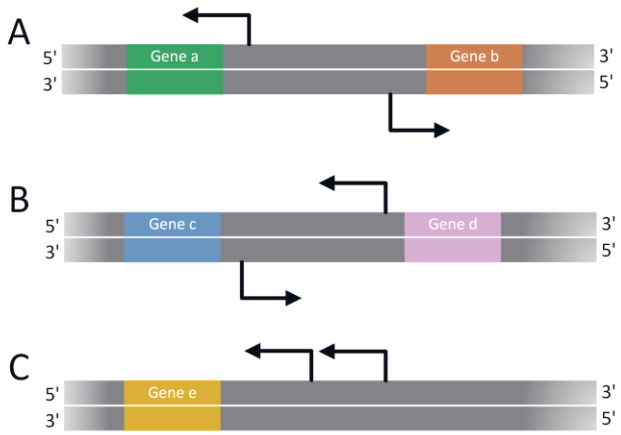

Figure 5. Promoter configurations that can lead to transcriptional interference. **(A)** Divergent, **(B)** Convergent, and **(C)** Tandem promoters. Black arrows indicate the transcription start sites. In the divergent and convergent configurations, the two genes are transcribed from opposite strands of the DNA. Interestingly, in (B) the genes may be located in between the promoters. If that is the case, they are considered converging transcripts if there are no terminators in between the genes.

They are also a common organizational motif of the human genome (Adachi & Lieber, 2002). Further, in bacteria, closely spaced promoters may arise from the fact that the probability of genes being co-transferred increases as the distance between them decreases (Lawrence, 2003; Lawrence & Roth, 1996).

In closely spaced promoters, TFs are not the only means of transcription regulation since interactions between the RNAPs may occur, causing transcription interference (TI). The term transcription interference refers to a direct suppressive perturbation of a transcriptional process on another transcriptional process (Shearwin et al., 2005). Transcriptional interference may be one of the oldest means by which closely spaced genes interact. However, since their expression kinetics remains largely unexplored in prokaryotes, their exact mechanisms of regulation are also not well catalogued.

Callen *et al.* (Callen et al., 2004) studied one set of phage, converging promoters (lysis-lysogeny switch from coliphage 186) using *in vivo* methods. This promoters' set is characterized by the 186 lytic promoter being stronger than the lysogenic promoter (*pL).* The study made different promoter arrangements to search for possible mechanisms of interference. Results showed that most of the interference is due to the overlapping of the open complex elongation over the *pL* promoter. In the same year, a repression mechanism acting by transcription interference was identified for the first time (Martens et al., 2004). In detail, the study focused on the *S. cerevisiae SER3* gene, which is highly repressed during growth phase. However, there is a high level of active transcription upstream *SER3*, in the same strand. This highly transcribed region was shown to code for a non-protein-coding RNA and was designated as *SRG1*. Experiments using derivatives of *SRG1,* with and without a transcription-termination region, revealed that when the terminator is present, there were no signs of repression of the *SER3* gene, supporting the existence of repression by transcription interference.

Several transcription interference mechanisms have been identified: Occlusion, where the access to the second promoter is blocked due to the elongation complex (Adhya & Gottesman, 1982), collisions between elongating RNAPs from each of the promoters (Prescott & Proudfoot, 2002; Ward & Murray; 1979), and occupancy of the downstream promoter, where an RNAP on the downstream promoter becomes unable to form a closed complex due to the RNAPs elongating from the upstream promoter passing through (Sneppen et al., 2005).

More recently, (Yeung et al., 2017) reported the expression of induced synthetic promoters arranged in all possible orientations, and the results agreeed with past studies (Brophy & Voigt, 2016), in that the configuration affects the overall kinetics.

However, knowledge on natural interference factors for each arrangement (divergent, convergent, and tandem) remain speculative. In **Publication III,** we contribute to this by studying the dynamics of many natural genes controlled by tandem promoters. We confirmed past hypotheses that transcriptional interference in tandem promoters is particularly strong when there is occlusion and differs with the occupancy times of the two promoters.

2.7 Effects of Temperature

The fitness of bacterial cells depends on their adaptability to environmental changes. This adaptation has two phases. First, the cell triggers a rapid response that initiates the adaptation. Next, the cell generates a more extended response to sustain survival under the stress (Arsène et al., 2000; Gunasekera et al., 2008; Phadtare & Inouye, 2004).

One of the most typical stresses that an organism faces in nature is temperature change. Bacteria have a cold-shock response when facing temperature downshifts (Phadtare & Inouye, 2004). Contrastingly, a temperature upshift triggers a heat-shock response (Craig & Schlesinger, 1985; Neidhardt et al., 1983).

The transcriptional response program of *E. coli* to heat shocks is controlled by σ^{32} (Grossman et al., 1987), which is one of the seven σ-subunits of RNAP (Maeda et al., 2000). The concentration of σ^{32}, under optimal growth conditions, is very low (Craig & Gross, 1991). However, at elevated temperatures, its expression increases by 8-fold (Craig & Gross, 1991), initiating a heat shock-response transcriptional program that synthesizes several proteins. These include molecular chaperones and proteases that counteract protein denaturation-related changes (Georgopoulos, 2006; Guisbert et al., 2004, 2008). In addition, the free chaperones can also bind to and inactivate σ^{32}, causing negative feedback regulation to active σ^{32} levels. The complete regulatory network controlling the synthesis, stability, and activity of σ^{32} is still largely unknown. Recent studies suggest that σ^{32} is also associated with the cell membrane, allowing it to be responsive to the folding status of the proteins within the inner cell membrane (Lim et al., 2013; Miyazaki et al., 2016).

Bacteria have also evolved complex 'cold-shock' responses to downshifts in temperature. The cold shock adaptation process includes cell growth stoppage for 3 to 6 hours (called acclimation phase). During that time, the synthesis of certain proteins is inhibited, while the synthesis of others is induced. The proteins whose synthesis increased counteract the effects of cold temperatures. After this acclimation phase, the expression of cold- inducible proteins declines and bulk protein production and growth are resumed at slower rates than in optimal conditions (Phadtare & Inouye, 2004)

Temperature downshifts have wide effects on cells, such as lowering membrane permeability (Cao-Hoang et al., 2010; Goncalves et al., 2018), stabilizing nucleic acid secondary structures, hampering protein folding, reducing translational rates, etc. (Phadtare & Inouye, 2008). To restore membrane flexibility, cells induce the synthesis of unsaturated fatty acids with high flexibility (Sinensky, 1974). Also, most of the cold-inducible proteins are nucleic acid chaperones from the Csps family, involved in RNA metabolism (Gualerzi et al., 2003). The main function of these proteins is to prevent the formation of RNA secondary structures or facilitate RNA degradation. In addition, the translation initiation factor IF2 and the protein Hsc66 were suggested to participate in correcting protein folding during cold shocks (Caldas et al., 2000; Lelivelt & Kawula, 1995).

Regarding the translational block, the protein PY has been identified as being involved in it, for certain RNAs. PY binds to the 30S subunit of ribosomes, preventing ribosomes from forming 70S initiation complexes. This protein was detected one hour after the cold shock (Di Pietro et al., 2013). Another possible cause for translational block is that RNA structures can hide the Shine-Dalgarno sequences, not allowing translation initiation. However, this is expected to be countered by the Csps protein family, as they prevent the synthesis of RNA secondary structures.

The regulatory network that triggers the cold shock response in *E. coli* is not completely understood. Most studies, described above, suggest that regulation relies on changes in RNA stability and in the translation machinery. Nevertheless, a genome-wide study has suggested that supercoiling may play a role in gene expression regulation. This is done by functioning as an intermediary messenger, transmitting environmental signals to the regulatory networks (Drolet, 2006; Peter et al., 2004). More recently, it has also been hypothesized that high sensitivity to supercoiling is a feature shared by many genes capable of short-term cold shock responses (Oliveira et al., 2019). In **Publication IV**, we investigated this hypothesis by performing genome-wide measurements under cold-shock and under Gyrase inhibition. In detail, we collected samples and analyzed the transcription profile by RNA-seq. Next, we compared the sets of short-term responsive genes to each of the perturbations. We found that their behavior is correlated. As such, supercoiling not only reflects environmental changes, but also acts as a mechanism of gene expression regulation under low temperatures.

2.8 Single-Cell Variability in Gene Expression

Isogenic cells, growing in the same environment, exhibit variability in both RNAs and protein levels (Elowitz et al., 2002; Kærn et al., 2005; McAdams & Arkin, 1997). Given its implications for cellular control and non-genetic uniqueness, this phenomenon has been a topic of study for years (Novick & Weiner, 1957; Berg, 1978; Kepler & Elston, 2001; Rigney & Schieve, 1977).

The variability in RNA and protein concentrations between genetically identical cells (defined as the standard deviation divided by the mean) mostly arises from stochasticity (i.e., noise) in gene expression. For example, transcription is intrinsically stochastic, since the binding of regulatory molecules to the promoter region is the result of random encounters between them. Since they can exist in relatively low numbers, the time they take to bind differs from one event to the next (Elowitz et al., 2002; Kærn et al., 2005; Paulsson, 2005). These variabilities result in temporal fluctuations in RNA and protein numbers in any given cell, and then variability also in number between different cells.

Besides this 'intrinsic' noise, additional variability emerges from an extrinsic component, such as potential differences between cells in the numbers of regulatory molecules and polymerases (Elowitz et al., 2002; Swain et al., 2002). To observe this, in 2002, Elowitz and colleagues constructed strains with a dual reporter system where two identical promoters were tagged with a different fluorescent protein (CFP and YFP, respectively) to measure the extrinsic and intrinsic noise at the single gene level. The intrinsic noise is given by the relative difference in fluorescence intensity of the two reporters, whereas the extrinsic noise is the correlated component between the two reporters (Elowitz et al., 2002).

Over the years, new methodologies appeared for measuring RNAs (Golding et al., 2005; Jones et al., 2014; Lenstra et al, 2016; So et al., 2011), proteins (Taniguchi et al., 2010; Stracy et al., 2015; Yu et al., 2006), and plasmids (Reyes-Lamothe et al., 2014) with single molecule sensitivity. In (Golding et al., 2005), using *in vivo* tagging of RNAs that allowed quantification of its numbers with single-molecule resolution, they showed evidence that transcription occurs in bursts (even when cells are fully induced). Transcription bursts are periods of high expression intensity followed by periods of low intensity. The frequency and size of these

bursts have an impact on the RNA and proteins, contributing to noise. Later in (Chong et al, 2014), it was suggested that these bursts are due to PSB.

Also, transcription and translation are multi-step processes whose kinetics are sequence-dependent. (Jones et al., 2014; Lutz et al., 2001; Saecker et al., 2011) and the duration of these steps can vary considerably, (Herbert et al., 2006; Tuller et al., 2010) independently of TF occupancy (Gama-Castro et al., 2011). In addition to gene expression stochasticity, asymmetries in the partitioning of RNA and proteins during cell division further enhance cell-to-cell variability in RNA and protein numbers (Baptista & Ribeiro, 2020).

In **Publication II,** we investigated how the cell-to-cell variability in RNA numbers differs with supercoiling levels**.** In **Publication III,** we studied how the cell-to-cell variability in protein numbers differs with promoter arrangement**.** Finally, in **Publication IV,** we investigated how temperature downshifts affect the cell-to-cell variability in protein numbers, as measured by the CV^2.

3 MODELS AND SIMULATIONS

This chapter provides an outline of the theoretical foundations of model design and simulation used in this thesis. First, it briefly reviews the foundations of stochastic models with chemical kinetics. Following that, it covers the fundamentals of gene networks' modeling. Finally, it provides background knowledge about the incorporation of complex biological systems and processes into models of gene expression.

3.1 Chemical Master Equation

> *"If large numbers of identical events occurred in the same cell, and they were statistically independent, relative fluctuations could be ignored and deterministic rate equations would suffice to describe dynamics. But the numbers are not large and the events are not independent."*
>
> (Paulsson, 2005)

This quotation, taken from the review by Johan Paulsson (Paulsson, 2005) of models of stochastic gene expression, conveys that cellular systems cannot be analyzed using traditional methods in which the variables change deterministically.

According to traditional chemical kinetics models, the evolution in time of a well-stirred and thermally equilibrated system can be given by the following set of ordinary differential equations (ODEs):

$$\frac{dX_i}{dt} = f_i(X_1,...X_N) \tag{3.1}$$

where X_i is, in a system with N species, the number of molecules of the chemical species S_i $(i=1,...,N)$, while the functions f_i control how X_i changes over time, depending on which reactions occur in the interval of time dt. The true behavior of a cellular system cannot be well described by Equation 3.1 when the number of molecules of the reactant species are small.

In a bacterium, a gene is frequently found in a single copy, its RNAs are also present in low amounts (usually less than 10 per gene (Taniguchi et al., 2010)), and its proteins are found in less than 100 molecules (Arkin et al., 1998; Blake et al., 2003; Elowitz et al., 2002; McAdams & Arkin, 1997; Paulsson, 2005). Therefore, to track the time evolution of these molecules with more accuracy, one should perform simulations where the population of each species changes with the occurrence of a reaction.

When modelling gene networks, assuming a well stirred system restricted to a fixed volume in thermal equilibrium, one can consider only the molecular populations of interest and model only the reactions that change their populations (Gillespie, 2007). Assuming we have a system of molecules of N chemical species interacting trough M chemical reactions R_j $(j = 1,...,M)$, where the number of molecules of each species (S_i), at a given moment, is $X_i(t) = x_i$. Then, the population size of each species is saved in a current state vector $x = (x_1,...,x_N)$ (Gillespie, 1977, 2007).

Each reaction R_j is characterized by the state-change vector v_j (which causes a change in the state vector x) and by its propensity, which is defined as (Gillespie, 2007):

$$a_j(x)dt \tag{3.2}$$

Expression 3.2 is the probability that, given the state vector x, the reaction R_j will occur in the infinitesimal time interval $[t, t+dt)$.

For unimolecular reactions (i.e., occurring as a result of processes internal to a single molecule) we consider the constant c_j, such that $c_j dt$ is the probability that a molecule of species S_j will undergo reaction R_j in the next dt. As such, one can write the reaction propensity as (Gillespie, 2007):

$$a_j(x) = c_j x_1 \tag{3.3}$$

For biomolecular reactions, provided that the system is well-stirred, the theory predicts that there is a constant c_j such that $c_j dt$ is the probability that a random pair of molecules (one S_1 and the other from S_2) react during the next dt. Given this, the propensity of the reaction can be written as (Gillespie, 2007):

$$a_j(x) = c_j x_1 x_2 \tag{3.4}$$

In the case of a bimolecular reaction where two molecules of the same species react, the number of unique molecule pairs able of reacting is $\frac{x_1(x_1-1)}{2}$, and the propensity function is thus:

$$a_j(x) = \frac{c_j x_1 (x_1 - 1)}{2} \tag{3.5}$$

Higher order reactions do not occur as "elementary events" and should instead be partitioned into simpler unimolecular or bimolecular reactions (Gillespie, 1992). Also, one should know that the constant c_j of uni- and bimolecular reactions are distinct. For unimolecular reactions c_j is numerically equal to the reaction rate constant while, for bimolecular reactions, c_j is a scaling factor dependent on the system volume (Gillespie, 1977, 1976, 1992).

From equation 3.2 one can derive the probability of the system existing in a particular state x at any point in time t, given the initial conditions $x = x_0$ and $t = t_0$ (Gillespie, 2007):

$$\frac{\partial P(x,t \mid x_0,t_0)}{\partial t} = \sum_{j=1}^{M} [a_j(x - v_j) P(x - v_j, t \mid x_0, t_0) - a_j(x) P(x,t \mid x_0,t_0)] \tag{3.6}$$

Equation 3.6 is known as the chemical master equation (CME). It describes the time evolution of the probabilities of changes in molecular abundances. The CME is a collection of coupled ODEs, with one ODE for each set of identical reactant molecules. In most cases, the CME cannot be solved explicitly, particularly for highly dimensional systems (i.e., systems in which there are several possible combinations of the reactant species' population sizes).

Several methodologies have been developed to obtain approximated solutions of the CME. These methods can be divided into three major classes: (i) methods that compute approximations of the CME solution by solving a truncated version of the original Markov process such as the finite state projection method (Munsky & Khammash, 2006) and the window abstraction method (Henzinger et al., 2009); (ii) methods that use asymptotic simplifications such as the Moment Closure method (Gómez-Uribe & Verghese, 2007; Hespanha & Singh, 2005); (iii) and a class of

methods that uses kinetic Monte Carlo approaches. These work by generating multiple realizations of $X(t)$ to obtain the statistics of these events (Gibson & Bruck, 2000; Gillespie, 1976, 2001). In **Publications II, III and IV** we used the later approach.

3.2 Stochastic Simulation Algorithm

The stochastic formulation of chemical kinetics considers a system's discreteness and stochasticity, which are key factors to consider while predicting the exact molecular population at a future time. The stochastic simulation of the CME is a popular method that requires only a chemical reactions model and a stochastic simulation algorithm (SSA)(Gillespie, 1977, 1976, 1992). The stochastic simulation algorithm (SSA) (Gillespie, 2007) is a computational method that uses a Monte Carlo approach to quantitatively model the temporal development of a chemical system, while accounting for the intrinsic fluctuations that deterministic models overlook (Gillespie, 1977).

The SSA works by answering two questions, iteratively: which reaction (j) will happen next, and in how much time (τ). This is based on the probability function $p(\tau, j \mid x,t)$ which is defined as the probability, given $X(t)=x$, that the next R_j reaction will happen in the infinitesimal time interval $[t+\tau, t+\tau+d\tau)$. The exact formula of $p(\tau, j \mid x,t)$ can be written as (Gillespie, 1977):

$$p(\tau, j \mid x,t) = a_j(x) e^{-a_0(x)\tau} \tag{3.7}$$

where,

$$a_0(x) = \sum_{j=1}^{M} a_j(x) \tag{3.8}$$

According to equation 3.7, the time τ until the next reaction occurs is an exponential random variable with mean $1/a_0(x)$ while the identification of which reaction it is, j, is an integer random variable with the probability $a_j(x)/a_0(x)$ (Gillespie, 1977).

Several Monte Carlo procedures can be implemented to generate the samples of τ and j, with the simplest being the direct method (DM). According to the DM, we generate two random numbers r_1 and r_2 from the uniform distribution in the interval [0, 1]. Next, r_1 and r_2 are used to generate τ and j as follows:

$$\tau = \frac{-\ln(r_1)}{a_0(x)} \tag{3.9}$$

$$j = \text{the smallest integer satisfying } \sum_{j'=1}^{j} a_{j'}(x) > r_2 a_0(x) \tag{3.10}$$

Using this or other methods, such as the next reaction method (Gibson & Bruck, 2000) or the logarithmic direct method (Li & Petzold, 2006), along with the start and stop simulation times and the initial population size vector of each species (x_0) one can implement the SSA and simulate trajectories of $X(t)$ (Gillespie, 2007).

In **Publications II, III and IV** we performed stochastic simulations of the models using the simulator SGNS2 (Ribeiro & Lloyd-Price, 2007), whose dynamics is driven by the SSA.

3.3 Models of Transcription in *Escherichia coli*

Advances in fluorescent live cell imaging, single-cell fluorescence microscopy, genetic engineering, and computational image processing have given new understanding of the transcription process *in vivo*, contributing to the development of stochastic models of gene expression, particularly for the model organism *E. coli*.

Most stochastic models of gene expression focus on transcription initiation (Baptista et al., 2022; De Jong, 2002; Gibson & Bruck, 2000; Ribeiro, 2010) since most regulatory mechanisms occur at this step (Browning & Busby, 2004; Djordjevic & Bundschuh, 2008). However, transcription can also be regulated during elongation (e.g., the tryptophan attenuation mechanism (Simão et al., 2005) and the accumulation of supercoiling (Rovinskiy et al., 2012)). Also, the existence of regulation of RNA levels as a function of cell growth rates has been reported (Esquerré et al., 2014). Nevertheless, unlike transcription initiation, there is no evidence for sequence dependent RNA regulation.

The simplest stochastic model of gene expression assumes a constitutive gene (i.e., always in active state, reaction 3.11) and models the production of RNAs and proteins as one-step processes (reactions 3.11 and 3.12, respectively). Usually, RNA and protein degradation are also modelled as one-step events (exponential decays) (reactions 3.13 and 3.14, respectively) (Munsky & Khammash, 2006). This model is shown in Table 1.

Table 1. One-step gene expression model. The rate constants k_1 and k_2 are the RNA and protein (*P*) production rates, respectively. The rate constants k_3 and k_4 are the RNA and protein degradation rates, respectively. The variable *Pro* stands for active promoter.

Event	Reaction	
Transcription	$Pro \xrightarrow{k_1} RNA$	(3.11)
Translation	$RNA \xrightarrow{k_2} P$	(3.12)
RNA degradation	$RNA \xrightarrow{k_3} \varnothing$	(3.13)
Protein degradation	$P \xrightarrow{k_4} \varnothing$	(3.14)

Next, one can introduce regulatory processes to this model. As described in section 2.6.1 one common mechanism of gene expression regulation is the binding of a TF to the promoter region (Jacob & Monod, 1961). For example, promoter activation may require the binding of an activator (reaction 3.15). Contrarily, repression may occur upon the binding of a repressor molecule to the promoter region (reaction 3.16).

Table 2. ON-OFF promoter transitioning model. A promoter can be either in an active or repressed state ('ON' and 'OFF', respectively). The binding of an activator molecule (*Act*) or the unbinding of a repressor (*Rep*) molecule is required for promoter activation. The rate constant k_{OFF}^{P} is the rate at which an active promoter is locked. The rate constant k_{ON}^{P} is the rate at which a locked promoter becomes active.

Event	Reaction	
Promoter activation	$Pro_{OFF} + Act \underset{k_{OFF}^{P}}{\overset{k_{ON}^{P}}{\rightleftharpoons}} Pro_{ON}.Act$	(3.15)
Promoter repression	$Pro_{OFF}.Rep \underset{k_{OFF}^{P}}{\overset{k_{ON}^{P}}{\rightleftharpoons}} Pro_{ON} + Rep$	(3.16)

The binding and unbinding events will make the promoter transitioning between 'ON' and 'OFF' states (i.e., active, and inactive, respectively) at exponentially distributed intervals (Gardiner, 2004). Example model reactions of these events are shown in Table 2.

In the model above, transcription is depicted as a one-step process, but transcription initiation is a multi-step process. In fact, the first *in vitro* studies focused on the regulation of the transcription initiation steps (binding, isomerization, and promoter clearance), in order to establish the first detailed models of transcription (McClure, 1985; Saecker et al., 2011).

The steps of transcription initiation were first identified by two methods: the abortive initiation assay and the *in vitro* transcription assay (McClure, 1980, 1985; McClure et al., 1978). These studied showed that the lag times to reach steady state were found to be dependent on the RNAP concentration. Also, they showed that transcription initiation can be well described as a two-step process (McClure, 1985). These steps are modelled in reactions 3.17 and 3.18 of Table 3.

The model in Table 3 involves the binding of the RNAP to an active promoter with a rate binding constant, k_{CC}, to form a closed complex (reaction 3.17). In detail, the rate constant k_{CC} involves the finding of the promoter by RNAP followed by DNA percolation until binding to the TSS (Bai et al., 2006; Wang & Greene, 2011). Next, the closed complex isomerizes (reaction 3.18) with a rate constant, k_{OC}, and forms the open complex (RP_O). Given that, the closed complex formation is reversible, multiple closed complexes are produced until one of them successfully unwinds the DNA and establishes a transcription bubble, forming a stable open complex (Bai et al., 2006; Baptista & Ribeiro, 2020). Finally, equation 3.19 models the promoter escape, which is the last stage of transcription initiation.

Table 3. Multi-step transcription model. In the reactions, *Pro* stands for promoter, RP_C stands for closed complex and RP_O for open complex.

Event	Reaction
Closed complex formation	$Pro_{ON} + RNAP \xleftrightarrow{k_{CC}} RP_C$ (3.17)
Open complex formation	$RP_C \xleftrightarrow{k_{OC}} RP_O$ (3.18)
Promoter escape and transcription elongation	$RP_O \xrightarrow{k_{escape}} Pro_{ON} + RNAP + RNA$ (3.19)

The escape process is marked by the generation of a series of abortive transcripts which reflect the instability of the transcribing complexes during the escaping process. As such, promoter escape is considered a rate limiting step during initiation, here represented by the rate constant k_{escape} (Duchi et al., 2016; Hsu et al., 1995; Liang et al., 1999). After the RNAP is released from the promoter, the elongation phase begins. Modelling transcription as a multi-step process has been shown to be relevant. For example, in (Startceva et al., 2019) it was shown that the rate constants of k_{CC} and k_{OC} affect the shape of the distribution of time intervals between consecutive RNA production events. Furthermore, because both phases are lengthy, regulatory mechanisms may act just in one or both (Mäkelä et al., 2017). As a result, an accurate model must account for both steps. Using the measured time intervals between single-molecule RNA productions at various RNAP concentrations, a recent study (Lloyd-Price et al., 2016), estimated the *in vivo* durations of the open and closed complex for the $P_{Lac/ara-1}$ promoter.

A different modelling approach uses delayed stochastic models where a delay is introduced to account for the time-length of transcription initiation and/or protein folding and activation (Ribeiro et al., 2006). This allowed for an easy investigation of the system dynamics through the examination of delay effects. This simple delayed stochastic model, modelling transcription and translation as delayed reactions, accurately reproduce the stochastic kinetics of protein production (Zhu et al., 2007) except when two-body effects exist (e.g., collisions between RNAPs) (Ribeiro et al., 2009; Roussel & Zhu, 2006).

Next, stochastic models at the single nucleotide level, at both the initiation and elongation stage, were developed (Mäkelä et al., 2011; Rajala et al., 2010; Ribeiro et al., 2009; Ribeiro et al., 2010). These models explicitly incorporated events such as pausing, arrest, misincorporation and editing, pyrophosphorolysis, premature termination, and account for the region occupied by an RNAp when on the DNA template. In Table 4 example model reactions of these events are shown.

To simulate elongation at the single nucleotide level, one needs to first model transcription initiation, where a promoter is found by an RNAP (reaction 3.20), followed by promoter escape (reaction 3.21), which initiates stepwise elongation (reaction 3.22). A new transcription initiation event can occur as soon as the promoter becomes vacant.

Table 4. Single-nucleotide transcription model. Shown are the chemical reactions representing the various processes to model transcription initiation, elongation, parallel and competing events at the nucleotide (*n*) level, termination, and RNA production. Pro stands for the promoter region, *RNAP* for the RNA polymerase, and *RNAP.Pro* for the promoter region when occupied by an RNAP. The variables A_n, O_n and U_n stand for the nth nucleotide when active, occupied, and unoccupied, respectively. Ranges of nucleotides are denoted as in $U_{[start,end]}$, which are a particular set of consecutive, unoccupied nucleotides from indexes start to end. O_{n_p}, $O_{n_{ar}}$ and $O_{n_{correcting}}$ represent a paused, arrested, or error correcting RNAP at position *n*, respectively. On the template, each RNAP occupies (2Δ+1) nucleotides, where Δ = 12 (Greive & von Hippel, 2005).

Event	Reaction	
Initiation and promoter complex formation	$Pro + RNAP \xrightarrow{k_1} RNAP.Pro$	(3.20)
Promoter escape	$RNAP.Pro + U_{[1,(\Delta+1)]} \xrightarrow{k_m} O_1 + Pro$	(3.21)
Elongation	$A_n + A_{n+\Delta+1} \xrightarrow{k_m} O_{n+1} + U_{n-\Delta}$	(3.22)
Activation	$O_{n+1} \xrightarrow{k_{act}} A_{n+1}$	(3.23)
Pausing	$O_n \underset{1/t_p}{\overset{k_{pause}}{\rightleftarrows}} O_{n_p}$	(3.24)
Pause release due to collision	$O_{n_p} + A_{n-2\Delta-1} \xrightarrow{k_{m_1}} O_n + A_{n-2\Delta-1}$	(3.25)
Pause induced by collision	$O_{n_p} + A_{n-2\Delta-1} \xrightarrow{k_{m_2}} O_{n_p} + O_{n-2\Delta-1_p}$	(3.26)
Arrests	$O_n \underset{1/d_{ar}}{\overset{k_{ar}}{\rightleftarrows}} O_{n_{ar}}$	(3.27)
Editing	$O_n \underset{1/d_{ed}}{\overset{k_{ed}}{\rightleftarrows}} O_{n_{correcting}}$	(3.28)
Premature termination	$O_n \xrightarrow{k_{pre}} RNAP + U_{[(n-\Delta),(n+\Delta)]}$	(3.29)
Pyrophosphorolysis	$O_n + U_{n-\Delta-1} \xrightarrow{k_{pyr}} O_{n-1} + U_{n+\Delta}$	(3.30)
Completion	$A_{n_{last}} \xrightarrow{k_f} RNA + RNAP + U_{n_{[last,last-\Delta]}}$	(3.31)
RNA degradation	$RNA \xrightarrow{k_3} \varnothing$	(3.32)

As the RNAP percolates the DNA from one nucleotide to the next (reaction 3.22), an intermediate activation step (reaction 3.23) is required. However, several events compete with the activation step, such as pausing (reaction 3.24), arrest (reaction

3.27), editing (reaction 3.28), premature termination (reaction 3.29) and pyrophosphorolysis (reaction 3.30). All these events, with the exception of premature termination, are modeled as reversible due to the transcription machinery's capacity to resolve them.

The model also accounts for pauses and pause escapes caused by RNAP collisions. (reactions 3.26 and 3.25, respectively). If there is no premature termination, elongation is completed, an RNA is generated, and the RNAP is released (reaction 3.31).

Both single-nucleotide level models and single-step multi-delayed stochastic models were shown to be accurate for low expression rates (Mäkelä et al., 2011; Rajala et al., 2010; Ribeiro et al., 2009; Roussel & Zhu, 2006), however for higher rates the two models differ. As such, when modelling highly transcribed genes both transcription and translation elongation should be modelled explicitly and coupled to mimic the mean and noise in RNAs and proteins (Mäkelä et al., 2011). Models at the single-nucleotide level, although being complex, allow the study of mechanisms at the sequence level. For example, in (Potapov et al., 2012), making use of a stochastic model of transcription and translation at the nucleotide level, the authors suggested that mean codon translation efficiencies near the RBS region are determinant for the dynamics of protein numbers.

In **Publication I**, we proposed a five-step model of the dynamics of transcription repression. The model includes the reversibility between ON and OFF promoter states (due to TFs), the closed complex formation, an initial stage of open complex formation, a fully formed open complex and promoter escape. In **Publication II** we used a stochastic model of transcription with stepwise elongation, at the single nucleotide level. In **Publication III** we started by modelling the dynamics of gene expression under the control of promoters in tandem formation assuming a four-step model of transcription. Finally, in **Publication IV** we assumed three distinct transcription initiation models: a one-step model, a two-step model, and an ON-OFF model, to compare their protein noise for the same mean protein expression level.

3.3.1 Modelling Supercoiling

Several gene expression models have been developed (McAdams & Arkin, 1997; Paulsson, 2005; Rajala et al., 2010; Ribeiro, 2010; Roberts et al., 2011), in which the explicit modelling of supercoiling was shown not to be determinant to mimic the systems' kinetics. However, for highly active operons the production of supercoiling causes transcriptional bursting (Chong et al., 2014; El Hanafi & Bossi, 2000). As such, incorporating this mechanism when modelling such systems becomes critical if one wants to emulate its noise and mean kinetics.

Modelling PSB at the transcription level can be done at different levels of detail, which will change the level of complexity required. A simple way of modeling the effects of PSB would be using the same approach as reaction 3.15. In this case, as positive supercoiling accumulates, the propensity of the promoter entering an OFF-state increases (Chong et al., 2014; Ma et al., 2013). This event can be modeled by reaction 3.33 (Baptista & Ribeiro, 2020). The escape from an OFF-state state can be modeled as in reaction 3.34. The propensities of these events are dynamic, in that they differ with the global level of PSB in the region of the DNA where the gene of interest is located (Chong et al., 2014; Ma & Wang, 2014).

Table 5. ON-OFF promoter transitioning model due to PSB. The model includes the promoter when active (Pro_{ON}) and when locked due to PSB (Pro_{OFF}). The rate constant k_{PSB} is the rate at which an active promoter is locked, given the presence of PSB. The rate constant k_G is the rate at which an inactive promoter becomes active, depending on the number of Gyrases (GYR) in the system.

Event	Reaction
Promoter locking due to PSB	$Pro_{ON} \xrightarrow{k_{PSB}} Pro_{OFF}$ (3.33)
Promoter unlocking	$Pro_{OFF} + GYR \xrightarrow{k_{GYR}} Pro_{ON}$ (3.34)

Since Gyrases relax positive supercoils (Gellert et al., 1976), the kinetics of reaction 3.34 is determined by their number, their rate of binding/unbinding to the DNA, and the rate at which positive supercoils are resolved (catalysis) (Reece et al., 2008). In this regard, more comprehensive models, which explicitly account for such parameters, have been developed (Bohrer & Roberts, 2016).

For example, a model of gene expression that accounts for a Gyrase binding site in the local DNA was developed in (Bohrer & Roberts, 2016). In this model (Table 6) two different species are defined (*RCoil* and *PCoil*), to track the supercoiling levels. The variable *RCoil* tracks the amount of the normal state of DNA ('regular') while *PCoil* tracks the amount of positive supercoiling. The total sum of *RCoil* and *PCoil* is constant throughout the simulation and equals the number of transcription events that can happen in the DNA domain before transcription locking occurs. In detail, in each transcription event there is an accumulation of *PCoil* (reaction 3.35), which can be relaxed by Gyrase intervention (reaction 3.36). To relax a *PCoil,* Gyrase must first bind to the DNA. Gyrase binding and unbinding to the DNA are modelled by reactions 3.37 and 3.38, respectively. Finally, translation, RNA and protein degradation are modelled as one-step processes (reactions 3.39-3.41).

This model assumes that negative supercoils are rapidly removed by Topo I (Cheng et al., 2003). When compared to solely ON-OFF and to two-step models, this model provides a more accurate representation of empirical data regarding the shape of the distribution of RNA numbers. Also, this model successfully captures the Fano factor of RNA numbers of highly transcribed genes (Geng et al., 2021; So et al., 2011; Taniguchi et al., 2010).

Table 6. Model of gene expression with local supercoiling effects in (Bohrer & Roberts, 2016). The variables *GYR* and *GYR'* stand for free Gyrase and Gyrase bound to the DNA, respectively. The species *RCoil* tracks the amount of 'regular' DNA state while *PCoil* tracks the amount of positive supercoiling.

Event	Reaction	
Transcription	$DNA + RCoil \rightarrow DNA + PCoil + RNA$	(3.35)
Relaxation of positive supercoils	$GYR' + PCoil \longrightarrow GYR' + RCoils$	(3.36)
Gyrase binding to DNA	$GYR \longrightarrow GYR'$	(3.37)
Gyrase unbinding	$GYR' \longrightarrow GYR$	(3.38)
Translation	$RNA \longrightarrow Protein$	(3.39)
RNA degradation	$RNA \longrightarrow \varnothing$	(3.40)
Protein degradation	$Protein \longrightarrow \varnothing$	(3.41)

In its current form, the model in Table 6 and similar ones (Ancona et al., 2019; Houdaigui et al., 2019), focus on quantifying the effect of supercoiling on transcription initiation. Other models focus instead on the effects of supercoiling

during transcription elongation (Chatterjee et al., 2021; Klindziuk & Kolomeisky, 2021; Sevier & Levine, 2017, 2018; Tripathi et al., 2022).

Recently, a comprehensive supercoiling-sensitive model including all transcription stages (initiation, elongation, and termination) was proposed (Figure 6) (Geng et al., 2021). This model includes an explicit description of the interaction of RNAP and DNA during all stages of transcription (Figure 6A-C), the formation of stochastic topological domains (Figure 6G), the diffusion of the supercoils within the domain (Figure 6D), and topoisomerase activities (Figure 6E and 6F).

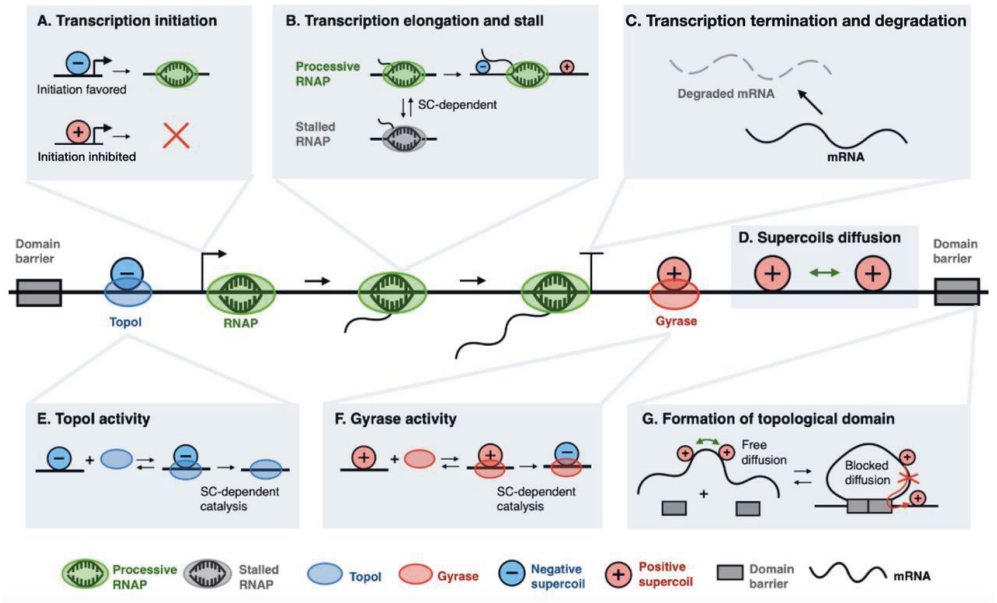

Figure 6. Illustration of a supercoiling model of transcription. **(A)** Transcription initiation is favored for a negatively supercoiled promoter (blue circle) over a positively supercoiled promoter (red circle). **(B)** RNAP translocation during elongation generates positive supercoiling downstream of the RNAP and negative supercoiling upstream of it. This affects the propensity of the RNAP to stall (grey) or continue elongation (green). **(C)** Transcription termination and subsequent RNA degradation. **(D)** Transcription and supercoils diffusion within a topological domain (delimited by domain barriers). Positive supercoils interact with Gyrase, whereas negative supercoils interact with Topoisomerase I (Topo I). **(E)** Topo I removes one negative supercoil per catalytic cycle. **(F)** Gyrase removes one positive supercoil and inserts one negative supercoil per catalytic cycle. **(G)** Formation and disintegration of a topological domain upon the binding and unbinding of nucleoid associated proteins (grey rectangles). Picture adapted with permission from (Geng et al., 2021).

The integration of all these events into the model allowed to reproduce quantitatively the empirical results on the collaborative dynamics of co-transcribing RNAPs (Kim et al., 2019). In detail, the group behavior exhibited by multiple RNAPs transcribing a DNA template within 3 kb can be explained by the annihilation of DNA supercoils between RNAPs. (Geng et al., 2021; Kim et al., 2019). Also, the existence of dynamic topological domains can reduce the RNA production within that domain, since having multiple genes in the same supercoiling domain results in each gene's expression influencing the expression of other genes.

Finally, the topological domain organization of the DNA also adds intrinsic noise to transcription, whose levels are dependent on the promoters'-strength (Geng et al., 2021). As such, modeling the interplay between transcription and supercoiling not only sheds light on how genes are regulated in bacteria, but also points the way towards experimental validation of supercoiling as a transcription regulator.

In **Publication II** we designed and implemented the first stochastic model of transcription with stepwise elongation, at the single nucleotide level, that, besides the events described in Table 4, it also includes the dynamic processes of accumulation and removal of positive supercoils.

3.3.2 Modelling Transcription Interference in Closely Spaced Promoters

In closely spaced promoters, transcription interference can occur through several mechanisms (Figure 7) (Callen et al., 2004; Shearwin et al., 2005; Sneppen et al., 2005). Out of the five transcription interference mechanisms in (Figure 7), three happen at the transcription initiation phase and two during elongation.

The first mechanism proposed is due to promoter competition (Figure 7A). In this mechanism, the occupation of one promoter by RNAP restricts the RNAP biding site of the second promoter. RNAP occupancy of the downstream promoter (also named 'sitting ducks', Figure 7B) happens when the elongation complex is slow to form and is hit by an RNAP transcribing from another promoter. Occlusion happens when the RNAP cannot bind to the promoter, because of the passage of an elongating RNAP (Figure 7C). However, this interference can be considered minor if the upstream promoter is not very strong (Shearwin et al., 2005). The mechanism of collision, (Figure 7D) refers to two converging elongating RNAPs.

Finally, the roadblock mechanism happens when a DNA-bound element (e.g., the DNA-bound Lac repressor) inhibits the progression of RNAP starting upstream of the binding site. An open complex on a promoter may also act as a roadblock.

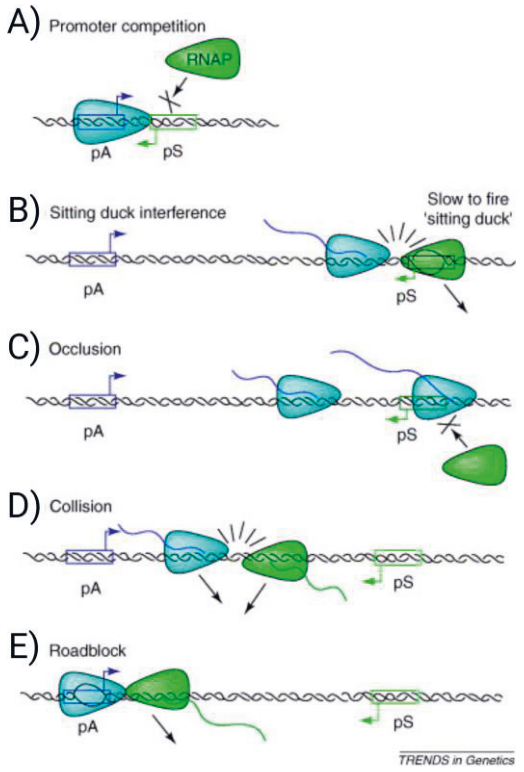

Figure 7. Mechanisms of transcriptional interference (TI). At least five mechanisms can generate TI: **(A)** promoter competition; **(B)** sitting duck interference; **(C)** occlusion; **(D)** collision; and **(E)** roadblock. The illustration shows two promoters (*pA* and *pS*) convergently oriented. All TI mechanisms presented may occur in promoters arrayed in a tandem format, except for the collision mechanism illustrated in D. For promoters arranged in a divergent configuration, only mechanism A applies. Picture adapted with permission from (Shearwin et al., 2005).

Palmer and colleagues (Palmer et al., 2009) proposed a model in which RNAPs move one base pair every time step. Three rate-limiting stages were included in the model: the reversible binding of RNAP to the promoter to create the closed complex, the isomerisation of the closed complex, and the open complex initiating elongation. However, stochastic simulations of the model were unable to explain the high interference of the weak promoter when located upstream the strong promoter, in a convergent configuration. This suggested the existence of another

mechanism of interference. After experimental validating by DNA footprinting that elongation complexes could be paused, a model with RNAP pausing on the downstream promoter was able to explain the *in vivo* data. As such, occlusion by pausing can be tuned to control the quantity of TI, by adjusting the pause-duration parameter. The fact that the quantity of repression can be tuned independently of promoter activity, indicates that the pauses are a robust and evolvable mechanism (Palmer et al., 2009).

Additionally, in (Martins et al., 2012) a single-nucleotide stochastic model of divergent and convergent closely spaced promoters was proposed (Figure 8). The results showed that interference between RNAPs in both these geometries increased the mean and standard deviation of the distribution of time intervals between RNA production events, from each TSS. Also, alterations in the distance between the TSSs led to sudden transitions in the RNA production kinetics, particularly when the configuration of the promoters changed from overlapped to non-overlapped.

Recently, a study investigated how closely spaced promoters are affected by the rate-limiting steps of each promoter's initiation kinetics (Häkkinen et al., 2019). The study predicted that the duration of such steps, namely the closed and open complex formation, can be used to control the kinetics of the temporal gaps between the transcription events of the opposing gene.

In addition to the two rate-limiting steps and transcription interference, when modelling gene pairs distancing few kilobases apart, one should also account with the mechanical features of the double DNA helix, namely the accumulation of supercoiling (Meyer et al., 2014; Yeung et al., 2017).

Finally, models incorporating supercoiling and promoter configurations demonstrated that, while RNAPs transcribing genes in tandem formation can cooperate, those transcribing genes in divergent or convergent configurations can act antagonistically, and that this behavior holds across a wide range of distance separations between TSSs (Tripathi et al., 2022; Yeung et al., 2017). These new models have been important in elucidating how the mechanical interaction between RNAPs and DNA can affect the transcriptional kinetics of closely spaced promoters.

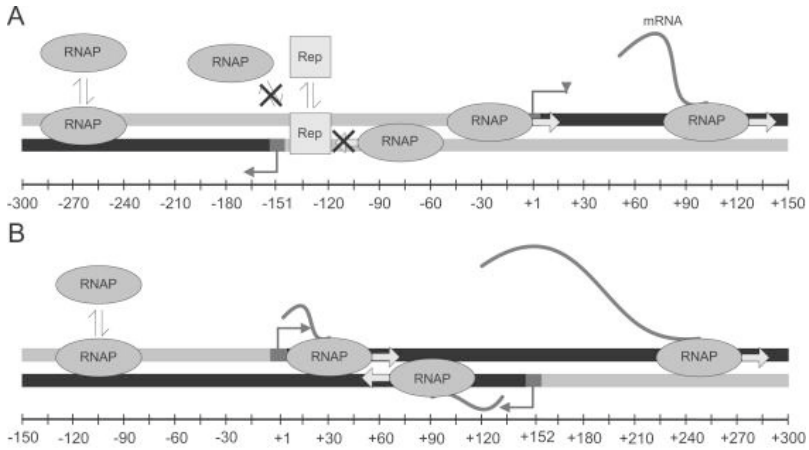

Figure 8. Schematic model of closely spaced promoters. **(A)** Divergent promoters. **(B)** Convergent promoters. Regions in black represent elongation regions. TSSs are represented by angled arrows. In the models, the first step is the binding of the RNAP to the DNA template. The RNAP progresses one nucleotide at a time in a certain direction. The RNAP was modelled to be able to unbind from the template strand at any time and, if multiple RNAPs are on the template, collisions can happen. If two RNAPs collide, one or both of them will fall off from the template, depending if they are both elongating or not. When the diffusing RNAP finds the TSS, the closed complex is formed followed by isomerization and formation of the open complex. After the open complex is formed, elongation can begin and the TSS is free for another RNAP to bind. The model also accounts with repression by one or more TFs (squared boxes). Picture adapted with permission from (Martins et al., 2012).

Based on these studies, in **Publication III** we explored stochastic models to characterize transcriptional interference in tandem overlapping and non-overlapping closely spaced promoters.

4 MATERIALS AND METHODS

This chapter presents an overview of the experimental and theoretical methods employed in this thesis. These methods comprise: fluorescence microscopy, single-cell RNA detection by the MS2 RNA detection system, image analysis for single-cell RNA quantification, single-cell protein expression measurements by flow cytometry, genome-wide analysis of differential gene expression by RNA-seq, measurements of relative gene expression levels by qPCR, use of Lineweaver-Burk plots to dissect the kinetics of the rate-limiting steps of transcription, and the use of external databases.

4.1 Fluorescent Proteins

The earliest account of bioluminescent species dates all the way back to the first century A.D., when the glowing jellyfish *Pulmo marinus* (now known as *Pelagia noctiluca*), living in the Mediterranean Sea, was first described (Harvey, 1957). Meanwhile, the first use of such organisms dates to the late 18th century, when coal miners used dried fish skins as light sources. This natural glowing phenomenon was termed 'fluorescence', in 1852 by the physicist George Stokes (Stokes, 1852).

However, the first study of fluorescent proteins did not occur until the 1960s, when Osamu Shimomura carried out the first experiments (Shimomura et al., 1962). Shimomura and colleagues studied the jellyfish *Aequorea*, out of which they extracted and purified the luminescent protein aequorin, along with a green fluorescence protein (GFP) (Shimomura, 1979, 2005). At the time of discovery, these two compounds were not accorded any special significance. Their potential only became apparent in later years.

Nowadays, fluorescent proteins are widely used in cell biology research, and are regarded as the foundation of fluorescence microscopy. This is mostly due to their versatility and specificity, along with the existence of molecular cloning methods that allow for the fusion of a fluorescent protein with a cellular protein of interest. Such capabilities contributed to the use of fluorescent proteins in different areas

such as studies of gene expression dynamics (Golding et al., 2005; Taniguchi et al., 2010; Yu et al., 2006), spatial localization of proteins and cellular compartments (Neeli-Venkata et al., 2016; Stracy et al., 2015; Weng et al., 2019), and molecular processes dynamics (Gupta et al., 2014; Mäkelä et al., 2021).

Over the years, fluorescence proteins have become widely used, and they now span the majority of the visible spectrum (Day & Davidson, 2009; Shaner et al., 2004). Furthermore, advancements in the area of fluorescent probing have led to the development and manipulation of a wide range of light-induced properties such as photoactivation, photoconversion, and photoswitching (Day & Davidson, 2009; Wu et al., 2011). These enable the change of fluorescence emission bandwidth or, in the case of photoswitchable fluorescent proteins, turn on or off the light emission, by employing appropriate lighting (Day & Davidson, 2009). These properties have been crucial in the development of super-resolution microscopy (Betzig et al., 2006; Rust et al., 2006; Wichmann & Hell, 1994).

When imaging fluorescent proteins, it is critical to choose the most suitable probes. For example, one should take into consideration brightness, wavelength, photostability, and toxicity (Shaner et al., 2005). The brightness (i.e., the fluorescence signal generated by the proteins) must be much greater than the cell background (referred to as cellular autofluorescence). To distinguish the two signals, it is recommended that the absorption and emission spectra of the chosen fluorophore do not overlap with the spectrum of cell autofluorescence (Ha & Tinnefeld, 2012). Additionally, the spectral shift between the excitation and emission wavelengths of the chosen fluorescent protein, referred to as the Stokes shift, should be considered (Valeur & Berberan-Santos, 2012). The larger the Stokes shift, the more distinct is the separation between the excitation and emission wavelengths.

When fluorescent proteins are excited over an extended period, they become photobleached (Shaner et al., 2005). The photobleaching rate of fluorescent proteins varies greatly. For example, when performing time-lapse experiments (where the capture of several images is required) photostability of the protein is essential. A measure of photostability is the photobleaching curve of a protein (Shaner et al., 2005).

Finally, wild-type fluorescent proteins are usually dimeric or tetrameric, which has been shown to be toxic for cells. As such, most wild-type fluorescent proteins need

to be optimized and engineered into monomers or tandem dimers (Shaner et al., 2004; Zhang et al., 2002).

In **Publication I-IV**, we used fluorescent probing. In detail, we use MS2-GFP (Golding et al., 2005) in **Publication I and II**, while in **Publication III and IV** we used a YFP fusion library with ~1000 native genes tagged with YFP (Taniguchi et al., 2010).

4.1.1 MS2 RNA Detection System

For many years, it was technically challenging to measure gene expression at the single-cell level. As such, most gene expression studies relied on population-level measurements, such as quantitative polymerase chain reaction (qPCR) (Raj & Van Oudenaarden, 2009). One of the technical advances that made single-cell measurements possible was the use of fluorescent proteins, such as GFP. The ease with which these molecules can be incorporated to measure gene expression allowed for the first studies of the mechanisms that generate transcriptional noise (Elowitz et al., 2002; Ozbudak et al., 2002).

One of the methods that allows real-time *in vivo* single RNA molecule studies is the MS2 RNA detection system. MS2 is a bacteriophage coat protein and was the first complete protein-coding gene sequence engineered (Jou et al., 1972). The MS2 RNA detection system was first used in yeast (Beach et al., 1999; Bertrand et al., 1998) and then adapted for *in vivo* usage in mammalian cells (Fusco et al., 2003) and *E. coli* (Golding et al., 2005; Golding & Cox, 2004).

The adaptation of this method for *E. coli* cells (Golding & Cox, 2004) is based on the synthetic engineering of a gene to transcribe a target RNA that contains multiple binding sites to the coat protein of the bacteriophage MS2. This gene is then induced to express in a cell that also contains a reporter system. This reporter system is responsible for the expression of the MS2 coat protein fused to GFP (MS2-GFP). The binding of the MS2-GFP molecules to the RNA multiple binding sites generates a fluorescent signal above the cell background level, which can be detected as bright spots by fluorescent microscopy (Figure 9B). Interestingly, when RNA molecules are bound to MS2-GFP, they become 'immortalized', meaning that they do not decay throughout the measurement (Golding et al., 2005).

One year later, Golding and colleagues (Golding et al., 2005) presented an upgraded version of this system, where the expression levels of the proteins translated from the target RNA could also be measured. In this version, the target RNA contains not only the MS2-GFP multiple binding sites, but also a distinct segment located after the RBS coding for RFP.

In **Publication I**, we used a variant of this system to detect single-cell, single-molecule RNAs. Specifically, one plasmid's reporter system that codes for the MS2-GFP molecules, under the control of the P_{BAD} promoter. On a second plasmid, the target gene (that codes for the target RNA containing the multiple MS2-GFP binding sites) is under the control of the $P_{LacO3O1}$ promoter. A schematic description of this construct is shown in Figure 9A. In **Publication II,** apart from using the same construct as in **Publication I**, we used a construct where the target gene is under the control of a plasmid-borne, P_{Lac} promoter. Moreover, in **Publication II**, we also engineered a strain where the target gene is integrated into the *lac* locus of the genome. Finally, in **Publication IV** we studied the genes' cold shock response by detecting changes in transcripts using RNA-seq. However, previous works have also studied the cold-shock response using the MS2 RNA detection system. (Oliveira et al., 2019).

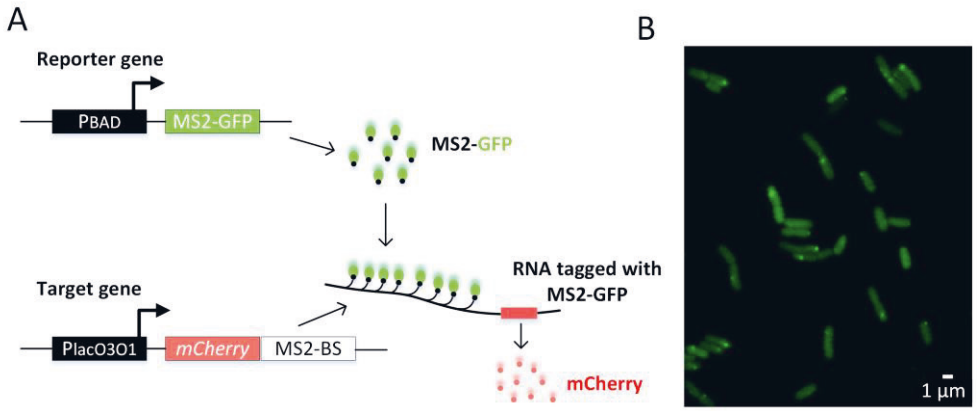

Figure 9. RNA quantification in individual cells by RNA tagging with MS2-GFP. **(A)** Schematic representation of the single-RNA MS2-GFP tagging detection system. Cells produce multiple MS2-GFP reporter proteins, under the control of P_{BAD}, while the production of RNAs target for MS2-GFP is under the control of $P_{LacO3O1}$. MS2-GFP molecules accumulate in the cytoplasm and bind to the target RNA upon its production. The mCherry region is translated into proteins that glow red after translation. **(B)** Example image of *E. coli* cells expressing RNAs tagged with MS2-GFP molecules. RNA molecules appear as fluorescent spots inside the cells.

4.2 Microscopy

Live cell fluorescence microscopy and transmitted light techniques are frequently used to study cellular activities. For optimal imaging, one should consider the optical properties of the image setup, the characteristics of the sample, and the viability of the specimen. Important optical characteristics to consider include the illumination wavelength and intensity, the spectra of filters and dichroic mirrors, the camera sensitivity to the emission wavelength, and the speed of acquisition (Shaner et al., 2005; Stephens & Allan, 2003). In terms of sample characteristics, one should evaluate the brightness of the signal and the speed of the process being observed. Also, regardless of the imaging technique, the cells' health and environment in the microscope stage must be considered. Microfluidics systems are one technique to ensure such control (Wang, et al., 2010).

The most often used fluorescence microscopy technique is wide-field microscopy. Widefield microscopes do not exclude light from any focal plane with the illumination bean spanning a volume of approximately $10 \times 10\ \mu m^2$ (Lang et al., 2006). This makes the acquisition time of widefield microscopes fast. However, the excitation of the entire specimen depth causes out-of-focus illumination decreasing the spatial resolution (Stephens & Allan, 2003).

To minimize the out-of-focus fluorescence and limit the duration of illumination, several other live cell fluorescence microscopy techniques were developed. These include confocal microscopy (Nakano, 2002; Pawley, 2006), total internal reflection florescence (TIRF) microscopy (Axelrod, 1981), and highly inclined and laminated optical sheets (HILO) (Konopka & Bednarek, 2008).

In **Publication I-IV**, cells were visualized using a Nikon Eclipse (Ti-E) inverted microscope equipped with a point scanning confocal microscope system, a HILO system, and a phase contrast system.

4.2.1 Scanning Confocal Microscopy

Confocal microscopes operate on the principle of focusing the illumination and detection optics on the same diffraction-limited point. In detail, a pinhole is used to reduce the out-of-focus light. This pinhole aperture assures that no light other than the illuminated point emission light reaches the detector (Pawley, 2006;

Stephens & Allan, 2003). The scanning of this diffraction-limited point over the sample (also denominated as point-scanning) allows building a complete image. Typically, modern confocal microscopes are controlled using an acousto-optic tunable filter, with the light beam scanning the stationary sample (Elliott, 2020). The acousto-optic tunable filter is responsible for turning lasers on and off, tuning laser power, and selecting the wavelength for imaging.

To build the complete image, a photomultiplier tube collects the fluorescence produced by the sample at each point. Photomultiplier tubes work as single-pixel cameras that amplify the emitted light using a photoelectric device. Next, a computer processes the change of the fluorescence over time into a two-dimensional picture (Elliott, 2020).

Overall, scanning confocal microscopes have a shallow depth of field, an overall better spatial resolution, and can collect serial optical sections from thick specimens (Pawley, 2006). To produce optical sections, the light beam scans the image in *x* and *y* directions of a single field of view, and then changes the z-stack and repeats the process. Upon the collection of images at several z-stacks, computer algorithms reconstruct a three-dimensional image of the sample (Elliott, 2020; Pawley, 2006).

However, confocal microscopy has fundamental limitations, such as the speed of data acquisition. To accommodate the need for fast acquisition, other types of confocal microscopes have been developed, such as the spinning disk and the swept field confocal microscopes (Castellano-Muñoz et al., 2012; Nakano, 2002).

In **Publications II-IV**, fluorescent proteins (GFP and YFP) were observed utilizing a point scanning confocal microscope to analyze both gene expression kinetics and/or protein spatial localization.

4.2.2 Highly Inclined and Laminated Optical Sheet Microscopy

Highly inclined and laminated optical sheet (HILO) microscopy is a technique in which the excitation light is refracted into the sample at a high inclination angle (Tokunaga et al., 2008). The inclination of the excitation beam minimizes the illumination area. Therefore, the background fluorescence from out of focus sections is reduced, allowing for better contrast and spatial resolution (Tokunaga et

al., 2008; Vignolini et al., 2018). HILO microscopy provides a means of performing fast widefield image acquisition with low signal/background ratio.

In **Publication I**, we employed HILO microscopy to detect MS2-GFP-tagged RNA molecules. This approach was selected because of its fast acquisition rates and high signal-to-noise ratio, as described above.

4.2.3 Phase-Contrast Microscopy

Microscopy transmitted light techniques are used to distinguish the morphological characteristics of the biological material such as cell size, structure, shape, and inclusion bodies. There are several different transmitted light techniques such as bright-field microscopy, differential interference contrast, and phase contrast (Stephens & Allan, 2003).

In all publications associated with this thesis, the transmitted light technique used was phase contrast. The working principle of phase contrast is the introduction of a phase shift in the light scattered by the sample (Zernike, 1942).

In detail, when the light wave is transmitted through a cell, its propagation speed decreases and therefore the scattered and background light will have a phase difference of approximately -90°. This generates a blurred image. However, the phase contrast technique uses phase-shift ring to also shift the background light by +90°. As a result, destructive interference occurs when background light and scattered light rays converge, making the cells darker than the background (Figure 1A) (Zernike, 1942, 1955).

4.3 Image Analysis and Data Extraction

Image analysis and signal processing techniques can be employed to retrieve single-cell data from microscopy images, such as cell area, the number of RNAs per cell, nucleoid size, and spatial location of Gyrase and RNAP. This section details the methodologies used in the publications associated with this thesis.

4.3.1 Cells and MS2 RNA Spots Segmentation

To process single-cell data from microscopy images, the cells must first be detected and segmented. The detection of cells is done from the phase-contrast images using automatic segmentation (Häkkinen et al., 2013).

The automatic segmentation algorithm identifies the cell regions by gradient path labeling (Mora et al., 2011) and uses classifiers for merging and discarding elements (Breiman et al., 2017). For each cell region identified, the algorithm draws a mask over it (Figure 10A). Next, principal component analysis is used to determine the position, orientation, and size of the cells inside each mask area.

After the automatic segmentation, the results go through manual correction, where the cells' borders are corrected, by removing, merging, or splitting (Häkkinen et al., 2013). Also, cells crossing the borders of the frames and their lineages are discarded.

To extract information, such as the number of RNAs per cell, the fluorescent images are aligned with the cell segmentation results. This step is essential given that the images are obtained from different detector systems and there is movement of cells during the measurements if the two images are not simultaneously taken.

The alignment consists of scaling and translation of the fluorescent image along the x- and y-axis. The optimal alignment result is the one that maximizes total fluorescence within the segmented cell areas. After the global alignment is automatically performed, manual local adjustments can also be applied (Häkkinen et al., 2013).

After image alignment, the MS2-GFP RNA spots in each cell must be detected and segmented to determine their intensities (Figure 10B). The automatic segmentation of the spots is performed using the Kernel Density Estimation (KDE) approach for spot detection (Ruusuvuori et al., 2010). This approach evaluates the probability density function of each spot's pixel intensities and finds a cut-off point that corresponds to the KDE's first local minimum (Otsu, 1979). Next, each pixel is evaluated and only segmented if its value exceeds the cut-off value.

To properly estimate the total MS2-GFP RNA spots fluorescence, one needs to account for the background fluorescence caused by the MS2-GFP molecules not bound to the RNA. This is accomplished by multiplying the average cell background fluorescence by the area of the spot, and then subtracting it from the total spot fluorescence intensity (Santinha et al., 2016).

The cell segmentation procedure was used in **Publications I-IV**. The RNA quantification method was used in **Publication I and II**. In addition, in **Publication II**, we performed time-lapse microscopy.

For time-lapse microscopy, two additional image analysis steps were done. First, before cell segmentation, the time-lapse images were temporally aligned using cross-correlation, to minimize the effects of cell drifting during measurement time. Second, a correspondence of the same cells between consecutive frames was established to create a temporal distribution of total fluorescence intensity, spot fluorescence intensity, and cell area for each cell.

4.3.2 Measurement of Time Intervals between RNA Production Events

The measurement of time interval distributions between consecutive events or waiting times has been shown to be important to understand events related to gene activation, DNA repair and cell-fate decision making (Norman et al., 2015; Uphoff et al., 2013). The use of MS2-GFP tagged RNAs over time can be used to quantify the kinetics of RNA production in single cells. In detail, after preliminary image processing, where cells and MS2-GFP RNA spots are segmented, one can extract the time intervals between consecutive RNA production events (Muthukrishnan et al., 2012).

The calculation of these time intervals is possible given that MS2-GFP tagged RNAs do not degrade over time (Golding et al., 2005). Thus, given that RNA production is not halted, one expects the total MS2-GFP RNA spot fluorescence intensity to increase over time with the production of new RNA molecules. Given that, the synthesis of a new RNA molecule results in a discrete increase in the cell's total spot intensity (Muthukrishnan et al., 2012).

To detect the discrete fluorescence intensity jumps in the MS2-GFP RNA spots, we fitted a monotone piecewise-constant function by least squares fit to the total

spot fluorescence of each cell, over time (Figure 10C) (Muthukrishnan et al., 2012). The plateaus in between distinct jumps of the total spots' fluorescence are utilized to determine the time intervals between successive RNA production events. Having this, in **Publication I**, we studied the statistics of the distributions of time intervals under various induction schemes.

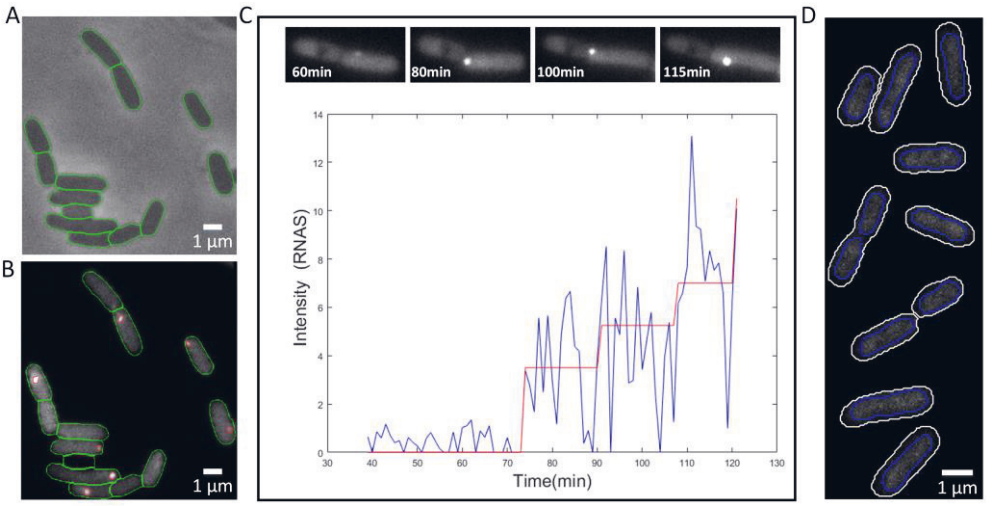

Figure 10. Cells, MS2-GFP RNA spots, and nucleoids segmentation. **(A)** Phase contrast image of *E. coli* cells along with results of automatic cell segmentation (green lines around the cells) (Häkkinen et al., 2013). Image acquired with a Nikon phase contrast system using a Nikon DS-Fi2 CCD camera **(B)** Confocal microscopy image of *E. coli* cells with a $P_{LacO3O1}$ promoter producing an RNA target for MS2-GFP. These results in MS2-GFP-RNA spots, which are detected and segmented by an automatic spot detection method (results of segmentation of spots and cells shown by red and green lines, respectively). Image acquired using confocal microscopy using a 100x objective, a 488 nm laser, and a 514/30 emission filter. **(C)** Example *E. coli* cell containing an MS2-GFP RNA spot, over time. The graph shows the total fluorescence intensity of the MS2-GFP RNA spot over time (blue line) and the result of the jump detection method (red line). **(D)** Visualization of nucleoids in *E. coli* cells with DAPI staining. Cell borders (white lines) were detected using an automatic cell segmentation algorithm (Häkkinen et al., 2013). Nucleoids were detected (blue lines) using an automatic nucleoid segmentation algorithm (Martins et al., 2018). Confocal microscopy was used to acquire this image, using a 100x objective, a 405 nm laser, and a 447/60 emission filter.

In **Publication II** the average time interval between consecutive RNA production events was instead estimated from the number of RNAs in cells obtained from microscopy images at two different time moments. First, for each time moment, from the histograms of fluorescence of MS2-GFP tagged RNAs, we estimated the

intensity of a single tagged RNA molecule (which should be approximately the same in both time moments). This is done by finding the intensity of the first histogram 'peak' (as it should correspond to the intensity of one MS2-GFP RNA molecule). Following that, we rounded the fluorescence intensity of each MS2-GFP RNA spot to the nearest integer, to estimate the number of RNA molecules contained in it, as in (Golding et al., 2005).

From the estimated mean number of RNAs per cell at the initial moment (t_0), and final moment (t), along with the rate of RNA dilution (k_d) due to cell division, we estimated the average rate of RNA production per cell (r_{rna}) using equation 4.1.

$$r_{rna} == \frac{k_d}{\ln 2} \cdot \frac{M - M_0 e^{-k_d \cdot t}}{\left[1 - e^{-k_d \cdot t}\right]} \tag{4.1}$$

Where M_0 and M are the mean number of RNAs per cell at moment t_0 and t, respectively. It is worth noting that one does not need to account for RNA degradation, since MS2-GFP RNAs are not degradable (Golding et al., 2005) and their fluorescence intensity is constant for the duration of the measurements (Tran et al., 2015). The standard error of the mean of the average rate of RNA production can be calculated using the Delta Method (Casella & Berger, 2002).

4.3.3 Nucleoid Visualization and Segmentation

To visualize, at the single-cell level, the cellular elements of interest we mostly made use of fluorescent proteins, such as GFP and YFP. However, fluorescent dyes capable of staining the desired cell structure may also be utilized (Davis & Gruebele, 2018; Schmidt et al., 1996).

As with fluorescent proteins, selecting the best appropriate fluorescent dye requires consideration of several parameters, such as the excitation and emission spectra, photostability, and compatibility with the target structure (Ha & Tinnefeld, 2012).

There are several fluorescent stains for nucleoid visualization, such as DAPI (4',6-diamidino-2-phenylindole), ethidium bromide, propidium bromide, among other (Chazotte, 2011; Nairin et al., 1982). However, DAPI has higher photostability when compared with other dyes (Chazotte, 2011). As such, in **Publication II and**

IV we used DAPI to visualize cells' nucleoids (Figure 10D) (Chazotte, 2011). After entering the cell, DAPI molecules attach to the A-T rich DNA sequences, forming a fluorescent complex (Kapuscinski, 1995) which is ~20-fold brighter than non-bound DAPI molecules (Chazotte, 2011).

To study the physical properties of the nucleoids (e.g., spatial location, area, and number) we first segmented the nucleoids using an automatic nucleoid segmentation software (Martins et al., 2018). To detect nucleoids the software applies a gradient path labeling algorithm that creates seeds based on the image intensity gradient (Ebright, 2000; Martins et al., 2018). Next, the segmentation is done based on a two-dimensional Gaussian profile and the global intensity of the neighboring pixels of the seed (Martins et al., 2018).

In addition, the maximum excitation wavelength of DAPI molecules is 450 nm which allows for the simultaneous labeling of other cellular structures (with e.g. GFP or YFP) (Kapuscinski, 1995). In **Publication III** we used this and, simultaneously with DAPI staining and strains from the YFP fusion library (Taniguchi et al., 2010), we quantified the co-localization of Gyrase and RNAP with the nucleoid(s).

4.4 Flow Cytometry

Cytometry refers to the measuring of physical and/or chemical features of individual cells (Shapiro, 2003). The first application of this method dates to the 1940's, when it was used to count and estimate the size of blood cells and bacteria (Shapiro, 2003).

In flow cytometer, the cell's characteristics are measured while cells travel through the equipment in a liquid stream (Shapiro, 2003). The purpose of the stream is to transport cells, one at a time, through the laser and detector systems. When a cell passes through the center of the laser beam, the detectors collect its optical signals (e.g., forward scatter light, side scatter light and emitted light).

Modern flow cytometers analyze tens of thousands of cells per minute, putting this approach squarely in the world of big data. Further, the raw single-cell data is provided in a format which is fast to process when compared with e.g., microscopy

data. The most frequent ways of presenting flow cytometry data are distributions, which are used to visualize a single parameter, and scatter plots, which are used to examine the correlation between two parameters.

However, similar to microscopy, when examining the single-cell distributions of protein expression levels measured using flow-cytometry, one needs to correct for the cell autofluorescence (Bahrudeen et al., 2019; Galbusera et al., 2020). For this, a measurement of control cells (i.e., without a fluorescent probe) is required.

Next, the four moments (mean, standard deviation, skewness, and kurtosis) of the single-cell distributions can be corrected. In detail, the mean fluorescence (*M*) can be corrected by applying equation 4.2 (Galbusera et al., 2020):

$$M_p = M_T - M_{cell} \tag{4.2}$$

Where, M_p is the mean cell fluorescence corrected for the cell autofluorescence (M_{cell}). Meanwhile, M_T is the total mean cell fluorescence measured by flow cytometry. Similarly, to correct the standard deviation (σ), one can apply equation 4.3 (Galbusera et al., 2020):

$$\sigma_p = \sqrt{\sigma_T^2 - \sigma_{cell}^2} \tag{4.3}$$

For correcting the skewness (*S*) of the distribution one can apply equation 4.4 as in (Bahrudeen et al., 2019) :

$$S_p = \frac{S_T \cdot {\sigma_T}^3 - S_{cell} \cdot {\sigma_{cell}}^3}{{\sigma_p}^3} \tag{4.4}$$

Finally, for correction of the kurtosis (*K*) one can apply equation 4.5.

$$K_p = \frac{K_T(\sigma_T)^4 - K_{cell}(\sigma_{cell})^4 - 6 \cdot \left(\sigma_p\right)^2 \cdot (\sigma_{cell})^2}{\left(\sigma_p\right)^4} \tag{4.5}$$

Along with correcting for cell autofluorescence, flow cytometry data should also be filtered for cell debris and doublets using gating techniques (Aghaeepour et al., 2013; Razo-Mejia et al., 2018).

In **Publication II-IV** we measured and corrected cells fluorescence levels (YFP or GFP) using an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA). We used a 488 nm laser for excitation and the fluorescein isothiocyanate detection channel (FITC-H) with a 530/30 nm filter for emission.

4.5 Quantitative Polymerase Chain Reaction

The quantitative polymerase chain reaction (qPCR) can be used to measure relative gene expression levels (Livak & Schmittgen, 2001; Schmittgen & Livak, 2008). To perform qPCR, the cells' RNA must first be extracted and converted to complementary DNA. (cDNA). Following that, the region of interest is amplified using a specific set of primers. The introduction of sequence-specific fluorescent probes enables real-time monitoring of this amplification during each cycle. As the amplification cycles occur, the total fluorescence increases. The number of cycles it takes to detect the signal from the sample is given by the threshold cycle value (C_T).

The method to analyze qPCR results is based on the C_T value (Livak & Schmittgen, 2001). In detail, a typical qPCR curve has an exponential phase and a plateau phase. The C_T measure is taken during the exponential phase where, for each cycle, the number of qPCR products doubles.

The method includes four steps. First, one averages the C_T of the repeats of each condition and the C_T of the respective reference conditions. Following this, the average C_T value of the target condition is normalized by the average C_T value of the reference condition (ΔC_T). This step allows normalizing the loading difference that could have occurred between conditions (Livak & Schmittgen, 2001). The next step is choosing the calibrator, which corresponds to the target condition for which the relative difference in expression is calculated. Once the calibrator is chosen, the ΔC_T value of the target condition is normalized by the ΔC_T value of the calibrator condition ($\Delta\Delta C_T$). Finally, to calculate the fold change between the target and calibrator, having in account that the product doubles for each amplification cycle, one should apply the equation below:

$$\text{Ratio (calibrator/target)} = \frac{2^{\Delta C_{T(calibrator)}}}{2^{\Delta C_{T(\text{target})}}} = 2^{\Delta\Delta C_T} \tag{4.6}$$

The result of equation 4.6 is the change in expression of the gene between the target and the calibrator, already normalized by the loading difference that could have occurred between conditions.

In **Publication II**, qPCR was employed to estimate the fold change in the expression levels of RNA coding for Gyrase, for different inducer concentrations, and to quantify the RNA production kinetics of $P_{LacO3O1}$ and P_{Lac} promoters, at different Gyrase concentrations.

4.6 RNA Sequencing

The ability to measure the transcriptome (all transcripts in a cell at the same time) is critical to understand the genome's global functioning. Measuring the transcriptome entails identifying a list of all transcript species (mRNAs, siRNAs, non-coding RNAs etc.) and quantifying their expression levels (Wang et al., 2009). Much effort has been done in the past fifty years to invent and improve technologies and sequencing protocols (see (Heather & Chain, 2016) comprehensive review) to quantify the transcriptome.

Microarray-based approaches are one of the first methods developed for identifying and quantifying the transcriptome as a whole. However, they require specific probes, limiting the ability to detect novel transcripts not covered by the probes (Wang et al., 2009). Additionally, these measurements by array hybridization techniques face technical challenges inherent to microarray probe performance, such as cross- and non-specific hybridization (Okoniewski & Miller, 2006; Royce et al., 2007), and the restricted detection range of individual probes (Wang et al., 2009). Finally, due to the narrow detection range, this approach performs poorly when quantifying low- and high-expression genes (Zhao et al., 2014).

Given the limitations of microarray methods, traditional sequencing technologies such as Sanger sequencing (Sanger et al., 1977) or tag-based methods (Harbers & Carninci, 2005; Velculescu et al., 1995) have been preferred. However, traditional Sanger sequencing has low throughput capabilities (i.e., is only capable of sequencing a single DNA fragment at a time) making it non-ideal for genome-wide studies. Additionally, although tag-based methods provide high throughput, a

substantial proportion of small tags cannot be accurately mapped to a single reference genome site (Wang et al., 2009).

These challenges were overcome by high-throughput technologies, such as RNA sequencing (RNA-seq). Since its development, approximately fifteen years ago (Emrich et al., 2007; Lister et al., 2008), RNA-seq has become the most widely used approach for determining the presence and amount of RNA in a biological sample.

The workflow of RNA-seq starts with the RNA extraction. Next, the RNA population is converted to cDNA fragments with sequencing adaptors added to each cDNA fragment. Each cDNA molecule is then sequenced (usually with a read depth of 10-30 million reads per sample) to obtain short sequence reads from one or both ends (single-end or pair-end sequencing). Following sequencing, the next steps are computational. First, the reads are aligned to a reference genome (or transcriptome) (Stark et al., 2019). Next, the reads that overlap transcripts are quantified, followed by filtering and normalization between samples. Finally, statistical analysis is performed to identify differential expressed genes (Stark et al., 2019).

As a technology, RNA-seq has several advantages: i) it is not restricted to recognizing transcripts expressed from known-sequence genes; ii) the transcription sites can be determined with single-base resolution; iii) it does not have an upper limit for quantification, as the expression levels correlate with the number of reads obtained; iv) has a large dynamic range of expression levels detectable; v) has high levels of reproducibility (Cloonan et al., 2008; Nagalakshmi et al., 2008)

The primary application of RNA-seq is genome-wide analysis of differential gene expression (Stark et al., 2019). Nevertheless, its use has also been extended to broader applications such as the study of RNA splicing (Wang et al., 2008) and of the function of non-coding and enhancer RNAs (Djebali et al., 2012; Li et al., 2016).

In **Publication III and IV** we used RNA-seq technology to perform a genome-wide analysis of differential gene expression in various conditions and over time.

4.7 Lineweaver-Burk Plot

Lineweaver-Burk plots were introduced to determine how an inhibitor competes with an enzyme (Lineweaver & Burk, 1934). In 1980, McClure used this strategy to quantify the duration of the rate-limiting steps of transcription initiation *in vitro* (i.e. RP_C and RP_O formation) (McClure, 1980).

This quantification was based on an abortive initiation reaction of promoter-bound RNAP. The abortive initiation assay consisted of the binding of the first two nucleoside triphosphates of an RNA sequence, when under RNAP saturation. After the binding of the nucleosides, a phosphodiester bond is created, and reaction products are produced. In the lack of additional nucleosides, the bound between the nucleosides is broken and initiation is aborted. After a period of time, a steady-state level of abortive product is reached (McClure et al., 1978). Assuming a two sequential states model (reaction 4.7), the total time until reaching steady state corresponds to the time it takes for RNAP to bind and isomerize into an open complex form (McClure, 1980).

$$Pro + RNAP \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} RP_C \xrightarrow{k_2} RP_O \tag{4.7}$$

The model in reaction 4.7 includes the promoter (Pro), RNAP ($RNAP$), closed complex formation (RP_C) and open complex formation (RP_O). From the derivation of the steady-state kinetics of the two-state model (equation 4.8) it is possible to estimate the rate of open and closed-complex formations. This is done from the calculation of the total time (τ_{TOTAL}) that it takes to reach steady state (equation 4.8), at different RNAP concentrations.

$$\tau_{TOTAL} = \underbrace{\frac{1}{k_2}}_{\tau_{OC}} + \underbrace{\frac{k_{-1} + k_2}{k_1 \cdot k_2 \cdot [RNAP]}}_{\tau_{CC}} \tag{4.8}$$

According to equation 4.8, the average time spent in RP_C formation (τ_{CC}) is proportional to [RNAP]$^{-1}$. Meanwhile, the time for RP_O formation (τ_{OC}) remains constant for different levels of [RNAP]. As such, the total time for transcription initiation should vary linearly with [RNAP]$^{-1}$. As such, the duration of the rate of open-complex and closed complex formation can be estimated using a Lineweaver-Burk plot between τ_{TOTAL} and the inverse of RNAP concentrations (Figure 11). In

the Lineweaver-Burk plot, the intersection of the best fit line with the *y*-axes gives an estimation for the time length of open complex formation (τ_{OC}), which equals $\frac{1}{k_2}$, according to equation 4.8. Meanwhile, the time for closed complex formation (τ_{CC}) can be estimated from $\tau_{OC} = \tau_{TOTAL} - \tau_{CC}$.

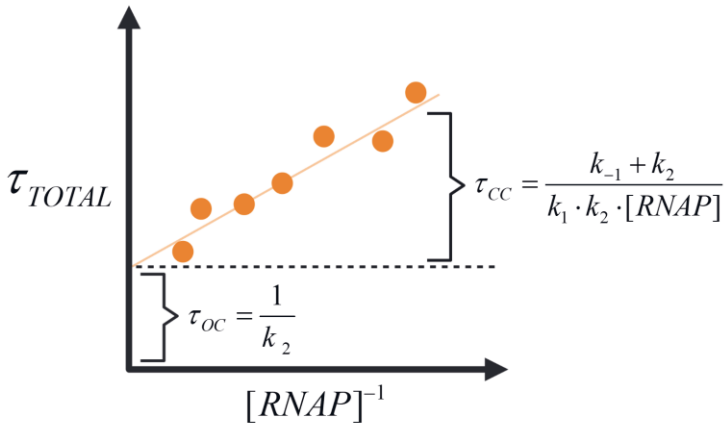

Figure 11. Schematic representation of a Lineweaver-Burk plot. The total time ($\tau_{TOTAL} = \tau_{OC} + \tau_{CC}$) is plotted against $[RNAP]^{-1}$. The intersection of the best fit line (orange line) with the *y*-axes gives an estimation for the time of open complex formation (τ_{OC}), which equals $\frac{1}{k_2}$ according to equation 4.8. For a certain RNAP concentration, the time length of closed complex formation (τ_{CC}) is $\frac{k_{-1} + k_2}{k_1 \cdot k_2 \cdot [RNAP]}$.

In **Publications I and II** a similar strategy was employed to quantify the *in vivo* kinetics of repression of the $P_{LacO3O1}$ promoter due to repressor molecules and due to PSB, respectively.

4.8 Research Output from Databases

In **Publication III and IV** we made use of data from three databases: RegulonDB (Santos-Zavaleta et al., 2019), gene ontology (Ashburner et al., 2000; The Gene Ontology Consortium., 2021) and NCBI (O'Leary et al., 2016).

From RegulonDB (Santos-Zavaleta et al., 2019)**,** we obtained the lists of all transcription units, operons, promoters, gene sequences and terminators that are empirically supported in the literature. From these lists, using *in-house* algorithms, in **Publication III** we identified genes controlled by tandem promoters (that differ by as few parameters as possible) along with their interactions with the gene regulatory network. In **Publication IV**, we also made use of these lists to study the operon organization, the sequence, and the interactions of our gene cohort of interest.

From the gene ontology database (Ashburner et al., 2000; The Gene Ontology Consortium., 2021), which contains gene's functional annotations, together with the classification systems (Mi et al., 2019), we performed enrichment analyses between the cohorts of interest and their biological processes. In particular, we investigated if there are any functional terms with more annotations in our cohort of interest than what would be expected by chance.

Finally, to measure the gene's evolutionary fitness we extracted information from 4133 reference bacterial DNAs from the NCBI database (O'Leary et al., 2016). Shortly, we used the number of DNAs that a gene is present in as a measure of its evolutionary fitness.

These curated biological databases are of high value, and became a critical component of our work by allowing us to relate the genes' dynamics to their sequence, function, structure, etc. To contribute to an open-access dissemination of results we have also deposited the data generated during our studies in certified repositories for open access, such as FlowRepository, Dryad and NCBI GEO.

5 RESULTS: SUMMARY AND CONCLUSIONS

In cells, transcription is a vital process. The regulatory mechanisms behind this process are the means by which cells adapt to changes, by activating or inhibiting, completely or partially, gene activities.

To comprehend the mechanisms regulating gene expression, apart from identifying the key molecules involved, one needs to characterize the kinetics. Here, we focused at the single-cell, single-gene level, on the dynamics of transcription repression mediated by: i) gene-specific regulatory molecules; ii) changes in the DNA supercoiling; iii) promoter arrangements, and iv) environmental changes.

In **Publication I**, we proposed a method to dissect the dynamics of transcription repression by TFs, based on *in vitro* dissection strategies of the kinetics of the underlying events of transcription. The new method for inferring the time-length spent by promoters in repressed state is based on *in vivo* measurements of RNA production kinetics at the single molecule level, at different repression strengths. In detail, we adapted a LineWeaver-Burk plot to infer from the data, the RNA production rate under no repression. This was then validated using a mutant strain lacking the repression mechanism (Figure 2, **Publication I**).

We found that the inverse of the RNA production rate varies linearly with the inverse of the inducer concentration, within a certain range. Further, on average, under full induction, the mean time interval between transcription events is ~1839 s, with the promoter spending 280 s (15%) of that time in an OFF state.

The method developed was only applied to the well-characterized $P_{LacO3O1}$ synthetic promoter and its repression mechanism. Nonetheless, we expect this technique to be applicable to a broad range of promoters. Additionally, one may use it to compare repressor/inducer binding rates, therefore contributing to our knowledge on how to optimize promoter efficiency.

In **Publication II**, we studied the dynamics of the events that result in transcription repression by PSB. Subsequently, we proposed a stochastic model of

transcription at the nucleotide level that includes all known events during transcription elongation, including PSB at the nucleotide level (Figure 1 and Table S2 of **Publication II**). Based on this model and its mean field approximation, we have developed a method for dissecting the rate-limiting steps due to PSB, and thus, allowing for the empirical quantification of this process' rate constants.

Namely, we again applied a LineWeaver-Burk plot and dissected the *in vivo* transcription rate in the absence of PSB (Figure 3C of **Publication II**). Next, we considered that the model assumes that, within the range of the Gyrase perturbation strengths, the kinetics of the steps in active transcription (e.g., closed and open complex formations) change much less than the kinetics of transcription locking. Based on this, we estimated the expected mean transcription rate for infinite Gyrase concentration. By confronting the estimations from the model, with the actual measurements of RNA production rates at the single RNA level, we estimated that the mean time spent in locked states due to PSB equaled ~735 s. We additionally estimated the number of transcription events between consecutive transcription locking events to be ~1.8 ± 0.84.

We validated the estimation of the mean time spent in locked states due to PSB using measurements of the kinetics of the same synthetic target gene when single-copy plasmid borne (rather than chromosome integrated), which we show to be free from PSB. We further validate the results with measurements of single RNA sensitivity of the time for the ceasing RNA production following the introduction in the media of Novobiocin, a Gyrase inhibitor (Figure 4 of **Publication II**). When analyzing the data, we assumed that there is only one copy of the target gene per cell. This has been validated by nucleoid DAPI staining and RT-qPCR in (Oliveira et al., 2019).

Having validated the methodology for dissecting the kinetics of locked states due to PSB, we investigated whether the dynamics of transcription locking due to PSB of a gene is affected by its own kinetics of basal transcription. This hypothesis follows from the reasoning that, if the basal transcription rate increases while Gyrase concentration is kept constant, then the time for Gyrase to resolve transcription locking due to PSB in between consecutive transcription events should decrease.

To validate this hypothesis, we replaced $P_{LacO3O1}$ by a weaker, native P_{Lac} promoter, and showed that its activity is less affected by changes in the concentration of Gyrases (Figure 5 of **Publication II**). Further, using the data from both constructs, we showed that it is possible to predict the state space of possible kinetics of transcription locking of genes in a given chromosomal location for varying basal transcription rates (Figure 6 of **Publication II**).

Overall, we expect that our findings will assist in resolving the kinetics of natural genes and circuits and, thus, provide better understanding of the transcription programs of *E. coli*. They should also contribute to the development of more efficient synthetic genetic circuits by accounting for the kinetics of PSB.

In **Publication III**, we studied the dynamics of transcription repression as a function of the promoters' spatial arrangement. In this work, tandem promoters were defined as two promoters in a 'head-to-tail' formation transcribing the same gene, in accordance with (Shearwin et al., 2005). To begin, we identified all native genes controlled by tandem promoters. We found a total of 102 genes controlled by tandem promoters in the genome of *E. coli,* whose transcription is not expected to suffer interference from the expression of neighboring genes (Section 'Selection of natural genes controlled by tandem promoters of S1 Appendix of **Publication III**).

Next, we collected RNA-seq data to investigate the effect of their input TFs. Under our measurement settings, we failed to find evidence that their dynamics are influenced by their input TFs (panel A in Figure D in the S2 Appendix of **Publication III**).

In addition, we searched for potential sequence-specific pauses in their DNA sequences and investigated their distances from oriC. Furthermore, we analyzed their biological functions and evolutionary conservation. We did not find any particular nucleotide sequence that may affect transcription elongation (Section 'Pause sequences' of S4 Appendix of **Publication III**). Also, no biological process was found to be overrepresented in the cohort of genes controlled by tandem promoters (S1 Table of **Publication III**). However, the mean conservation level of this cohort of genes is high, suggesting that they perform important biological roles (section 'Gene Conservation of S1 Appendix of **Publication III**).

Further, out of the 102 genes, we collected by flow cytometry the single-cell distributions of protein numbers of 30 of them. Considering the protein statistics, as well as all findings above, we proposed a novel analytical model for the kinetics of expression of genes regulated by tandem promoters. The model extends a 4-step standard model of transcription (Friedman et al., 2013; McClure, 1985) by including the mechanisms that lead to TI between the promoters.

Based on the model, if the TSSs are sufficiently close (dTSS < 35 bp), the binding of RNAP to a TSS will occlude the other TSS, preventing the closed complex formation. Meanwhile, if the TSSs are distant (dTSS > 35 bp), interference may still occur since the elongating RNAP starting from the upstream promoter can collide with RNAPs bound to the TSS of the downstream promoter. The effects of each of these occurrences is described using linear functions of the TSSs occupancy times, whereas the effect of dTSS is modeled using a continuous step function. All model rate constants are either based on known empirical parameter values or are extracted by best fitting to the empirical data.

From information on the distance between the two promoters, the model was shown to mimic the dynamics of gene expression (i.e., changes in mean and in variability) of tandem closely spaced promoters due to transcription interference. Consequently, our model can predict, prior to engineering, the dynamics of pairs of genes controlled by synthetically assembled tandem promoters.

In **Publication IV**, we characterized the transcription repression mechanism triggered by cold shock. Namely, we investigate what features allow some genes to become repressed, while other genes remain unresponsive during such perturbations.

First, we used RNA-seq to search for genes which are repressed by cold shock. We identified more than 300 short-term cold shock repressed genes (CSR). Next, we investigated their features and degree of TF regulation during cold shock. Namely, we investigated their ontology, promoter AT-richness, operon organization, and evolutionary fitness.

The function of CSR genes was found to be mainly associated with metabolism and response to stimulus, in accordance with (Gadgil et al., 2005; Phadtare & Inouye, 2004). Meanwhile, we failed to find a relationship between responses to cold shock, TF interactions, and promoter AT-richness. Moreover, while a few of

the identified CSR genes are likely responsive because they are in an operon with upstream CSR genes, the bulk appears to be independently sensitive to cold shock. Finally, their evolutionary fitness variability is lower than expected by chance, suggesting that the genes responsible for the cold-shock response have co-evolved.

Next, we used a YFP fusion library, in conjunction with a flow cytometer, to measure the effects of cold shock in the single-cell protein statistics of 30 CSR genes. Following cold shock, CSR genes rapidly decreased their expression level, while their noise relative to the mean expression increased (Figure 4 of **Publication IV**). This increase in noise is consistent with the emergence of transient locking in transcription.

Afterwards, we performed RNAseq under Gyrase inhibition and showed that many CSR genes are also supercoiling sensitive (Figure 6C of **Publication IV**). Focusing on dual responsive genes, we showed biophysical evidence that their temperature responsiveness is derived from their supercoiling sensitive responsiveness. In detail, we showed that the cells' nucleoid density increases (Figure 7A of **Publication IV**), their energy levels become depleted (Figure S13 of **Publication IV**), and the colocalization of Gyrases and the nucleoid increases (Figure 7D of **Publication IV**), in agreement with increased time length for resolving supercoils.

Finally, we proposed a general model of how cold shock responsiveness emerges from DNA relaxation (Figure 6A of **Publication IV**).

We expect our results to be of significance to resolve one of biology's current long-standing questions: which mechanisms control the ubiquitous cell response to cold and other stresses. Our results suggest that DNA relaxation is a key mechanism responsible for the genome-wide, short-term response of CSR genes. However, since not all CSR gene responses to cold shock can be explained by supercoiling sensitivity, we do not expect this to be the single mechanism governing the short-term response to cold shock.

Overall, the results of **Publications I-IV** contribute to the body of knowledge on the quantitative effects of the regulatory mechanisms of gene expression in *E. coli*. The focus was on the physical-based mechanisms, specifically, DNA torsion, promoters' spatial organization, and temperature-sensitivity in gene expression.

6 DISCUSSION

This thesis employed a multi-disciplinary strategy to investigate repression mechanisms, such as promoters' proximity and DNA relaxation, at the single-cell level in live *E. coli* cells. Using empirical data, we developed stochastic biophysical models of gene regulation. The new models capture the enhanced understanding on the influences of the phenomena on natural genes.

In detail, both **Publication I and II** propose strategies for assessing the time spent by synthetic promoters in OFF state due to two distinct repression mechanisms, respectively. The data is from single-cell, single-RNA microscopy images, using tagging of synthetic and native genes with fluorescent probes for RNA and proteins, respectively. Meanwhile, **Publication III and IV** investigate the behavior of gene cohorts, using genome-wide, RNA-seq data and single-cell flow-cytometry data of strain libraries with fluorescently tagged proteins.

This study produced several significant results. First, to our knowledge, for the first time we obtained the rates of the promoter kinetics of ON-OFF, from data from live individual cells. Second, the methods developed for this estimation can be used to characterize the transcription kinetics of other genes subject to repression of similar nature. Third, we obtained the first model of transcription that accounts for positive supercoils' accumulation and removal, at the nucleotide level.

Finally, we contributed to the solution of a challenging puzzle regarding the mechanisms used by cells to respond to stresses (Drolet, 2006). In particular, we showed that DNA supercoiling fluctuations ultimately contribute to the triggering of a multitude of genes that leads to physiological changes during cold shock. Specifically, we showed that correlations in genome-wide responses to a stress and to an antibiotic blocking a transcription step may identify if that step is involved in the stress response. This strategy may be applicable to identify factors that regulate genome-wide responses large-scale perturbations.

Conversely, it was evident that some of the methods have limitations. For instance, one limitation is that the OFF times quantification strategy requires that, for a

certain range of the state space, the transcription rates change linearly with the concentration of molecules involved. For example, the method proposed in **Publication II** is only valid within the range for which the transcription rate of a chromosome-integrated gene changes linearly with the inverse of Gyrase concentration. This limits the applicability of this approach to a subset of genes, since this change is unlikely to be linear when a gene is regulated by more than one TF (which is the case for at least 23 % of all genes in *E. coli* (Almeida et al., 2022)).

Another limitation pertains the estimate of the effect of changes in DNA relaxation levels. In detail, for these effects to be accurately estimated, one must account not only for the basal transcription activity of the gene, but also for the mean transcription activity within its DNA loop. However, the latter is not trivial because the locations of topological domains are dynamic (Lioy et al., 2018; Postow et al., 2004; Wang et al., 2013). More recent works have tried to account for this. For example, (Geng et al., 2021) proposes a model that, unlike ours, simulates multiple genes simultaneously and includes a stochastic formation and dissociation of topological barriers. Due to that, it can estimate the changes in gene-gene correlations due to supercoiling accumulation.

Nevertheless, considerable efforts remain necessary before the effects of supercoiling on transcription can be precisely quantified. For example, empirical data on the rates of formation and dissolution of topological domains formed by nucleoid associated proteins remain unknown. This is crucial information, as topological domains will only be influential if their rates are of similar order of magnitude as the transcription rates. Also, the dependency between DNA local transcription and the creation of topological domains in *E. coli* should be investigated, given that such correlation occurs in other bacteria (Le & Laub, 2016). Relevantly, assuming constant supercoiling diffusion rates along the DNA is not accurate since small DNA curvatures may suffice to generate supercoiling accumulation (Nelson, 1999). Further, the coexistence of negative and positive supercoiling makes it challenging to empirically distinguish between a reduction in negative supercoiling and a buildup in positive supercoiling. This hampers the modeling, since Topo I binding/unbinding rates are unknown. Finally, while not explicitly examined in this work, processes such as DNA replication, recombination and segregation also alter DNA supercoiling, hence indirectly impacting on transcription (Alberts, 2003; Higgins, 2007a).

In the long run, understanding bacterial mechanisms of adaptation to stresses will assist industrial and biomedical applications. For instance, when *E. coli* triggers its cold shock response, its growth rate is greatly reduced. Because of this, perishables have longer lifetime in cold conditions. However, cold conditions are energetically costly. To limit microbial proliferation and subsequent food spoilage in the absence of cold (thus, with lower energy costs), one could externally trigger natural bacterial programs to adapt to cold. Regarding biomedical applications, certain proteins involved in cold shock adaptation have been linked to pathogenic bacteria's virulence (Barria et al., 2013). In detail, the knockout of genes responsible to produce cold-shock proteins led to less virulent strains (Schärer et al., 2013). Thus, controlling such genes may allow modulating bacteria virulence levels.

In addition to temperature fluctuations, there are other stresses that bacteria commonly experience and must adapt to. For example, upon nutrient limitations bacteria growth rates decrease (Shehata & Marr, 1971), which enhances survival rates (Biselli et al., 2020). So far, research on these growth rates have relied mostly on proteomics data (Klumpp et al., 2008). It should be of interest to perform new studies of what regulates these physiological changes, using transcriptomics as the main data.

Overall, this project has generated new research questions, paving the way for future projects in synthetic biology and microbiology. For example, our small-scale models (the closely spaced interference and the PSB at the single-gene level) can be expanded to a genome-wide level to investigate how the 'micro phenomena' expands at larger scales. Such large-scale models could include not only single-gene features (gene location, configuration, orientation, operon organization etc.), but also account for TF-gene interactions and other GRN features, such as the type of interaction (activation or repression). These larger models, encompassing intrinsic and global features, may be able to explain the complexity required for the timely selection of short- and long-term gene cohorts.

Finally, this project may contribute to the design of synthetic circuits. Current designs are often limited to the use of promoters whose kinetics have been thoroughly studied. Our studies provided insight into new methodologies to build synthetic genes with a wide variety of predictable kinetics, by assembling promoters in a tandem formation at a specific nucleotide distance or by tuning the genes' supercoiling sensitivity (e.g., to be used as a source of noise). For the latter to be

implemented one can alter the genes' location (e.g., by placing genes in DNA locations known to have high expression levels) (Bryant et al., 2014), alter the promoter spacer and discriminator sequences (Klein et al., 2021 and Forquet et al., 2021), or rearrange genes in closely spaced configurations (Yeung et al., 2017). In the long-term, these techniques could contribute to the design of circuits with pre-defined dynamics for bio-industrial processes.

7 BIBLIOGRAPHY

Adachi, N., & Lieber, M. R. (2002). Bidirectional gene organization: A common architectural feature of the human genome. *Cell*, *109*(7), 807–809. https://doi.org/10.1016/s0092-8674(02)00758-4

Adhya, S., & Gottesman, M. (1982). Promoter occlusion: transcription through a promoter may inhibit its activity. *Cell*, 29(3), 939-944. https://doi.org/10.1016/0092-8674(82)90456-1

Aghaeepour, N., Finak, G., The FlowCAP Consortium, The DREAM Consortium, Hoos, H., Mosmann, T. R., Brinkman, R., Gottardo. R., & Scheuermann, H. R. (2013). Critical assessment of automated flow cytometry data analysis techniques. *Nature Methods 2013 10:3*, *10*(3), 228–238. https://doi.org/10.1038/nmeth.2365

Alberts, B. (2003). DNA replication and recombination. *Nature*, *421*(6921), 431–435. https://doi.org/10.1038/nature01407

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., & Walter, P. (2002). Molecular biology of the cell. *Garland Science*.

Almeida, B. L. B., Bahrudeen, M. N. M., Chauhan, V., Dash, S., Kandavalli, V., Häkkinen, A., Lloyd-Price, J., Palma, C. S. D., Baptista, I. S. C., Gupta, A., Kesseli, J., Dufour, E., Smolander, O.-P., Nykter, M., Auvinen, P., Jacobs, H. T., Oliveira, S. M. D., & Ribeiro, A. S. (2022). The transcription factor network of E. coli steers global responses to shifts in RNAP concentration. *bioRxiv*, 2022.03.07.483226. https://doi.org/10.1101/2022.03.07.483226

Ancona, M., Bentivoglio, A., Brackley, C. A., Gonnella, G., & Marenduzzo, D. (2019). Transcriptional Bursts in a Nonequilibrium Model for Gene Regulation by Supercoiling. *Biophysical Journal*, *117*(2), 369–376. https://doi.org/10.1016/j.bpj.2019.04.023

Andersen, K. B., & Von Meyenburg, K. (1980). Are growth rates of Escherichia coli in batch cultures limited by respiration? *Journal of Bacteriology*, *144*(1), 114. https://doi.org/10.1128/jb.144.1.114-123.1980

Antoun, A., Pavlov, M. Y., Lovmar, M., & Ehrenberg, M. (2006). How initiation factors tune the rate of initiation of protein synthesis in bacteria. *The EMBO Journal*, *25*(11), 2539–2550. https://doi.org/10.1038/sj.emboj.7601140

Arkin, A., Ross, J., & McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics*, *149*(4), 1633–1648. https://doi.org/10.1093/genetics/149.4.1633

Arsène, F., Tomoyasu, T., & Bukau, B. (2000). The heat shock response of Escherichia coli. *International Journal of Food Microbiology*, *55*(1–3), 3–9.

https://doi.org/10.1016/S0168-1605(00)00206-3

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., & Sherlock, G. (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics 2000 25:1*, *25*(1), 25–29. https://doi.org/10.1038/75556

Avery, O. T., Macleod, C. M., & McCarty, M. (1944). Studies on the chemical nature of the substance inducing transformation of pneumococcal types : induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type iii. *The Journal of Experimental Medicine*, *79*(2), 137. https://doi.org/10.1084/jem.79.2.137

Axelrod, D. (1981). Cell-substrate contacts illuminated by total internal reflection fluorescence. *The Journal of Cell Biology*, *89*(1), 141–145. http://www.ncbi.nlm.nih.gov/pubmed/7014571

Baeshen, M. N., Al-Hejin, A. M., Bora, R. S., Ahmed, M. M. M., Ramadan, H. A. I., Saini, K. S., Baeshen, N. A., & Redwan, E. M. (2015). Production of Biopharmaceuticals in E. coli: Current Scenario and Future Perspectives. *Journal of Microbiology and Biotechnology*, 25(7), 953–962. https://doi.org/10.4014/jmb.1412.12079

Bahrudeen, M. N. M., Chauhan, V., Palma, C. S. D., Oliveira, S. M. D., Kandavalli, V. K., & Ribeiro, A. S. (2019). Estimating RNA numbers in single cells by RNA fluorescent tagging and flow cytometry. *Journal of Microbiological Methods*, 105745. https://doi.org/10.1016/j.mimet.2019.105745

Bai, L., Santangelo, T. J., & Wang, M. D. (2006). Single-molecule analysis of RNA polymerase transcription. *Annual Review of Biophysics and Biomolecular Structure, 35*, 343–360.https://doi.org/10.1146/annurev.biophys.35.010406.150153

Baptista, I. S. C., & Ribeiro, A. S. (2020). Stochastic models coupling gene expression and partitioning in cell division in Escherichia coli. *Biosystems*, *193–194*, 104154. https://doi.org/10.1016/j.biosystems.2020.104154

Baptista, I. S. C., Kandavalli, V., Chauhan, V., Bahrudeen, M. N. M., Almeida, B. L. B., Palma, C. S. D., Dash, S., & Ribeiro, A. S. (2022). Sequence-dependent model of genes with dual σ factor preference. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 194812. https://doi.org/10.1016/J.BBAGRM.2022.194812

Barria, C., Malecki, M., & Arraiano, C. M. (2013). Bacterial adaptation to cold. Microbiology, 159, 2437–2443. https://doi.org/10.1099/min.0.052209-0

Beach, D. L., Salmon, E. D., & Bloom, K. (1999). Localization and anchoring of mRNA in budding yeast. *Current Biology , 9*(11), 569–578. https://doi.org/10.1016/S0960-9822(99)80260-7

Beck, C. F., & Warren, R. A. (1988). Divergent promoters, a common form of gene organization. *Microbiological Reviews*, *52*(3), 318–326. https://doi.org/10.1128/mr.52.3.318-326.1988

Berg, O. G. (1978). A model for the statistical fluctuations of protein numbers in a microbial population. *Journal of Theoretical Biology*, *71*(4), 587–603. https://doi.org/10.1016/0022-5193(78)90326-0

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. and Cohen, S .N. (2002) Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proceedings of the National Academy of Sciences, 99, 9697 LP – 9702. https://doi.org/10.1073/pnas.112318199

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H., & Long, R. M. (1998). Localization of ASH1 mRNA Particles in Living Yeast. *Molecular Cell*, *2*(4), 437–445. https://doi.org/10.1016/s1097-2765(00)80143-4

Betzig, E., Patterson, G. H., Sougrat, R., Lindwasser, O. W., Olenych, S., Bonifacino, J. S., Davidson, M. W., Lippincott-Schwartz, J., & Hess, H. F. (2006). Imaging intracellular fluorescent proteins at nanometer resolution. *Science*, *313*(5793), 1642–1645. https://doi.org/10.1126/science.1127344

Beuchat, L. R. (1996). Pathogenic Microorganisms Associated with Fresh Produce. *Journal of Food Protection*, *59*(2), 204–216. https://doi.org/10.4315/0362-028x-59.2.204

Beuchat, L. R. (2002). Ecological factors influencing survival and growth of human pathogens on raw fruits and vegetables. *Microbes and Infection*, *4*(4), 413–423. https://doi.org/10.1016/s1286-4579(02)01555-1

Biselli, E., Schink, S. J., & Gerland, U. (2020). Slower growth of Escherichia coli leads to longer survival in carbon starvation due to a decrease in the maintenance rate. Molecular Systems Biology, 16(6). https://doi.org/10.15252/MSB.20209478

Blake, W. J., Kærn, M., Cantor, C. R., & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature, 422:6932*, 633–637. https://doi.org/10.1038/nature01546

Blattner, F. R., Plunkett, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B., & Shao, Y. (1997). The complete genome sequence of Escherichia coli K-12. *Science*, *277*(5331), 1453–1462. https://doi.org/10.1126/science.277.5331.1453

Bohrer, C. H., & Roberts, E. (2016). A biophysical model of supercoiling dependent transcription predicts a structural aspect to gene regulation. *BMC Biophysics*, *9*(1). https://doi.org/10.1186/S13628-016-0027-0

Boles, T. C., White, J. H., & Cozzarelli, N. R. (1990). Structure of plectonemically supercoiled DNA. *Journal of molecular biology*, 213(4), 931-951. https://doi.org/10.1016/S0022-2836(05)80272-4

Borukhov, S., & Nudler, E. (2008). RNA polymerase: the vehicle of transcription. *Trends in Microbiology*, *16*(3), 126–134. https://doi.org/10.1016/j.tim.2007.12.006

Borukhov, S., & Severinov, K. (2002). Role of the RNA polymerase sigma subunit in transcription initiation. *Research in Microbiology*, *153*(9), 557–562. https://doi.org/10.1016/s0923-2508(02)01368-2

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (2017). Classification and regression trees. *Classification and Regression Trees*, 1–358. https://doi.org/10.1201/9781315139470

Brophy, J. A. N., & Voigt, C. A. (2016). Antisense transcription as a tool to tune gene expression. *Molecular Systems Biology*, *12*(1), 854. http://doi.org/10.15252/msb.20156540

Browning, D. F., & Busby, S. J. W. (2004). The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, *2*(1), 57–65. https://doi.org/10.1038/nrmicro787

Browning, D. F., & Busby, S. J. W. (2016). Local and global regulation of transcription initiation in bacteria. *Nature Reviews. Microbiology*, *14*(10), 638–650. https://doi.org/10.1038/nrmicro.2016.103

Bryant, J. A., Sellars, L. E., Busby, S. J. W. and Lee, D. J. (2014) Chromosome position effects on gene expression in Escherichia coli K-12. Nucleic Acids Res., 42, 11383–11392. https://doi.org/10.1093/nar/gku828

Burns, H., & Minchin, S. (1994). Thermal energy requirement for strand separation during transcription initiation: the effect of supercoiling and extended protein DNA contacts. *Nucleic Acids Research*, *22*(19), 3840. https://doi.org/10.1093/nar/22.19.3840

Caldas, T., Laalami, S., & Richarme, G. (2000). Chaperone properties of bacterial elongation factor EF-G and initiation factor IF2. *The Journal of Biological Chemistry*, *275*(2), 855–860. https://doi.org/10.1074/jbc.275.2.855

Callen, B. P., Shearwin, K. E., & Egan, J. B. (2004). Transcriptional interference between convergent promoters caused by elongation over the promoter. *Molecular Cell*, *14*(5), 647–656. https://doi.org/10.1016/j.molcel.2004.05.010

Cao-Hoang, L., Dumont, F., Marechal, P. A., & Gervais, P. (2010). Inactivation of Escherichia coli and Lactobacillus plantarum in relation to membrane permeabilization due to rapid chilling followed by cold storage. *Archives of Microbiology*, *192*(4), 299–305. https://doi.org/10.1007/s00203-010-0555-y

Casella, G., & Berger, R. L. (2002). *Statistical inference*. Thomson Learning.

Castellano-Muñoz, M., Peng, A. W., Salles, F. T., & Ricci, A. J. (2012). Swept Field Laser Confocal Microscopy for Enhanced Spatial and Temporal Resolution in Live-Cell Imaging. *Microscopy and Microanalysis*, *18*(4), 753–760. https://doi.org/10.1017/s1431927612000542

Chamberlin, M. J. (1974). The selectivity of transcription. *Annual Review of Biochemistry*, *43*(0), 721–775. https://doi.org/10.1146/annurev.bi.43.070174.003445.

Chargaff, E., Lipshitz, R., Green, C., & Hodes, M. E. (1951). The composition of the deoxyribonucleic acid of salmon sperm. *The Journal of Biological Chemistry*, *192*(1), 223–230.

Charlebois, D. A., Hauser, K., Marshall, S., & Balázsi, G. (2018). Multiscale effects of heating and cooling on genes and gene networks. *Proceedings of the National Academy of Sciences*, *115*(45), E10797–E10806. https://doi.org/10.1073/pnas.1810858115

Chatterjee, P., Goldenfeld, N., & Kim, S. (2021). DNA Supercoiling Drives a Transition between Collective Modes of Gene Synthesis. *Physical Review Letters*, *127*(21), 218101. https://doi.org/10.1103/PhysRevLett.127.218101

Chazotte, B. (2011). Labeling nuclear DNA using DAPI. *Cold Spring Harbor Protocols*, *2011*(1). https://doi.org/10.1101/pdb.prot5556

Chen, S. H., Chan, N. L., & Hsieh, T. S. (2013). New mechanistic and functional insights into DNA topoisomerases. *Annual Review of Biochemistry*, *82*, 139–170. https://doi.org/10.1146/annurev-biochem-061809-100002

Cheng, B., Zhu, C.-X., Ji, C., Ahumada, A., & Tse-Dinh, Y.-C. (2003). Direct interaction between Escherichia coli RNA polymerase and the zinc ribbon domains of DNA topoisomerase I. *The Journal of Biological Chemistry*, *278*(33), 30705–30710. https://doi.org/10.1074/jbc.M303403200

Cheung, K. J., Badarinarayana, V., Selinger, D. W., Janse, D., & Church, G. M. (2003). A Microarray-Based Antibiotic Screen Identifies a Regulatory Role for Supercoiling in the Osmotic Stress Response of Escherichia coli. *Genome Research*, *13*(2), 206–215. https://doi.org/10.1101/gr.401003

Chong, S., Chen, C., Ge, H., & Xie, X. S. (2014). Mechanism of Transcriptional Bursting in Bacteria. *Cell*, *158*(2), 314–326. https://doi.org/10.1016/j.cell.2014.05.038

Cloonan, N., Forrest, A. R. R., Kolle, G., Gardiner, B. B. A., Faulkner, G. J., Brown, M. K., Taylor, D. F., Steptoe, A. L., Wani, S., Bethel, G., Robertson, A. J., Perkins, A. C., Bruce, S. J., Lee, C. C., Ranade, S. S., Peckham, H. E., Manning, J. M., McKernan, K. J., & Grimmond, S. M. (2008). Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nature Methods, 5:7*, *5*(7), 613–619. https://doi.org/10.1038/nmeth.1223

Craig, E. A., & Gross, C. A. (1991). Is hsp70 the cellular thermometer? *Trends in Biochemical Sciences*, *16*(C), 135–140. https://doi.org/10.1016/0968-0004(91)90055-z

Craig, E. A., & Schlesinger, M. J. (1985). The heat shock response. *Critical Reviews in Biochemistry*, *18*(3), 239–280. https://doi.org/10.3109/10409238509085135

Crick, F. H. C., Barnett, L., Brenner, S., & Watts-Tobin, R. J. (1961). General Nature of the Genetic Code for Proteins. *Nature, 192*, 1227–1232. https://doi.org/10.1038/1921227a0

Crick, F. H. C. (1970). Central Dogma of Molecular Biology. *Nature*, 227, 561–563. https://doi.org/10.1038/227561a0

Dahm, R. (2005). Friedrich Miescher and the discovery of DNA. *Developmental Biology*, *278*(2), 274–288. https://doi.org/10.1016/j.ydbio.2004.11.028

Dangkulwanich, M., Ishibashi, T., Bintu, L., & Bustamante, C. (2014). Molecular mechanisms of transcription through single-molecule experiments. *Chemical*

Reviews, *114*(6), 3203–3223. https://doi.org/10.1021/cr400730x

Davis, C. M., & Gruebele, M. (2018). Labeling for Quantitative Comparison of Imaging Measurements in Vitro and in Cells. *Biochemistry*, *57*(13), 1929–1938. https://doi.org/10.1021/acs.biochem.8b00141

Day, R. N., & Davidson, M. W. (2009). The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*, *38*(10), 2887–2921. https://doi.org/10.1039/b901966a

DeHaseth, P. L., Zupancic, M. L., & Record Jr, M. T. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. *Journal of* bacteriology, 180(12), 3019-3025. https://doi.org/10.1128/jb.180.12.3019-3025.1998

De Jong, H. (2002). Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*, *9*(1), 67–103. https://doi.org/10.1089/10665270252833208

De Vries, R. (2010). DNA condensation in bacteria: Interplay between macromolecular crowding and nucleoid proteins. *Biochimie*, 92(12), 1715–1721. https://doi.org/10.1016/j.biochi.2010.06.024

Dillon, S. C., & Dorman, C. J. (2010). Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8(3), 185–195. https://doi.org/10.1038/nrmicro2261

Di Pietro, F., Brandi, A., Dzeladini, N., Fabbretti, A., Carzaniga, T., Piersimoni, L., Pon, C. L., & Giuliodori, A. M. (2013). Role of the ribosome-associated protein PY in the cold-shock response of Escherichia coli. *MicrobiologyOpen*, *2*(2), 293. https://doi.org/10.1002/mbo3.68

Djebali, S., Davis, C. A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., Xue, C., Marinov, G. K., Khatun, J., Williams, B. A., Zaleski, C., Rozowsky, J., Röder, M., Kokocinski, F., Abdelhamid, R. F., *et al. (*2012). Landscape of transcription in human cells. *Nature ,489,* 101–108. https://doi.org/10.1038/nature11233

Djordjevic, M., & Bundschuh, R. (2008). Formation of the Open Complex by Bacterial RNA Polymerase—A Quantitative Model. *Biophysical Journal*, *94*(11), 4233–4248. https://doi.org/10.1529/biophysj.107.116970

Dorman, C. J. (2019). DNA supercoiling and transcription in bacteria: a two-way street. *BMC Molecular and Cell Biology*, *20*(1), 26. https://doi.org/10.1186/s12860-019-0211-6

Dorman, C. J. (2013). Genome architecture and global gene regulation in bacteria: making progress towards a unified model? *Nature Reviews Microbiology* 2013, 11(5), 349–355. https://doi.org/10.1038/nrmicro3007

Dorman, C. J., & Dorman, M. J. (2016). DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophysical Reviews*, *8*, 89–100. https://doi.org/10.1007/s12551-016-0238-2

Drlica, K. (1992). Control of bacterial DNA supercoiling. *Molecular Microbiology*, *6*(4), 425–433. https://doi.org/10.1111/j.1365-2958.1992.tb01486.x

Drolet, M. (2006). Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology. *Molecular Microbiology*, *59*(3), 723–730. https://doi.org/10.1111/J.1365-2958.2005.05006.x

Duchi, D., Bauer, D. L. V., Fernandez, L., Evans, G., Robb, N., Hwang, L. C., Gryte, K., Tomescu, A., Zawadzki, P., Morichaud, Z., Brodolin, K., & Kapanidis, A. N. (2016). RNA Polymerase Pausing during Initial Transcription. *Molecular Cell*, *63*(6), 939–950. https://doi.org/10.1016/j.molcel.2016.08.011

Ebright, R. H. (2000). RNA Polymerase: Structural Similarities Between Bacterial RNA Polymerase and Eukaryotic RNA Polymerase II. *Journal of Molecular Biology*, *304*(5), 687–698. https://doi.org/10.1006/jmbi.2000.4309

Ebright, R. H., & Busby, S. (1995). The Escherichia coli RNA polymerase alpha subunit: structure and function. *Current Opinion in Genetics & Development*, *5*(2), 197–203. https://doi.org/10.1016/0959-437x(95)80008-5

El Hanafi, D., & Bossi, L. (2000). Activation and silencing of leu-500 promoter by transcription-induced DNA supercoiling in the Salmonella chromosome. *Molecular Microbiology*, *37*(3), 583–594. https://doi.org/ 10.1046/j.1365-2958.2000.02015.x

Eliasson, A., Bernander, R., Dasgupta, S., & Nordström, K. (1992). Direct visualization of plasmid DNA in bacterial cells. *Molecular Microbiology*, *6*(2), 165–170. https://doi.org/10.1111/j.1365-2958.1992.tb01997

Elliott, A. D. (2020). Confocal Microscopy: Principles and Modern Practices. *Current Protocols in Cytometry*, *92*(1), e68. https://doi.org/10.1002/cpcy.68

Elowitz, M. B., Levine, A. J., Siggia, E. D., & Swain, P. S. (2002). Stochastic Gene Expression in a Single Cell. *Science*, *297*(5584), 1183–1186. https://doi.org/10.1126/science.1070919

Emrich, S. J., Barbazuk, W. B., Li, L., & Schnable, P. S. (2007). Gene discovery and annotation using LCM-454 transcriptome sequencing. *Genome Research*, *17*(1), 69–73. https://doi.org/10.1101/gr.5145806

Epshtein, V., & Nudler, E. (2003). Cooperation Between RNA Polymerase Molecules in Transcription Elongation. *Science*, *300*(5620), 801–805. https://doi.org/10.1126/science.1083219

Erie, D. A., Hajiseyedjavadi, O., Young, M. C., & von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science*, *262*(5135), 867–873. https://doi.org/ 10.1126/science.8235608

Escherich, T. (1884). Klinisch-therapeutische beobachtungen aus der cholera-epidemie in Neapel. *Mun Med Wochenschrift*, 31, 561-4.

Escherich, T. (1988). The intestinal bacteria of the neonate and breast-fed infant. 1884. *Reviews of Infectious Diseases*, *10*(6), 1220–1225. https://doi.org/10.1093/clinids/10.6.1220

Esquerré, T., Laguerre, S., Turlan, C., Carpousis, A. J., Girbal, L., & Cocaign-

Bousquet, M. (2014). Dual role of transcription and transcript stability in the regulation of gene expression in Escherichia coli cells cultured on glucose at different growth rates. *Nucleic Acids Research*, *42*(4), 2460–2472. https://doi.org/10.1093/nar/gkt1150

Feklístov, A., Sharon, B. D., Darst, S. A., & Gross, C. A. (2014). Bacterial sigma factors: a historical, structural, and genomic perspective. *Annual Review of Microbiology*, *68*, 357–376. https://doi.org/10.1146/annurev-micro-092412-155737

Finney, A. H., Blick, R. J., Murakami, K., Ishihama, A., & Stevens, A. M. (2002). Role of the C-terminal domain of the alpha subunit of RNA polymerase in LuxR-dependent transcriptional activation of the lux operon during quorum sensing. *Journal of Bacteriology*, *184*(16), 4520–4528. https://doi.org/10.1128/jb.184.16.4520-4528.2002

Fogg, J. M., Judge, A. K., Stricker, E., Chan, H. L., & Zechiedrich, L. (2021). Supercoiling and looping promote DNA base accessibility and coordination among distant sites. *Nature Communications*, *12*(1), 1–16. https://doi.org/10.1038/s41467-021-25936-2

Forquet,R., Pineau,M., Nasser,W., Reverchon,S. and Meyer,S. (2021) Role of the Discriminator Sequence in the Supercoiling Sensitivity of Bacterial Promoters. *mSystems*, 6, e0097821. https://doi.org/10.1128/mSystems.00978-21

Friedman, L. J., Mumm, J. P., & Gelles, J. (2013). RNA polymerase approaches its promoter without long-range sliding along DNA. *Proceedings of the National Academy of Sciences*, *110*(24), 9740–9745. https://doi.org/10.1073/pnas.1300221110

Fujita, K., Iwaki, M., & Yanagida, T. (2016). Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. *Nature Communications*, *7*, 13788. https://doi.org/10.1038/ncomms13788

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H., & Bertrand, E. (2003). Single mRNA Molecules Demonstrate Probabilistic Movement in Living Mammalian Cells. *Current Biology*, *13*(2), 161–167. https://doi.org/10.1016/s0960-9822(02)01436-7

Gadgil, M., Kapur, V., & Hu, W. S. (2005). Transcriptional response of Escherichia coli to temperature shift. *Biotechnology Progress*, *21*(3), 689–699. https://doi.org/10.1021/bp049630l

Galbusera, L., Bellement-Theroue, G., Urchueguia, A., Julou, T., & van Nimwegen, E. (2020). Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. *PLOS ONE*, *15*(10). https://doi.org/10.1371/journal.pone.0240233

Gama-Castro, S., Salgado, H., Peralta-Gil, M., Santos-Zavaleta, A., Muniz-Rascado, L., Solano-Lira, H., Jimenez-Jacinto, V., Weiss, V., Garcia-Sotelo, J. S., Lopez-Fuentes, A., Porron-Sotelo, L., Alquicira-Hernandez, S., Medina-Rivera, A., Martinez-Flores, I., Alquicira-Hernandez, K., Martinez-Adame, R., Bonavides-Martinez, C., Miranda-Rios, J., Huerta, A. M., *et al.* (2011).

RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units). *Nucleic Acids Research*, *39*(Database issue), D98–D105. https://doi.org/10.1093/nar/gkq1110

Garcia, H. G., Sanchez, A., Kuhlman, T., Kondev, J., & Phillips, R. (2010). Transcription by the numbers redux: experiments and calculations that surprise. *Trends in Cell Biology*, *20*(12), 723–733. https://doi.org/10.1016/j.tcb.2010.07.002

Gardiner, C. W. (2004). Handbook of Stochastic Methods: for Physics, Chemistry, and the Natural Sciences*, Springer (third ed.)*, New York.

Gellert, M., Mizuuchi, K., O'Dea, M. H., Itoh, T., & Tomizawa, J. I. (1977). Nalidixic acid resistance: A second genetic character involved in DNA gyrase activity. *Proceedings of the National Academy of Sciences*, *74*(11), 4772. https://doi.org/10.1073/pnas.74.11.4772

Gellert, M., Mizuuchi, K., O'Dea, M. H., & Nash, H. A. (1976). DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences*, *73*(11), 3872–3876. https://doi.org/10.1073/pnas.73.11.3872

Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOld mine. *Nucleic Acids Research*, *49*(D1), D325–D334. https://doi.org/10.1093/nar/gkaa1113

Geng, Y., Bohrer, C. H., Yehya, N., Hendrix, H., Shachaf, L., Liu, J., Xiao, J., & Roberts, E. (2021). A spatially resolved stochastic model reveals the role of supercoiling in transcription regulation. *BioRxiv*. https://doi.org/10.1101/2021.12.29.474406

Georgopoulos, C. (2006). Toothpicks, serendipity and the emergence of the Escherichia coli DnaK (Hsp70) and GroEL (Hsp60) chaperone machines. *Genetics*, *174*(4), 1699–1707. https://doi.org/10.1534/genetics.104.68262

Gibson, M. A., & Bruck, J. (2000). Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *Journal of Physical Chemistry A*, *104*(9), 1876–1889. https://doi.org/10.1021/jp993732q

Gillespie, D. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81:2340–2361. https://doi.org/10.1021/j100540a008

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, *22*(4), 403–434. https://doi.org/10.1016/0021-9991(76)90041-3

Gillespie, D. T. (1992). A rigorous derivation of the chemical master equation. *Physica A: Statistical Mechanics and Its Applications*, *188*(1–3), 404–425. https://doi.org/10.1016/0378-4371(92)90283-v

Gillespie, D. T. (2001). Approximate accelerated stochastic simulation of chemically reacting systems. *The Journal of Chemical Physics*, *115*(4), 1716. https://doi.org/10.1063/1.1378322

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual Review of*

Physical Chemistry, *58*, 35–55. https://doi.org/10.1146/annurev.physchem.58.032806.104637

Giuliodori, A. M., Brandi, A., Gualerzi, C. O., & Pon, C. L. (2004). Preferential translation of cold-shock mRNAs during cold adaptation. *RNA*, *10*(2), 265. https://doi.org/10.1261/rna.5164904

Golding, I., & Cox, E. C. (2004). RNA dynamics in live Escherichia coli cells. *Proceedings of the National Academy of Sciences*, *101*(31), 11310–11315. https://doi.org/10.1073/pnas.0404443101

Golding, I., Paulsson, J., Zawilski, S. M., & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), 1025–1036. https://doi.org/10.1016/j.cell.2005.09.031

Golding, I. (2011). Decision Making in Living Cells: Lessons from a Simple System. *Annual Review of Biophysics*, *40*(1), 63–80. https://doi.org/10.1146/annurev-biophys-042910-155227

Goldman, S. R., Ebright, R. H., & Nickels, B. E. (2009). Direct Detection of Abortive RNA Transcripts in Vivo. *Science*, *324*(5929), 927–928. https://doi.org/10.1126/science.1169237

Gómez-Uribe, C. A., & Verghese, G. C. (2007). Mass fluctuation kinetics: capturing stochastic effects in systems of chemical reactions through coupled mean-variance computations. *The Journal of Chemical Physics*, *126*(2). https://doi.org/10.1063/1.2408422

Goncalves, N. S., Startceva, S., Palma, C. S., Bahrudeen, M. N., Oliveira, S. M., & Ribeiro, A. S. (2018). Temperature-dependence of the single-cell variability in the kinetics of transcription activation in Escherichia coli. *Physical Biology*, 15(2), 026007. https://doi.org/10.1088/1478-3975/aa9ddf

Gourse, R. L., Ross, W., & Gaal, T. (2000). UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition. *Molecular Microbiology*, *37*(4), 687–695. https://doi.org/10.1046/j.1365-2958.2000.01972.x

Greive, S. J., & von Hippel, P. H. (2005). Thinking quantitatively about transcriptional regulation. *Nature Reviews Molecular Cell Biology*, *6*(3), 221–232. https://doi.org/10.1038/nrm1588

Gries, T. J., Kontur, W. S., Capp, M. W., Saecker, R. M., & Thomas Recor, M. (2010). One-step DNA melting in the RNA polymerase cleft opens the initiation bubble to form an unstable open complex. *Proceedings of the National Academy of Sciences*, *107*(23), 10418–10423. https://doi.org/10.1073/pnas.1000967107

Grossman, A. D., Straus, D. B., Walter, W. A., & Gross, C. A. (1987). Sigma 32 synthesis can regulate the synthesis of heat shock proteins in Escherichia coli. *Genes & Development*, *1*(2), 179–184. https://doi.org/10.1101/gad.1.2.179

Gualerzi, C. O., Giuliodori, A. M., & Pon, C. L. (2003). Transcriptional and post-transcriptional control of cold-shock genes. *Journal of Molecular Biology*, *331*(3), 527–539. https://doi.org/10.1016/s0022-2836(03)00732-0

Guisbert, E., Herman, C., Lu, C. Z., & Gross, C. A. (2004). A chaperone network controls the heat shock response in E. coli. *Genes & Development*, *18*(22), 2812–2821. https://doi.org/10.1101/gad.1219204

Guisbert, E., Yura, T., Rhodius, V. A., & Gross, C. A. (2008). Convergence of molecular, modeling, and systems approaches for an understanding of the Escherichia coli heat shock response. *Microbiology and Molecular Biology Reviews* , *72*(3), 545–554. https://doi.org/10.1128/mmbr.00007-08

Gunasekera, T. S., Csonka, L. N., & Paliy, O. (2008). Genome-wide transcriptional responses of Escherichia coli K-12 to continuous osmotic and heat stresses. *Journal of Bacteriology*, *190*(10), 3712–3720. https://doi.org/10.1128/jb.01990-07

Gunnelius, L., Hakkila, K., Kurkela, J., Wada, H., Tyystjärvi, E., & Tyystjärvi, T. (2014). The omega subunit of the RNA polymerase core directs transcription efficiency in cyanobacteria. *Nucleic Acids Research*, *42*(7), https://doi.org/10.1093/nar/gku084

Gupta, A., Lloyd-Price, J., Neeli-Venkata, R., Oliveira, S. M. D., & Ribeiro, A. S. (2014). In vivo kinetics of segregation and polar retention of MS2-GFP-RNA complexes in Escherichia coli. *Biophysical Journal*, *106*(9), 1928–1937. https://doi.org/10.1016/j.bpj.2014.03.035

Guptasarma, P. (1996). Cooperative relaxation of supercoils and periodic transcriptional initiation within polymerase batteries. *BioEssays*, *18*(4), 325–332. https://doi.org/10.1002/bies.950180411

Ha, T., & Tinnefeld, P. (2012). Photophysics of Fluorescent Probes for Single-Molecule Biophysics and Super-Resolution Imaging. *Annual Review of Physical Chemistry*, *63*, 595–617. https://doi.org/10.1146/annurev-physchem-032210-103340

Häkkinen, A., Muthukrishnan, A.-B., Mora, A., Fonseca, J. M., & 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. https://doi.org/10.1093/bioinformatics/btt194

Häkkinen, A., Oliveira, S. M. D., Neeli-Venkata, R., & Ribeiro, A. S. (2019). Transcription closed and open complex formation coordinate expression of genes with a shared promoter region. *Journal of The Royal Society Interface*, *16*, 20190507. https://doi.org/10.1098/rsif.2019.0507

Hammar, P., Leroy, P., Mahmutovic, A., Marklund, E. G., Berg, O. G., & Elf, J. (2012). The lac repressor displays facilitated diffusion in living cells. *Science, 336*(6088), 1595–1598. https://doi.org/10.1126/science.1221648

Harbers, M., & Carninci, P. (2005). Tag-based approaches for transcriptome research and genome annotation. *Nature Methods*, *2*(7), 495–502. https://doi.org/10.1038/nmeth768

Harden, T. T., Herlambang, K. S., Chamberlain, M., Lalanne, J. B., Wells, C. D., Li, G. W., Landick, R., Hochschild, A., Kondev, J., & Gelles, J. (2020). Alternative transcription cycle for bacterial RNA polymerase. *Nature*

Communications, *11*(1), 1–11. https://doi.org/10.1038/s41467-019-14208-9

Harden, T. T., Wells, C. D., Friedman, L. J., Landick, R., Hochschild, A., Kondev, J., & Gelles, J. (2016). Bacterial RNA polymerase can retain σ^{70} throughout transcription. *Proceedings of the National Academy of Sciences*, *113*(3), 602–607. https://doi.org/10.1073/pnas.1513899113

Harvey, E. N. (1957). A history of luminescence from the earliest times until 1900. American Philosophical Society.

Haugen, S. P., Ross, W., & Gourse, R. L. (2008). Advances in bacterial promoter recognition and its control by factors that do not bind DNA. *Nature Reviews Microbiology*, *6*(7), 507–519. https://doi.org/10.1038/nrmicro1912

Hawley, D. K., Johnson, A. D., & McClure, W. R. (1985). Functional and physical characterization of transcription initiation complexes in the bacteriophage lambda OR region. *Journal of Biological Chemistry*, *260*(14), 8618–8626. https://doi.org/10.1016/s0021-9258(17)39517-0

Heather, J. M., & Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics*, *107*(1), 1–8. https://doi.org/10.1016/j.ygeno.2015.11.003

Helmann, J. D., & Chamberlin, M. J. (1988). Structure and function of bacterial sigma factors. *Annual Review of Biochemistry*, *57*, 839–872. https://doi.org/10.1146/annurev.bi.57.070188.004203

Henzinger, T. A., Mateescu, M., & Wolf, V. (2009). Sliding Window Abstraction for Infinite Markov Chains. *Lecture Notes in Computer Science*, *5643*, 337–352. https://doi.org/10.1007/978-3-642-02658-4_27

Herbert, K. M., La Porta, A., Wong, B. J., Mooney, R. A., Neuman, K. C., Landick, R., & Block, S. M. (2006). Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell*, *125*(6), 1083–1094. https://doi.org/10.1016/j.cell.2006.04.032

Herbert, K. M., Zhou, J., Mooney, R. A., Porta, A. La, Landick, R., & Block, S. M. (2010). E. coli NusG Inhibits Backtracking and Accelerates Pause-Free Transcription by Promoting Forward Translocation of RNA Polymerase. *Journal of Molecular Biology*, *399*(1), 17–30. https://doi.org/10.1016/j.jmb.2010.03.051

Herbert, M., Kolb, A., & Buc, H. (1986). Overlapping promoters and their control in Escherichia coli: The gal case. *Proceedings of the National Academy of Sciences*, *83*(9), 2807–2811. https://doi.org/10.1073/pnas.83.9.2807

Hespanha, J. P., & Singh, A. (2005). Stochastic models for chemically reacting systems using polynomial stochastic hybrid systems. *International Journal of Robust and Nonlinear Control*, *15*(15), 669–689. https://doi.org/10.1002/rnc.1017

Higgins, N. P. (2007a). Chromosome Structure. *eLS*. https://doi.org/10.1002/9780470015902.a0001486.pub2

Higgins, N. P. (2007b). Under DNA stress, gyrase makes the sign of the cross. Nature Structural & Molecular Biology, 14(4), 256-258.

https://doi.org/10.1038/nsmb0407-256

Houdaigui, B. El, Forquet, R., Hindré, T., Schneider, D., Nasser, W., Reverchon, S., & Meyer, S. (2019). Bacterial genome architecture shapes global transcriptional regulation by DNA supercoiling. *Nucleic Acids Research*, *47*(11), 5648–5657. https://doi.org/10.1093/nar/gkz300

Hraiky, C., Raymond, M. A., & Drolet, M. (2000). RNase H overproduction corrects a defect at the level of transcription elongation during rRNA synthesis in the absence of DNA topoisomerase I in Escherichia coli. *Journal of Biological Chemistry*, 275(15), 11257-11263. https://doi.org/10.1074/jbc.275.15.11257

Hsieh, L. S., Burger, R. M., & Drlica, K. (1991). Bacterial DNA supercoiling and [ATP]/[ADP]. Changes associated with a transition to anaerobic growth. *Journal of Molecular Biology*, *219*(3), 443–450. https://doi.org/10.1016/0022-2836(91)90185-9

Hsu, L. M. (2002). Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta*, *1577*(2), 191–207. https://doi.org/10.1016/s0167-4781(02)00452-9

Hsu, L. M., Vo, N. V., & Chamberlin, M. J. (1995). Escherichia coli transcript cleavage factors GreA and GreB stimulate promoter escape and gene expression in vivo and in vitro. *Proceedings of the National Academy of Sciences*, *92*(25), 11588–11592. https://doi.org/10.1073/pnas.92.25.11588

Huertas, P., & Aguilera, A. (2003). Cotranscriptionally formed DNA: RNA hybrids mediate transcription elongation impairment and transcription-associated recombination. *Molecular cell*, 12(3), 711-721. https://doi.org/10.1016/j.molcel.2003.08.010

Jacob, F., & Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *Journal of Molecular Biology*, *3*(3), 318–356. https://doi.org/10.1016/s0022-2836(61)80072-7

Jacob, F., Perrin, D., Sanchez, C., & Monod, J. (1960). Operon: a group of genes with the expression coordinated by an operator. *Comptes Rendus Hebdomadaires Des Seances de l'Academie Des Sciences*, *250*, 1727–1729.

Jeon, Y. H., Yamazaki, T., Otomo, T., Ishihama, A., & Kyogoku, Y. (1997). Flexible linker in the RNA polymerase alpha subunit facilitates the independent motion of the C-terminal activator contact domain. *Journal of Molecular Biology*, *267*(4), 953–962. https://doi.org/10.1006/jmbi.1997.0902

Jin. D. J., Burgess, R. R., Richardson, J. P., & Gross, C. A. (1992). Termination efficiency at rho-dependent terminators depends on kinetic coupling between RNA polymerase and rho. *Proceedings of the National Academy of Sciences*, *89*(4), 1453–1457. https://doi.org/10.1073/pnas.89.4.1453

Jones, D. L., Brewster, R. C., & Phillips, R. (2014). Promoter architecture dictates cell-to-cell variability in gene expression. *Science*, *346*(6216), 1533–1536. https://doi.org/10.1126/science.1255301

Jou, W. M., Haegeman, G., Ysebaert, M., Fiers, W., Jou, W. M., Haegeman, G.,

Ysebaert, M., & Fiers, W. (1972). Nucleotide Sequence of the Gene Coding for the Bacteriophage MS2 Coat Protein. *Nature*, *237*(5350), 82–88. https://doi.org/10.1038/237082a0

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: From theories to phenotypes. In *Nature Reviews Genetics*, 6(6), 451–464. https://doi.org/10.1038/nrg1615

Kapanidis, A. N., Margeat, E., Ho, S. O., Kortkhonjia, E., Weiss, S., & Ebright, R. H. (2006). Initial Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism. *Science*, *314*(5802), 1144–1147. https://doi.org/10.1126/science.1131399

Kapuscinski, J. (1995). DAPI: a DNA-specific fluorescent probe. *Biotechnic & Histochemistry* , *70*(5), 220–233. https://doi.org/10.3109/10520299509108199

Karpen, M. E., & DeHaseth, P. L. (2015). Base flipping in open complex formation at bacterial promoters. Biomolecules, 5(2), 668-678. https://doi.org/10.3390/biom5020668

Kepler, T. B., & Elston, T. C. (2001). Stochasticity in Transcriptional Regulation: Origins, Consequences, and Mathematical Representations. *Biophysical Journal*, *81*(6), 3116–3136. https://doi.org/10.1016/S0006-3495(01)75949-8

Kim, S., Beltran, B., Irnov, I., & Jacobs-Wagner, C. (2019). Long-Distance Cooperative and Antagonistic RNA Polymerase Dynamics via DNA Supercoiling. *Cell*, *179*(1), 106-119.e16. https://doi.org/10.1016/j.cell.2019.08.033

Kisselev, L. L., & Buckingham, R. H. (2000). Translational termination comes of age. *Trends in Biochemical Sciences*, *25*(11), 561–566. https://doi.org/10.1016/s0968-0004(00)01669-8

Klein,C.A., Teufel,M., Weile,C.J. and Sobetzko,P. (2021) The bacterial promoter spacer modulates promoter strength and timing by length, TG-motifs and DNA supercoiling sensitivity. *Scientific reports*, 11, 24399. https://doi.org/10.1038/s41598-021-03817-4

Klindziuk, A., & Kolomeisky, A. B. (2021). Long-Range Supercoiling-Mediated RNA Polymerase Cooperation in Transcription. *The Journal of Physical Chemistry B*, *125*(18), 4692–4700. https://doi.org/10.1021/acs.jpcb.1c01859

Klumpp, S., & Hwa, T. (2008). Growth-rate-dependent partitioning of RNA polymerases in bacteria. Proceedings of the National Academy of Sciences, 105(51), 20245–20250. https://doi.org/10.1073/PNAS.0804953105

Konopka, C. A., & Bednarek, S. Y. (2008). Variable-angle epifluorescence microscopy: a new way to look at protein dynamics in the plant cell cortex. *The Plant Journal*, *53*(1), 186–196. https://doi.org/10.1111/j.1365-313x.2007.03306.x

Korbel, J. O., Jensen, L. J., Von Mering, C., & Bork, P. (2004). Analysis of genomic context: Prediction of functional associations from conserved bidirectionally transcribed gene pairs. *Nature Biotechnology*, *22*(7), 911–917. https://doi.org/10.1038/nbt988

Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A., & Darst, S. A. (2000). A structural model of transcription elongation. *Science*, *289*(5479), 619–625. https://doi.org/10.1126/science.289.5479.619

Lang, E., Baier, J., & Köhler, J. (2006). Epifluorescence, confocal and total internal reflection microscopy for single-molecule experiments: a quantitative comparison. *Journal of Microscopy*, *222*(2), 118–123. https://doi.org/10.1111/j.1365-2818.2006.01579.x

Lawrence, J. G. (2003). Gene organization: selection, selfishness, and serendipity. *Annual Review of Microbiology*, *57*, 419–440. https://doi.org/10.1146/annurev.micro.57.030502.090816

Lawrence, J. G., & Roth, J. R. (1996). Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics*, *143*(4), 1843–1860. https://doi.org/10.1093/genetics/143.4.1843

Le, T. B., & Laub, M. T. (2016). Transcription rate and transcript length drive formation of chromosomal interaction domain boundaries. *The EMBO Journal*, *35*(14), 1582–1595. https://doi.org/10.15252/embj.201593561

Lee, D. J., Minchin, S. D., & Busby, S. J. W. (2012). Activating Transcription in Bacteria. *Annual Review of Microbiology*, *66*, 125–152. https://doi.org/10.1146/annurev-micro-092611-150012

Lelivelt, M. J., & Kawula, T. H. (1995). Hsc66, an Hsp70 homolog in Escherichia coli, is induced by cold shock but not by heat shock. *Journal of Bacteriology*, *177*(17), 4900–4907. https://doi.org/10.1128/jb.177.17.4900-4907.1995

Lenstra, T. L., & Larson, D. R. (2016). Single-Molecule mRNA Detection in Live Yeast. *Current Protocols in Molecular Biology*, 113. https://doi.org/10.1002/0471142727.mb1424S113

Levin, M. D. (2003). Noise in gene expression as the source of non-genetic individuality in the chemotactic response of Escherichia coli. *FEBS Letters*, *550*(1–3), 135–138. https://doi.org/10.1016/s0014-5793(03)00857-3

Lewin, B. (2008). Genes IX. *Jones and Bartlett Publishers*, USA.

Li, H., & Petzold, L. (2006). Logarithmic direct method for discrete stochastic simulation of chemically reacting systems. *Journal of Chemical Physic*s, 16, 1-11.

Li, W., Notani, D., & Rosenfeld, M. G. (2016). Enhancers as non-coding RNA transcription units: recent insights and future perspectives. *Nature Reviews Genetics*, *17*(4), 207–223. https://doi.org/10.1038/nrg.2016.4

Liang, S.-T., Bipatnath, M., Xu, Y.-C., Chen, S.-L., Dennis, P., Ehrenberg, M., & Bremer, H. (1999). Activities of constitutive promoters in Escherichia coli. *Journal of Molecular Biology*, *292*(1), 19–37. https://doi.org/10.1006/jmbi.1999.3056

Lim, B., Miyazaki, R., Neher, S., Siegele, D. A., Ito, K., Walter, P., Akiyama, Y., Yura, T., & Gross, C. A. (2013). Heat shock transcription factor σ^{32} co-opts the signal recognition particle to regulate protein homeostasis in E. coli. *PLOS Biology*, *11*(12). https://doi.org/10.1371/journal.pbio.1001735

Lima, C. D., Wang, J. C., & Mondragón, A. (1994). Three-dimensional structure of

the 67K N-terminal fragment of E. coli DNA topoisomerase I. *Nature*, *367*(6459), 138–146. https://doi.org/10.1038/367138a0

Lineweaver, H., & Burk, D. (1934). The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society*, *56*(3), 658–666. https://doi.org/10.1021/ja01318a036

Lioy, V. S., Cournac, A., Marbouty, M., Duigou, S., Mozziconacci, J., Espéli, O., Boccard, F., & Koszul, R. (2018). Multiscale Structuring of the E. coli Chromosome by Nucleoid-Associated and Condensin Proteins. *Cell*, *172*(4), 771-783. https://doi.org/10.1016/j.cell.2017.12.027

Lister, R., O'Malley, R. C., Tonti-Filippini, J., Gregory, B. D., Berry, C. C., Millar, A. H., & Ecker, J. R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. *Cell*, *133*(3), 523–536. https://doi.org/10.1016/j.cell.2008.03.029

Liu, L. F., Liu, C. C., & Alberts, B. M. (1980). Type II DNA topoisomerases: enzymes that can unknot a topologically knotted DNA molecule via a reversible double-strand break. *Cell*, *19*(3), 697–707. https://doi.org/10.1016/S0092-8674(80)80046-8

Liu, L. F., & Wang, J. C. (1987). Supercoiling of the DNA template during transcription. *Proceedings of the National Academy of Sciences*, *84*(20), 7024–7027. https://doi.org/10.1073/pnas.84.20.7024

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2−ΔΔCT Method. *Methods*, *25*(4), 402–408. https://doi.org/10.1006/meth.2001.1262

Lloyd-Price, J., Startceva, S., Kandavalli, V., Chandraseelan, J. G., Goncalves, N., Oliveira, S. M. D., Häkkinen, A., & Ribeiro, A. S. (2016). Dissecting the stochastic transcription initiation process in live Escherichia coli. *DNA Research*, *23*(3), 203–214. https://doi.org/10.1093/dnares/dsw009

López-Maury, L., Marguerat, S., & Bähler, J. (2008). Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nature Reviews Genetics*, *9*(8), 583–593. https://doi.org/10.1038/nrg2398

Lutz, R., Lozinski, T., Ellinger, T., & Bujard, H. (2001). Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Research*, *29*(18), 3873–3881.https://doi.org/10.1093/nar/29.18.3873

Ma, J., Bai, L., & Wang, M. D. (2013). Transcription Under Torsion. *Science*, *340*(6140), 1580–1583. https://doi.org/10.1126/science.1235441

Ma, J., & Wang, M. (2014). Interplay between DNA supercoiling and transcription elongation. *Transcription*, *5*(3), e28636. https://doi.org/10.4161/trns.28636

Maeda, H., Fujita, N., & Ishihama, A. (2000). Competition among seven Escherichia coli σ subunits: relative binding affinities to the core RNA polymerase. *Nucleic Acids Research*, *28*(18), 3497–3503. https://doi.org/10.1093/nar/28.18.3497

Mäkelä, J., Kandavalli, V., & Ribeiro, A. S. (2017). Rate-limiting steps in

transcription dictate sensitivity to variability in cellular components. *Scientific Reports*, *7*(1), 10588. https://doi.org/10.1038/s41598-017-11257-2

Mäkelä, J., Lloyd-Price, J., Yli-Harja, O., & Ribeiro, A. S. (2011). Stochastic sequence-level model of coupled transcription and translation in prokaryotes. *BMC Bioinformatics*, *12*(1), 121. https://doi.org/10.1186/1471-2105-12-121

Mäkelä, J., Uphoff, S., & Sherratt, D. J. (2021). Nonrandom segregation of sister chromosomes by Escherichia coli MukBEF. *Proceedings of the National Academy of Sciences*, *118*(33). https://doi.org/10.1073/pnas.2022078118

Malik, M., Bensaid, A., Rouviere-Yaniv, J., & Drlica, K. (1996). Histone-like protein HU and bacterial DNA topology: suppression of an HU deficiency by gyrase mutations. *Journal of Molecular Biology*, 256(1), 66–76. https://doi.org/10.1006/jmbi.1996.0068

Manley, S., Gillette, J. M., Patterson, G. H., Shroff, H., Hess, H. F., Betzig, E., & Lippincott-Schwartz, J. (2008). High-density mapping of single-molecule trajectories with photoactivated localization microscopy. *Nature Methods*, 5(2), 155–157. https://doi.org/10.1038/nmeth.1176

Mansilla, M. C., Cybulski, L. E., Albanesi, D., & De Mendoza, D. (2004). Control of membrane lipid fluidity by molecular thermosensors. *Journal of Bacteriology*, *186*(20), 6681–6688. https://doi.org/10.1128/jb.186.20.6681-6688.2004

Marr, M. T., & Roberts, J. W. (2000). Function of transcription cleavage factors GreA and GreB at a regulatory pause site. *Molecular Cell*, *6*(6), 1275–1285. https://doi.org/10.1016/S1097-2765(00)00126-x

Martens, J. A., Laprade, L., & Winston, F. (2004). Intergenic transcription is required to repress the Saccharomyces cerevisiae SER3 gene. *Nature*, *429*(6991), 571–574. https://doi.org/10.1038/nature02538

Martin, F. H., & Tinoco, I. (1980). DNA-RNA hybrid duplexes containing oligo(dA:rU) sequences are exceptionally unstable and may facilitate termination of transcription. *Nucleic Acids Research*, *8*(10), 2295. https://doi.org/10.1093/nar/8.10.2295

Martínez-Antonio, A., & Collado-Vides, J. (2003). Identifying global regulators in transcriptional regulatory networks in bacteria. *Current Opinion in Microbiology*, *6*(5), 482–489. https://doi.org/10.1016/j.mib.2003.09.002

Martins, L., Mäkelä, J., Häkkinen, A., Kandhavelu, M., Yli-Harja, O., Fonseca, J. M., & Ribeiro, A. S. (2012). Dynamics of transcription of closely spaced promoters in Escherichia coli, one event at a time. *Journal of Theoretical Biology*, *301*, 83–94. https://doi.org/10.1016/j.jtbi.2012.02.015

Martins, L., Neeli-Venkata, R., Oliveira, S. M. D., Häkkinen, A., Ribeiro, A. S., & Fonseca, J. M. (2018). SCIP: a single-cell image processor toolbox. *Bioinformatics*, *34*(24), 4318–4320. https://doi.org/10.1093/bioinformatics/bty505

McAdams, H. H., & Arkin, A. (1997). Stochastic mechanisms in gene expression. *Proceedings of the National Academy of Sciences*, *94*(3), 814–819. https://doi.org/10.1073/pnas.94.3.814

McAdams, H. H., & Arkin, A. (1999). It's a noisy business! Genetic regulation at the nanomolar scale. *Trends in Genetics*, *15*(2), 65–69. https://doi.org/10.1016/s0168-9525(98)01659-x

McClure, W. R. (1980). Rate-limiting steps in RNA chain initiation. *Proceedings of the National Academy of Sciences*, *77*(10), 5634–5638. https://doi.org/10.1073/pnas.77.10.5634

McClure, W. R. (1985). Mechanism and Control of Transcription Initiation in Prokaryotes. *Annual Review of Biochemistry*, *54*(1), 171–204. https://doi.org/10.1146/annurev.bi.54.070185.001131

McClure, W. R., Cech, C. L., & Johnston, D. E. (1978). A steady state assay for the RNA polymerase initiation reaction. *Journal of Biological Chemistry*, *253*(24), 8941–8948.

McLeod, S. M., & Johnson, R. C. (2001). Control of transcription by nucleoid proteins. *Current Opinion in Microbiology*, *4*(2), 152–159. https://doi.org/10.1016/S1369-5274(00)00181-8

Mekler, V., Kortkhonjia, E., Mukhopadhyay, J., Knight, J., Revyakin, A., Kapanidis, A. N., Niu, W., Ebright, Y. W., Levy, R., & Ebright, R. H. (2002). Structural organization of bacterial RNA polymerase holoenzyme and the RNA polymerase-promoter open complex. *Cell*, *108*(5), 599–614. https://doi.org/10.1016/S0092-8674(02)00667-0

Meyer, S., Beslon, G., Coulon, A., Vallin, E., & Morin, V. (2014). Torsion-Mediated Interaction between Adjacent Genes. *PLOS Computational Biology*, *10*(9), e1003785. https://doi.org/10.1371/journal.pcbi.1003785

Miller, O. L., Hamkalo, B. A., & Thomas, C. A. (1970). Visualization of bacterial genes in action. *Science*, *169*(943), 392–395. https://doi.org/10.1126/science.169.3943.392

Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., & Alon, U. (2002). Network motifs: Simple building blocks of complex networks. *Science*, *298*(5594), 824–827. https://doi.org/ 10.1126/science.298.5594.824

Missiakas, D., Georgopoulos, C., & Raina, S. (1993). The Escherichia coli heat shock gene htpY: mutational analysis, cloning, sequencing, and transcriptional regulation. *Journal of Bacteriology*, *175*(9), 2613–2624. https://doi.org/10.1128/jb.175.9.2613-2624.1993

Miyazaki, R., Yura, T., Suzuki, T., Dohmae, N., Mori, H., & Akiyama, Y. (2016). A Novel SRP Recognition Sequence in the Homeostatic Control Region of Heat Shock Transcription Factor σ^{32}. *Scientific Reports, 6, 24147* https://doi.org/10.1038/srep24147

Moore, L. J., & Kiley, P. J. (2001). Characterization of the Dimerization Domain in the FNR Transcription Factor. *Journal of Biological Chemistry*, *276*(49), 45744–45750. https://doi.org/10.1074/jbc.m106569200

Mora, A. D., Vieira, P. M., Manivannan, A., & Fonseca, J. M. (2011). Automated drusen detection in retinal images using analytical modelling algorithms. *BioMedical Engineering OnLine*, *10*, 59. https://doi.org/10.1186/1475-925x-10-

59

Müller-Hill, B. (1996). *The lac Operon : a short history of a genetic paradigm*. de Gruyter, Hawthorne, NY. ISBN 3-11014830-7

Munsky, B., & Khammash, M. (2006). The finite state projection algorithm for the solution of the chemical master equation. *The Journal of Chemical Physics*, *124*(4), 044104. https://doi.org/10.1063/1.2145882

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., & Darst, S. A. (2002). Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex. *Science*, *296*(5571), 1285–1290. https://doi.org/10.1126/science.1069595

Muthukrishnan, A. B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O., & Ribeiro, A. S. (2012). Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells. *Nucleic Acids Research*, *40*(17), 8472–8483. https://doi.org/10.1093/nar/gks583

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., & Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. *Science*, *320*(5881), 1344–1349. https://doi.org/10.1126/science.1158441

Nairin, R. S., Dodson, M. L., & Humphrey, R. M. (1982). Comparison of ethidium bromide and 4'6-diamidino-2-phenylindole as quantitative fluorescent stains for DNA in agarose gels. *Journal of Biochemical and Biophysical Methods*, *6*(2), 95–103. https://doi.org/10.1016/0165-022x(82)90055-0

Nakano, A. (2002). Spinning-disk confocal microscopy - A cutting-edge tool for imaging of membrane traffic. *Cell Structure and Function*, *27*(5), 349–355. https://doi.org/10.1247/csf.27.349

Neeli-Venkata, R., Startceva, S., Annila, T., & Ribeiro, A. S. (2016). Polar Localization of the Serine Chemoreceptor of Escherichia coli Is Nucleoid Exclusion-Dependent. *Biophysical Journal*, *111*(11), 2512–2522. https://doi.org/10.1016/j.bpj.2016.10.024

Neidhardt, F. C., VanBogelen, R. A., & Lau, E. T. (1983). Molecular cloning and expression of a gene that controls the high-temperature regulon of Escherichia coli. *Journal of Bacteriology*, *153*(2), 597–603. https://doi.org/10.1128/jb.153.2.597-603.1983

Nelson, P. (1999). Transport of torsional stress in DNA. *Proceedings of the National Academy of Sciences*, *96*(25), 14342–14347. https://doi.org/10.1073/pnas.96.25.14342

Nickels, B. E., Dove, S. L., Murakami, K. S., Darst, S. A., & Hochschild, A. (2002). Protein-protein and protein-DNA interactions of sigma70 region 4 involved in transcription activation by lambdacI. *Journal of Molecular Biology*, *324*(1), 17–34. https://doi.org/10.1016/s0022-2836(02)01043-4

Nöllmann, M., Stone, M. D., Bryant, Z., Gore, J., Crisona, N. J., Hong, S. C., Mitelheiser, S., Maxwell, A., Bustamante, C., & Cozzarelli, N. R. (2007). Multiple modes of Escherichia coli DNA gyrase activity revealed by force and

torque. *Nature Structural & Molecular Biology*, *14*(4), 264–271. https://doi.org/10.1038/nsmb1213

Norman, T. M., Lord, N. D., Paulsson, J., & Losick, R. (2015). Stochastic Switching of Cell Fate in Microbes. *Annual Review of Microbiology*, 69(1), 381–403. https://doi.org/10.1146/annurev-micro-091213-112852

Novick, A., & Weiner, M. (1957). Enzyme Induction as an all-or-none phenomenon. *Proceedings of the National Academy of Sciences*, *43*(7), 553–566. https://doi.org/10.1073/pnas.43.7.553

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., *et al.* (2016). Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research*, *44*(D1), D733–D745. https://doi.org/10.1093/nar/gkv1189

Okoniewski, M. J., & Miller, C. J. (2006). Hybridization interactions between probesets in short oligo microarrays lead to spurious correlations. *BMC Bioinformatics*, *7*, 276. https://doi.org/10.1186/1471-2105-7-276

Oliveira, S. M. D., Goncalves, N. S. M., Kandavalli, V. K., Martins, L., Neeli-Venkata, R., Reyelt, J., Fonseca, J. M., Lloyd-Price, J., Kranz, H., & Ribeiro, A. S. (2019). Chromosome and plasmid-borne PLacO3O1 promoters differ in sensitivity to critically low temperatures. *Scientific Reports*, *9*(1), 4486. https://doi.org/10.1038/s41598-019-39618-z

Oliveira, S. M. D., Häkkinen, A., Lloyd-Price, J., Tran, H., Kandavalli, V., & Ribeiro, A. S. (2016). Temperature-Dependent Model of Multi-step Transcription Initiation in Escherichia coli Based on Live Single-Cell Measurements. *PLOS Computational Biology*, *12*(10). https://doi.org/10.1371/journal.pcbi.1005174

Oppenheim, A. B., Kobiler, O., Stavans, J., Court, D. L., & Adhya, S. (2005). Switches in bacteriophage lambda development. *Annual Review of Genetics*, *39*, 409–429. https://doi.org/10.1146/annurev.genet.39.073003.113656

Osbourn, A. E., & Field, B. (2009). Operons. *Cellular and Molecular Life Sciences*, *66*(23), 3755–3775. https://doi.org/10.1007/s00018-009-0114-3

Otsu, N. (1979). Threshold selection method from gray-level histograms. *IEEE Transactions on Systems, Man, and Cybernetics*, *9*(1), 62–66. https://doi.org/10.1109/tsmc.1979.4310076

Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D., & Van Oudenaarden, A. (2002). Regulation of noise in the expression of a single gene. *Nature Genetics*, *31*(1), 69–73. https://doi.org/10.1038/ng869

Palade, G. E. (1955). A small particulate component of the cytoplasm. *The Journal of Biophysical and Biochemical Cytology*, *1*(1), 59-68. https://doi.org/10.1083/jcb.1.1.59

Palmer, A. C., Ahlgren-Berg, A., Egan, J. B., Dodd, I. B., & Shearwin, K. E. (2009).

Potent Transcriptional Interference by Pausing of RNA Polymerases over a Downstream Promoter. *Molecular Cell*, *34*(5), 545–555. https://doi.org/10.1016/j.molcel.2009.04.018

Pan, G., Li, J., Zhou, Y., Zheng, H., & Pei, D. (2006). A negative feedback loop of transcription factors that controls stem cell pluripotency and self-renewal. *The FASEB Journal*, *20*(10), 1730–1732. https://doi.org/10.1096/fj.05-5543fje

Paulsson, J. (2005). Models of stochastic gene expression. *Physics of Life Reviews*, *2*(2), 157–175. https://doi.org/10.1016/j.plrev.2005.03.003

Pawley, J. B. (2006). Handbook of biological confocal microscopy. Springer, Boston, MA. ISBN: 978-1-4757-5348-6

Pelletier, J., Halvorsen, K., Ha, B. Y., Paparcone, R., Sandler, S. J., Woldringh, C. L., Wong, W. P., & Jun, S. (2012). Physical manipulation of the Escherichia coli chromosome reveals its soft nature. *Proceedings of the National Academy of Sciences*, 109(40). https://doi.org/10.1073/pnas.1208689109

Pérez-Rueda, E., & Collado-Vides, J. (2000). The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12. *Nucleic Acids Research*, *28*(8), 1838–1847. https://doi.org/10.1093/nar/28.8.1838

Peter, B. J., Arsuaga, J., Breier, A. M., Khodursky, A. B., Brown, P. O., & Cozzarelli, N. R. (2004). Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. *Genome Biology*, *5*(11), R87. https://doi.org/10.1186/gb-2004-5-11-r87

Phadtare, S., & Inouye, M. (2004). Genome-Wide Transcriptional Analysis of the Cold Shock Response in Wild-Type and Cold-Sensitive, Quadruple-csp-Deletion Strains of Escherichia coli. *Journal of Bacteriology*, *186*(20), 7007–7014. https://doi.org/10.1128/jb.186.20.7007-7014.2004

Phadtare, S., & Inouye, M. (2008). The Cold Shock Response. *EcoSal Plus*, *3*(1). https://doi.org/10.1128/ecosalplus.5.4.2

Plank, M., Wadhams, G. H., & Leake, M. C. (2009). Millisecond timescale slimfield imaging and automated quantification of single fluorescent protein molecules for use in probing complex biological processes. *Integrative Biology*, 1(10), 602–612. https://doi.org/10.1039/b907837a

Postow, L., Crisona, N. J., Peter, B. J., Hardy, C. D., & Cozzarelli, N. R. (2001). Topological challenges to DNA replication: Conformations at the fork. *Proceedings of the National Academy of Sciences*, *98*(15), 8219–8226. https://doi.org/10.1073/pnas.111006998

Postow, L., Hardy, C. D., Arsuaga, J., & Cozzarelli, N. R. (2004). Topological domain structure of the Escherichia coli chromosome. *Genes & Development*, *18*(14), 1766–1779. https://doi.org/10.1101/gad.1207504

Potapov, I., Mäkelä, J., Yli-Harja, O., & Ribeiro, A. S. (2012). Effects of codon sequence on the dynamics of genetic networks. *Journal of Theoretical Biology*, *315*, 17–25. https://doi.org/10.1016/j.jtbi.2012.08.029

Prescott, E. M., & Proudfoot, N. J. (2002). Transcriptional collision between convergent genes in budding yeast. *Proceedings of the National Academy of*

Sciences, 99(13), 8796-8801. https://doi.org/10.1073/pnas.132270899

Pruss, G. J., & Drlica, K. (1986). Topoisomerase I mutants: the gene on pBR322 that encodes resistance to tetracycline affects plasmid DNA supercoiling. *Proceedings of the National Academy of Sciences*, *83*(23), 8952–8956. https://10.1073/pnas.83.23.8952

Pruss, G. J., & Drlica, K. (1989). DNA supercoiling and prokaryotic transcription. *Cell*, 56(4), 521-523. https://10.1016/0092-8674(89)90574-6

Rajala, T., Häkkinen, A., Healy, S., Yli-Harja, O., & Ribeiro, A. S. (2010). Effects of transcriptional pausing on gene expression dynamics. *PLOS Computational Biology*, *6*(3), 29–30. https://doi.org/10.1371/journal.pcbi.1000704

Ramakrishnan, V. (2002). Ribosome structure and the mechanism of translation. *Cell*, *108*(4), 557–572. https://doi.org/10.1016/S0092-8674(02)00619-0

Rangel, J. M., Sparling, P. H., Crowe, C., Griffin, P. M., & Swerdlow, D. L. (2005). Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002. *Emerging Infectious Diseases*, *11*(4), 603–609. https://doi.org/10.3201/eid1104.040739

Rao, C. V., Wolf, D. M., & Arkin, A. P. (2002). Control, exploitation and tolerance of intracellular noise. *Nature*, *420*(6912), 231–237. https://doi.org/10.1038/nature01258

Razo-Mejia, M., Barnes, S. L., Belliveau, N. M., Chure, G., Einav, T., Lewis, M., & Phillips, R. (2018). Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. *Cell Systems*, *6*(4), 456-469. https://doi.org/10.1016/j.cels.2018.02.004

Reece, R. J., Maxwell, A., & Wang, J. C. (2008). DNA Gyrase: Structure and Function. *Critical Reviews in Biochemistry and Molecular Biology*, *26*(3–4), 335–375. https://doi.org/10.3109/10409239109114072

Revyakin, A., Liu, C., Ebright, R. H., & Strick, T. R. (2006). Abortive initiation and productive initiation by RNA polymerase involve DNA scrunching. *Science*, *314*(5802), 1139–1143. https://doi.org/10.1126/science.1131398

Reyes-Lamothe, R., Sherratt, D. J., & Leake, M. C. (2010). Stoichiometry and Architecture of Active DNA Replication Machinery in Escherichia coli. *Science*, 328(5977), 498–501. https://doi.org/10.1126/science.1185757

Rhee, K. Y., Opel, M., Ito, E., Hung, S. p, Arfin, S. M., & Hatfield, G. W. (1999). Transcriptional coupling between the divergent promoters of a prototypic LysR-type regulatory system, the ilvYC operon of Escherichia coli. *Proceedings of the National Academy of Sciences*, *96*(25), 14294–14299. https://doi.org/10.1073/pnas.96.25.14294

Ribeiro, A. S., Smolander, O. P., Rajala, T., Häkkinen, A., & Yli-Harja, O. (2009). Delayed stochastic model of transcription at the single nucleotide level. *Journal of Computational Biology*, *16*(4), 539–553. https://doi.org/10.1089/cmb.2008.0153

Ribeiro, A. S., & Lloyd-Price, J. (2007). SGN Sim, a stochastic genetic networks simulator. *Bioinformatics*, 23(6), 777-779.

https://doi.org/10.1093/bioinformatics/btm004

Ribeiro, A., Zhu, R., & Kauffman, S. A. (2006). A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics. *Journal of Computational Biology*, *13*(9), 1630–1639. https://doi.org/10.1089/cmb.2006.13.1630

Ribeiro, A. S., Häkkinen, A., Healy, S., & Yli-Harja, O. (2010). Dynamical effects of transcriptional pause-prone sites. *Computational Biology and Chemistry*, *34*(3), 143–148. https://doi.org/10.1016/j.compbiolchem.2010.04.003

Ribeiro, A. S. (2010). Stochastic and delayed stochastic models of gene expression and regulation. *Mathematical Biosciences*, *223*(1), 1–11. https://doi.org/10.1016/j.mbs.2009.10.007

Riggs, A. D., Bourgeois, S., & Cohn, M. (1970). The lac repressor-operator interaction. 3. Kinetic studies. *Journal of Molecular Biology*, *53*(3), 401–417. https://doi.org/10.1016/0022-2836(70)90074-4

Rigney, D. R., & Schieve, W. C. (1977). Stochastic model of linear, continuous protein synthesis in bacterial populations. *Journal of Theoretical Biology*, *69*(4), 761–766. https://doi.org/10.1016/0022-5193(77)90381-2

Ringquist, S., Shinedling, S., Barrick, D., Green, L., Binkley, J., Stormo, G. D., & Gold, L. (1992). Translation initiation in Escherichia coli: sequences within the ribosome-binding site. *Molecular Microbiology*, *6*(9), 1219–1229. https://doi.org/10.1111/J.1365-2958.1992.tb01561.x

Roberts, E., Magis, A., Ortiz, J. O., Baumeister, W., & Luthey-Schulten, Z. (2011). Noise Contributions in an Inducible Genetic Switch: A Whole-Cell Simulation Study. *PLOS Computational Biology*, *7*(3), e1002010. https://doi.org/10.1371/journal.pcbi.1002010

Rogozina, A., Zaychikov, E., Buckle, M., Heumann, H., & Sclavi, B. (2009). DNA melting by RNA polymerase at the T7A1 promoter precedes the rate-limiting step at 37°C and results in the accumulation of an off-pathway intermediate. *Nucleic Acids Research*, *37*(16), 5390. https://doi.org/10.1093/nar/gkp560

Ross, W., Gosink, K. K., Salomon, J., Igarashi, K., Zou, C., Ishihama, A., Severinov, K., & Gourse, R. L. (1993). A Third Recognition Element in Bacterial Promoters: DNA Binding by the α Subunit of RNA Polymerase. *Science*, *262*(5138), 1407–1413. https://doi.org/10.1126/science.8248780

Roussel, M. R., & Zhu, R. (2006). Validation of an algorithm for delay stochastic simulation of transcription and translation in prokaryotic gene expression. *Physical Biology*, *3*(4), 274–284. https://doi.org/10.1088/1478-3975/3/4/005

Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z., & Higgins, N. P. (2012). Rates of Gyrase Supercoiling and Transcription Elongation Control Supercoil Density in a Bacterial Chromosome. *PLOS Genetics*, *8*(8), e1002845. https://doi.org/10.1371/journal.pgen.1002845

Royce, T. E., Rozowsky, J. S., & Gerstein, M. B. (2007). Toward a universal microarray: prediction of gene expression through nearest-neighbor probe sequence identification. *Nucleic Acids Research*, *35*(15), e99. https://doi.org/10.1093/nar/gkm549

Ruff, E. F., Drennan, A. C., Capp, M. W., Poulos, M. A., Artsimovitch, I., & Record, M. T. (2015). E. coli RNA Polymerase Determinants of Open Complex Lifetime and Structure. *Journal of Molecular Biology*, *427*(15), 2435–2450. https://doi.org/10.1016/j.jmb.2015.05.024

Rust, M. J., Bates, M., & Zhuang, X. (2006). Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). *Nature Methods*, *3*(10), 793–795. https://doi.org/10.1038/nmeth929

Ruusuvuori, P., Aijo, T., Chowdhury, S., Garmendia-Torres, C., Selinummi, J., Birbaumer, M., Dudley, A. M., Pelkmans, L., & Yli-Harja, O. (2010). Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics*, *11*(1), 248. https://doi.org/10.1186/1471-2105-11-248

Sacerdot, C., Fayat, G., Dessen, P., Springer, M., Plumbridge, J. A., Grunberg-Manago, M., & Blanquet, S. (1982). Sequence of a 1.26-kb DNA fragment containing the structural gene for E. coli initiation factor IF3: presence of an AUU initiator codon. *The EMBO Journal*, *1*(3), 311. https://doi.org/10.1002/j.1460-2075.1982.tb01166.x

Saecker, R. M., Record, M. T., & deHaseth, P. L. (2011). Mechanism of Bacterial Transcription Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent Open Complexes, and Initiation of RNA Synthesis. *Journal of Molecular Biology*, *412*(5), 754–771. https://doi.org/10.1016/j.jmb.2011.01.018

Saito, K., Green, R., & Buskirk, A. R. (2020). Translational initiation in E. Coli occurs at the correct sites genome-wide in the absence of mrna-rrna base-pairing. *eLife*, *9*. https://doi.org/10.7554/eLife.55002

Samul, R., & Leng, F. (2007). Transcription-coupled Hypernegative Supercoiling of Plasmid DNA by T7 RNA Polymerase in Escherichia coli Topoisomerase I-Deficient Strains. *Journal of Molecular Biology*, *374*(4), 925–935. https://doi.org/10.1016/j.jmb.2007.10.011

Sanchez, A., Osborne, M. L., Friedman, L. J., Kondev, J., & Gelles, J. (2011). Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *The EMBO Journal*, *30*(19), 3940–3946. https://doi.org/10.1038/emboj.2011.273

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*, *74*(12), 5463. https://doi.org/10.1073/pnas.74.12.5463

Santinha, J., Martins, L., Häkkinen, A., Lloyd-Price, J., Oliveira, S. M. D., Gupta, A., Annila, T., Mora, A., Ribeiro, A. S., & Fonseca, J. R. (2016). iCellFusion: Tool for Fusion and Analysis of Live-Cell Images from Time-Lapse Multimodal Microscopy. *In Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes,* 71–99. https://doi.org/10.4018/978-1-4666-8811-7.ch004

Santos-Zavaleta, A., Salgado, H., Gama-Castro, S., Sánchez-Pérez, M., Gómez-

Romero, L., Ledezma-Tejeida, D., García-Sotelo, J. S., Alquicira-Hernández, K., Muñiz-Rascado, L. J., Peña-Loredo, P., Ishida-Gutiérrez, C., Velázquez-Ramírez, D. A., Del Moral-Chávez, V., Bonavides-Martínez, C., Méndez-Cruz, C.-F., Galagan, J., & Collado-Vides, J. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in *E. coli* K-12. *Nucleic Acids Research*, *47*(D1), D212–D220. https://doi.org/10.1093/nar/gky1077

Schlax, P J, Capp, M. W., & Record, M. T. (1995). Inhibition of transcription initiation by lac repressor. *Journal of Molecular Biology*, *245*(4), 331–350. https://doi.org/10.1006/jmbi.1994.0028

Schmidt, T., Schütz, G. J., Baumgartner, W., Gruber, H. J., & Schindler, H. (1996). Imaging of single molecule diffusion. *Proceedings of the National Academy of Sciences*, *93*(7), 2926–2929. https://doi.org/10.1073/pnas.93.7.2926

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature Protocols*, *3*(6), 1101–1108. https://doi.org/10.1038/nprot.2008.73

Sevier, S. A., & Levine, H. (2017). Mechanical Properties of Transcription. *Physical Review Letters*, *118*, 268101. https://doi.org/10.1103/PhysRevLett.118.268101

Sevier, S. A., & Levine, H. (2018). Properties of gene expression and chromatin structure with mechanically regulated elongation. *Nucleic Acids Research*, *46*(12), 5924–5934. https://doi.org/10.1093/nar/gky382

Shaner, N. C., Campbell, R. E., Steinbach, P. A., Giepmans, B. N. G., Palmer, A. E., & Tsien, R. Y. (2004). Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology*, *22*(12), 1567–1572. https://doi.org/10.1038/nbt1037

Shaner, N. C., Steinbach, P. A., & Tsien, R. Y. (2005). A guide to choosing fluorescent proteins. *Nature Methods,* 2, 905–909 https://doi.org/10.1038/nmeth819

Shapiro, H. M. (2003). Practical Flow Cytometry. *John Wiley & Sons.* https://doi.org/10.1002/0471722731

Schärer, K., Stephan, R., & Tasara, T. (2013). Cold shock proteins contribute to the regulation of Listeriolysin O production in Listeria monocytogenes. *Foodborne pathogens and disease*, 10(12), 1023-1029. https://doi.org/10.1089/fpd.2013.1562

Shearwin, K. E., Callen, B. P., & Egan, J. B. (2005). Transcriptional interference – a crash course. *Trends in Genetics*, *21*(6), 339. https://doi.org/10.1016/j.tig.2005.04.009

Shehata, T. E., & Marr, A. G. (1971). Effect of nutrient concentration on the growth of Escherichia coli. Journal of Bacteriology, 107(1), 210–216. https://doi.org/10.1128/JB.107.1.210-216.1971

Sheridan, S. D., Opel, M. L., & Hatfield, G, W. (2001). Activation and repression of transcription initiation by a distant DNA structural transition. *Molecular Microbiology*, *40*(3), 684–690. https://doi.org/10.1046/j.1365-

2958.2001.02416.x

Sheridan, S. D., Benham, C. J., & Hatfield, G. W. (1998). Activation of gene expression by a novel DNA structural transmission mechanism that requires supercoiling-induced DNA duplex destabilization in an upstream activating sequence. *The Journal of Biological Chemistry*, *273*(33), 21298–21308. https://doi.org/10.1074/jbc.273.33.21298

Shimomura, O. (1979). Structure of the chromophore of *Aequorea* green fluorescent protein. *FEBS Letters*, *104*(2), 220–222. https://doi.org/10.1016/0014-5793(79)80818-2

Shimomura, O. (2005). The discovery of aequorin and green fluorescent protein. *Journal of Microscopy*, *217*(1), 3–15. https://doi.org/10.1111/j.0022-2720.2005.01441.x

Shimomura, O., Johnson, F. H., & Saiga, Y. (1962). Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *Journal of Cellular and Comparative Physiology*, *59*, 223–239. https://doi.org/10.1002/jcp.1030590302

Simão, E., Remy, E., Thieffry, D., & Chaouiya, C. (2005). Qualitative modelling of regulated metabolic pathways: application to the tryptophan biosynthesis in E.coli. *Bioinformatics*, *21, Suppl 2:ii190-6*. https://doi.org/10.1093/bioinformatics/bti1130

Sinensky, M. (1974). Homeoviscous Adaptation—A Homeostatic Process that Regulates the Viscosity of Membrane Lipids in Escherichia coli. *Proceedings of the National Academy of Sciences*, *71*(2), 522. https://doi.org/10.1073/pnas.71.2.522

Sleator, R. D., & Hill, C. (2002). Bacterial osmoadaptation: the role of osmolytes in bacterial stress and virulence. *FEMS Microbiology Reviews*, *26*(1), 49–71. https://doi.org/10.1111/J.1574-6976.2002.tb00598.x

Sneppen, K., Dodd, I. B., Shearwin, K. E., Palmer, A. C., Schubert, R. A., Callen, B. P., & Egan, J. B. (2005). A mathematical model for transcriptional interference by RNA polymerase traffic in Escherichia coli. *Journal of Molecular Biology*, *346*(2), 399–409. https://doi.org/10.1016/j.jmb.2004.11.075

So, L. H., Ghosh, A., Zong, C., Sepúlveda, L. A., Segev, R., & Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. *Nature Genetics*, *43*(6), 554–560. https://doi.org/10.1038/ng.821

Sokurenko, E. V., Chesnokova, V., Dykhuizen, D. E., Ofek, I., Wu, X. R., Krogfelt, K. A., Struve, G., Schembri, M. A., & Hasty, D. L. (1998). Pathogenic adaptation of Escherichia coli by natural variation of the FimH adhesin. *Proceedings of the National Academy of Sciences*, *95*(15), 8922. https://doi.org/10.1073/pnas.95.15.8922

Sokurenko, E. V., Courtney, H. S., Maslow, J., Siitonen, A., & Hasty, D. L. (1995). Quantitative differences in adhesiveness of type 1 fimbriated Escherichia coli due to structural differences in fimH genes. *Journal of Bacteriology*, *177*(13), 3680–3686. https://doi.org/10.1128/jb.177.13.3680-3686.1995

Song, E., Uhm, H., Munasingha, P. R., Hwang, S., Seo, Y.-S., Kang, J. Y., Kang, C., & Hohng, S. (2022). Rho-dependent transcription termination proceeds via three routes. *Nature Communications*, 13(1), 1–12. https://doi.org/10.1038/s41467-022-29321-5

Stark, R., Grzelak, M., & Hadfield, J. (2019). RNA sequencing: the teenage years. *Nature Reviews Genetics*, *20*(11), 631–656. https://doi.org/10.1038/s41576-019-0150-2

Startceva, S., Kandavalli, V. K., Visa, A., & Ribeiro, A. S. (2019). Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, *1862*(2), 119–128. https://doi.org/10.1016/j.bbagrm.2018.12.005

Stein, R. A., Deng, S., & Higgins, N. P. (2005). Measuring chromosome dynamics on different time scales using resolvases with varying half-lives. *Molecular Microbiology*, *56*(4), 1049. https://doi.org/10.1111/j.1365-2958.2005.04588.x

Stephens, D. J., & Allan, V. J. (2003). Light microscopy techniques for live cell imaging. *Science*, *300*(5616), 82–86. https://doi.org/10.1126/science.1082160

Stevens, A. (1960). Incorporation of the adenine ribonucleotide into RNA by cell fractions from E. coli B. *Biochemical and Biophysical Research Communications*, 3(1), 92–96. https://doi.org/10.1016/0006-291x(60)90110-8

Stokes, G. G. (1852). On the Change of Refrangibility of Light. *Philosophical Transactions of the Royal Society of London*, *142*, 463-562.

Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., & Kapanidis, A. N. (2015). Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proceedings of the National Academy of Sciences*, *112*(32), E4390–E4399. https://doi.org/10.1073/pnas.1507592112

Stracy, M., Wollman, A. J. M., Kaja, E., Gapinski, J., Lee, J.-E., Leek, V. A., McKie, S. J., Mitchenall, L. A., Maxwell, A., Sherratt, D. J., Leake, M. C., & Zawadzki, P. (2019). Single-molecule imaging of DNA gyrase activity in living Escherichia coli. *Nucleic Acids Research*, 47(1), 210–220. https://doi.org/10.1093/nar/gky1143

Swain, P. S., Elowitz, M. B., & Siggia, E. D. (2002). Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences*, *99*(20), 12795–12800. https://doi.org/10.1073/pnas.162041399

Tabib-Salazar, A., Mulvenna, N., Severinov, K., Matthews, S. J., & Wigneshweraraj, S. (2019). Xenogeneic regulation of the bacterial transcription machinery. *Journal of molecular biology*, 431(20), 4078-4092. https://doi.org/10.1016/j.jmb.2019.02.008

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A., & Xie, X. S. (2010). Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Science*, *329*(5991), 533–538. https://doi.org/10.1126/science.1188308

Thattai, M., & Van Oudenaarden, A. (2004). Stochastic gene expression in

fluctuating environments. *Genetics*, *167*(1), 523–530. https://doi.org/10.1534/genetics.167.1.523

Thomas, E., & Williams. *The History of Miner Lamps*. Welshminerslamps. Retrieved March 7, 2022, from https://www.welshminerslamps.com/lamp-history/

Tilden, J., Young, W., McNamara, A. M., Custer, C., Boesel, B., Lambert-Fair, M. A., Majkowski, J., Vugia, D., Werner, S. B., Hollingsworth, J., & Morris, J. G. (1996). A new route of transmission for Escherichia coli: infection from dry fermented salami. *American Journal of Public Health*, *86*(8), 1142–1145. https://doi.org/10.2105/ajph.86.8_pt_1.1142

Tokunaga, M., Imamoto, N., & Sakata-Sogawa, K. (2008). Highly inclined thin illumination enables clear single-molecule imaging in cells. *Nature Methods*, *5*(2), 159–161. https://doi.org/10.1038/nmeth1171

Tran, H., Oliveira, S. M. D., Goncalves, N., & Ribeiro, A. S. (2015). Kinetics of the cellular intake of a gene expression inducer at high concentrations. *Molecular BioSystems*, *11*(9), 2579–2587. https://doi.org/10.1039/c5mb00244c

Travers, A., & Muskhelishvili, G. (2005). DNA supercoiling — a global transcriptional regulator for enterobacterial growth? *Nature Reviews Microbiology*, *3*(2), 157–169. https://doi.org/10.1038/nrmicro1088

Trinklein, N. D., Aldred, S. F., Hartman, S. J., Schroeder, D. I., Otillar, R. P., & Myers, R. M. (2004). An Abundance of Bidirectional Promoters in the Human Genome. *Genome Research*, *14*(1), 62–66. https://doi.org/10.1101/gr.1982804

Tripathi, S., Brahmachari, S., Onuchic, J. N., & Levine, H. (2022). DNA supercoiling-mediated collective behavior of co-transcribing RNA polymerases. *Nucleic Acids Research*, *50*(3), 1269–1279. https://doi.org/10.1093/nar/gkab1252

Tsao, Y. P., Wu, H. Y., & Liu, L. F. (1989). Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies. *Cell*, *56*(1), 111–118. https://doi.org/10.1016/0092-8674(89)90989-6

Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., Pan, T., Dahan, O., Furman, I., & Pilpel, Y. (2010). An evolutionarily conserved mechanism for controlling the efficiency of protein translation. *Cell*, *141*(2), 344–354. https://doi.org/10.1016/j.cell.2010.03.031

Uphoff, S., Reyes-Lamothe, R., De Leon, F. G., Sherratt, D. J., & Kapanidis, A. N. (2013). Single-molecule DNA repair in live bacteria. *Proceedings of the National Academy of Sciences*, 110(20), 8063–8068. https://doi.org/10.1073/pnas.1301804110

Valeur, B., & Berberan-Santos, M. N. (2012). Molecular Fluorescence: Principles and Applicationse (2nd ed.). Wiley-VCH. https://doi.org/10.1002/9783527650002

Velculescu, V. E., Zhang, L., Vogelstein, B., & Kinzler, K. W. (1995). Serial Analysis of Gene Expression. *Science*, *270*(5235), 484–487. https://doi.org/10.1126/science.270.5235.484

Viard, T., & de la Tour, C. B. (2007). Type IA topoisomerases: A simple puzzle? *Biochimie*, *89*(4), 456–467. https://doi.org/10.1016/j.biochi.2006.10.013

Vignolini, T., Curcio, V., Gardini, L., Capitanio, M., & Pavone, F. S. (2018). Characterization of highly inclined optical sheet microscopy for localization microscopy. *IET Conference Publications*, 1-4, https://doi.org/10.1049/cp.2018.1662

Vinograd, J., Lebowitz, J., Radloff, R., Watson, R., & Laipis, P. (1965). The twisted circular form of polyoma viral DNA. *Proceedings of the National Academy of Sciences*, 53(5), 1104. https://doi.org/10.1073/pnas.53.5.1104

Vos, S. M., Tretter, E. M., Schmidt, B. H., & Berger, J. M. (2011). All tangled up: how cells direct, manage and exploit topoisomerase function. *Nature Reviews Molecular Cell Biology*, *12*, 827–841. https://doi.org/10.1038/nrm3228

Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., Kingsmore, S. F., Schroth, G. P., & Burge, C. B. (2008). Alternative isoform regulation in human tissue transcriptomes. *Nature*, *456*(7221), 470–476. https://doi.org/10.1038/nature07509

Wang, F., & Greene, E. C. (2011). Single-molecule studies of transcription: From one RNA polymerase at a time to the gene expression profile of a cell. *Journal of Molecular Biology*, 412(5), 814–831. https://doi.org/10.1016/j.jmb.2011.01.024

Wang, J. C. (1971). Interaction between DNA and an Escherichia coli protein omega. *Journal of Molecular Biology*, *55*(3). https://doi.org/10.1016/0022-2836(71)90334-2

Wang, P., Robert, L., Pelletier, J., Dang, W. L., Taddei, F., Wright, A., & Jun, S. (2010). Robust Growth of Escherichia coli. *Current Biology*, 20(12), 1099–1103. https://doi.org/10.1016/j.cub.2010.04.045

Wang, X., Llopis, P. M., & Rudner, D. Z. (2013). Organization and segregation of bacterial chromosomes. *Nature Reviews Genetics*, *14*(3), 191–203. https://doi.org/10.1038/nrg3375

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, *10*(1), 57–63. https://doi.org/10.1038/nrg2484

Ward, D. F., & Murray, N. E. (1979). Convergent transcription in bacteriophage λ: interference with gene expression. *Journal of molecular biology*, 133(2), 249-266. https://doi.org/10.1016/0022-2836(79)90533-3

Watson, J., & Crick, F. (1953). Molecular Structure of Nucleic Acids. *Nature*, 171(4356), 737–738.

Weinstein-Fischer, D., Elgrably-Weiss, M., & Altuvia, S. (2000). Escherichia coli response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. *Molecular Microbiology*, 35(6), 1413–1420. https://doi.org/10.1046/J.1365-2958.2000.01805.x

Weng, X., Bohrer, C. H., Bettridge, K., Lagda, C. A., Cagliero, C., Jin, J. D., & Xiao, J. (2019). Spatial organization of RNA polymerase and its relationship

with transcription in Escherichia coli. *Proceedings of the National Academy of Sciences,* 116(40), 20115-20123. https://doi.org/10.1073/pnas.1903968116

Whitmarsh, A. J., & Davis, R. J. (2000). Regulation of transcription factor function by phosphorylation. *Cellular and Molecular Life Sciences* , 57(8–9), 1172–1183. https://doi.org/10.1007/PL00000757

Wichmann, J., & Hell, S. W. (1994). Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics Letters,* 19(11), 780-782. https://doi.org/10.1364/ol.19.000780

Williams, N. L., & Maxwell, A. (1999). Probing the two-gate mechanism of DNA gyrase using cysteine cross-linking. *Biochemistry*, 38(41), 13502–13511. https://doi.org/10.1021/bi9912488

Wilson, K. S., & von Hippel, P. H. (1994). Stability of Escherichia coli transcription complexes near an intrinsic terminator. *Journal of Molecular Biology*, *244*(1), 36–51. https://doi.org/10.1006/jmbi.1994.1702

Wu, B., Piatkevich, K. D., Lionnet, T., Singer, R. H., & Verkhusha, V. V. (2011). Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. *Current Opinion in Cell Biology*, *23*(3), 310–317. https://doi.org/10.1016/j.ceb.2010.12.004

Wu, H. Y., Shyy, S. H., Wang, J. C., & Liu, L. F. (1988). Transcription generates positively and negatively supercoiled domains in the template. *Cell*, *53*(3), 433–440. https://doi.org/10.1016/0092-8674(88)90163-8

Yamanaka, K. (1999). Cold shock response in Escherichia coli. *Journal of Molecular Microbiology and Biotechnology*, *1*(2), 193–202.

Yarchuk, O., Jacques, N., Guillerez, J., & Dreyfus, M. (1992). Interdependence of translation, transcription and mRNA degradation in the lacZ gene. *Journal of Molecular Biology*, *226*(3), 581–596. https://doi.org/10.1016/0022-2836(92)90617-s

Yeung, E., Dy, A. J., Martin, K. B., Ng, A. H., Del Vecchio, D., Beck, J. L., Collins, J. J., & Murray, R. M. (2017). Biophysical Constraints Arising from Compositional Context in Synthetic Gene Networks. *Cell Systems*, *5*(1), 11-24.e12. https://doi.org/10.1016/j.cels.2017.06.001

Young, B. A., Gruber, T. M., & Gross, C. A. (2002). Views of Transcription Initiation. *Cell*, *109*(4), 417–420. https://doi.org/10.1016/S0092-8674(02)00752-3

Yu, J., Xiao, J., Ren, X., & Xie, X. S. (2006). Probing Gene Expression in Live Single Escherichia coli Cells – One Molecule at a Time. *Science*, *311*(5767), 1600–1603. https://doi.org/10.1126/science.1119623

Zaychikov, E., Denissova, L., & Heumann, H. (1995). Translocation of the Escherichia coli transcription complex observed in the registers 11 to 20: "jumping" of RNA polymerase and asymmetric expansion and contraction of the "transcription bubble." *Proceedings of the National Academy of Sciences*, *92*(5), 1739–1743. https://doi.org/10.1073/pnas.92.5.1739

Zechiedrich, E. L., Khodursky, A. B., Bachellier, S., Schneider, R., Chen, D., Lilley,

D. M. J., & Cozzarelli, N. R. (2000). Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli. *Journal of Biological Chemistry*, *275*(11), 8103–8113. https://doi.org/10.1074/jbc.275.11.8103

Zernike, F. (1942). Phase contrast, a new method for the microscopic observation of transparent objects. *Physica*, *9*(7), 686–698. https://doi.org/10.1016/s0031-8914(42)80035-x

Zernike, F. (1955). How I discovered phase contrast. *Science*, *121*(3141), 345–349. https://doi.org/10.1126/science.121.3141.345

Zhang, J., Campbell, R. E., Ting, A. Y., & Tsien, R. Y. (2002). Creating new fluorescent probes for cell biology. *Nature Reviews Molecular Cell Biology*, *3*(12), 906–918. https://doi.org/10.1038/nrm976

Zhang, T., Shi, X. C., Xia, Y., Mai, L., & Tremblay, P. L. (2019). Escherichia coli adaptation and response to exposure to heavy atmospheric pollution. *Scientific Reports, 9*(1), 1–13. https://doi.org/10.1038/s41598-019-47427-7

Zhao, S., Fung-Leung, W. P., Bittner, A., Ngo, K., & Liu, X. (2014). Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLOS ONE*, *9*(1), e78644. https://doi.org/10.1371/journal.pone.0078644

Zhu, R., Ribeiro, A. S., Salahub, D., & Kauffman, S. A. (2007). Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models. *Journal of Theoretical Biology*, *246*(4), 725–745. https://doi.org/10.1016/j.jtbi.2007.01.021

Zimmer, C. (2008). Microcosm: E-coli and the New Science of Life. *Pantheon Books*. ISBN: 037542430x

PUBLICATIONS

PUBLICATION I

A strategy for dissecting the kinetics of transcription repression mechanisms

C.S.D. Palma, S. Startceva, R. Neeli-Venkata, M. Zare, N.S.M. Goncalves, J.M. Fonseca, S.M.D. Oliveira, and A.S. Ribeiro

EMBEC & NBC. IFMBE Proceedings, 65. Springer, Singapore, 2017
DOI: 10.1007/978-981-10-5122-7_274

A strategy for dissecting the kinetics of transcription repression mechanisms

Cristina S.D. Palma[1,2], Sofia Startceva[1], Ramakanth Neeli-Venkata[1], Marzieh Zare[1], Nadia S.M. Goncalves[1], Jose M. Fonseca[2], Samuel M.D. Oliveira[1], and Andre S. Ribeiro[1,2]

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

[2] Computational Intelligence Group of CTS/UNINOVA. Faculdade de Ciências e Tecnologia da Univ. Nova de Lisboa, Portugal

Abstract— **Promoters in *Escherichia coli* include an 'OFF' state, during which transcription is halted. Here, we propose a novel empirical method for assessing the time-length spent by promoters in this state. It relies on direct measurements of RNA production kinetics at the single molecule level at different induction levels, followed by an estimation of the RNA production rate under infinite induction, which is then compared to this rate under real, maximum induction. We apply it to the LacO3O1 promoter and infer that, under full induction, on average, 15% of the time between successful transcription events is spent in the OFF state. We verify this result by comparing the kinetics of a mutant strain lacking repressor molecules with that of the inferred rate under infinite induction. We expect this strategy of dissecting the kinetics of transcription repression to be applicable to a wide number of promoters in *E. coli*.**

Keywords— **Transcription, Induction, τ Plot, OFF state.**

I. Introduction

Novel experimental techniques of microscopy and fluorescent molecular probes have led to the rapid acquisition of invaluable data on the dynamics of gene expression in live cells. One particularly valuable development has been the engineering of the MS2-GFP protein that has the ability to bind specific RNA sequences and, thus, provided multiple binding sites, detecting individual RNA molecules as soon as they are produced in live cells [1]. This technique allows both estimating RNA numbers in individual cells of a population at a given time [1] as well as obtaining RNA production time intervals [2]. This data greatly increased our knowledge on the *in vivo* dynamics of transcription.

Recently, a technique was developed for dissecting the dynamics of active transcription [3]. This method uses measurements of the RNA production dynamics from cells with differing RNA polymerase (RNAP) concentrations, which allows estimating what would be the rate of transcription in cells with infinite RNAP concentration.

Here, we propose a novel, similar methodology that uses data from cells with differing intracellular inducer concentrations, to further dissect the kinetics of transcription initiation. In particular, we focus on the promoter OFF state.

II. Methods

A. Cells, Plasmids, and Chemicals

We use *E. coli* strain BW25113 ($lacI^+$ $rrnB_{T14}$ $\Delta lacZ_{WJ16}$ *hsdR514* $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$) [4], which have the constitutive promoters P_{lacI+} and P_{araC} producing, respectively, LacI repressors for the LacO3O1 promoter [5] and AraC repressors for the BAD promoter. We also use the deletion mutant strain JW0336 (BW25113 *ΔlacI*), lacking the ability to express LacI repressor molecules.

In both strains, we placed a single-copy plasmid pBELOBAC11 carrying a $P_{lacO3O1}$ promoter [6] controlling the production of an RNA coding for 48 binding sites for MS2-GFP proteins (48BS). We also introduced a medium-copy plasmid, pZA25, with the reporter gene, P_{BAD}-MS2-GFP, responsible for producing the fusion protein MS2d-GFP, generously provided by Orna Amster-Choder (Hebrew University of Jerusalem, Israel) [7]. The activity of $P_{lacO3O1}$ is regulated by the repressor LacI and the inducer Isopropyl β-D-1-thiogalactopyranoside (IPTG). The activity of P_{BAD} is regulated by AraC and the inducer L-arabinose.

This system has been used to measure the distribution of time intervals between RNA production events due to its ability to detect individual RNAs, as the MS2-GFP proteins rapidly bind to newly formed RNAs, which can be seen as fluorescent foci under a fluorescence microscope [1-3].

B. Growth Conditions

Cells were grown overnight in LB medium supplemented with appropriate antibiotics (34 μg/ml of Chloramphenicol and 50 μg/ml of Kanamycin) with shaking at 250 rpm. We subsequently made subcultures by diluting the stationary-phase culture into fresh M9 medium, supplemented with Glycerol (0.4% final concentration) along with the appropriate antibiotics. Cells were placed in the incubator until reaching OD600 of ~0.25. For the reporter plasmid activation, we added 0.4% of L-arabinose to the culture, which was incubated at 37°C for 60 minutes. Next, for the target plasmid, specific concentrations of IPTG (0, 5, 25, 50, 100, 250, 500, and 1000 μM) were added. Cells were then incubated for 120

minutes. In the end, cells were collected by centrifugation at 8000 × g for 1 minute, and diluted in fresh M9 medium. From this, 5 µL of cells were added to an M9 glycerol agarose gel pad, prior to microscopy observation.

C. Microscopy

Cells were imaged by a 488 nm argon laser (Melles-Griot), and an emission filter (HQ514/30, Nikon), using a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with a 100× Apo TIRF (1.49 NA, oil) objective. To obtain single-time-point images, we used a C2+ (Nikon) confocal laser-scanning system. Meanwhile, in time-lapse measurements, cells were imaged by Highly Inclined and Laminated Optical sheet (HILO) microscopy, using an EMCCD camera (iXon3 897, Andor Technology). In both cases, phase-contrast images were acquired by a CCD camera (DS-Fi2, Nikon), for purposes of cell segmentation. The software for image acquisition was NIS-Elements (Nikon, Japan).

In time-lapse microscopy, cells were constantly supplied with fresh media with IPTG/L-arabinose during image acquisition, at the same concentration as in liquid culture, by a micro-perfusion peristaltic pump (Bioptechs) at 0.3 ml/min. Images were captured for 2 hours, 1 per minute in the case of fluorescence and 1 per 5 minutes in the case of phase-contrast. Also, cells were kept in a temperature-controlled chamber (FCS2, Bioptechs) at optimal temperature (37 °C).

D. Image Analysis

Microscopy images were processed as in [3]. First, cells were detected from phase contrast images and then aligned with the confocal images. Fluorescent spots and their intensities were detected from the confocal images as in [8].

For population analysis, the intensity of one spot was calculated from the histogram of total intensity of fluorescent spots per cell, normalized by the intensity of the first spike on this histogram, as in [1].

For time series analysis, jumps in each cell's spot intensity over time were detected using a least-deviation jump-detection method [8]. Given the noise in the time series, the intensity of 'one RNA' was first selected automatically, and then corrected by manual inspection of the total foreground spot intensities, as in [3]. To avoid spurious jumps, we disregarded jumps within 5 min of the beginning or end of a cell's lifetime [3]. Censored intervals were calculated as the time from the last RNA production in a cell until the last time point when a jump could have been observed [8].

E. Model of transcription

RNA production starts with the freeing of the promoter for transcription, followed by the closed complex formation, which includes multiple binding events of RNAPs until the promoter commits to the next step, the open complex formation, and finally to the promoter escape and RNA production. We represent this in reactions (1) and (2). In these, the various k's represent the rate constants at which these events take place:

$$P_{OFF} \underset{k_{off}*[\mathrm{Rep}]}{\overset{k_{on}}{\rightleftarrows}} P_{ON} \quad (1)$$

$$P_{ON} \underset{k_{-1}}{\overset{k_1*[\mathrm{RNAP}]}{\rightleftarrows}} P_{cc} \xrightarrow{k_2} P_{oc}{}^{i} \xrightarrow{k_3} P_{oc}{}^{f} \xrightarrow{k_4} P_{ON} + RNA \quad (2)$$

In (1), it is represented the reversibility between ON and OFF states of the promoter P (represented by P_{ON} and P_{OFF}, respectively), due to, e.g., unbinding/binding of repressors (respectively). Another cause for OFF states could be the accumulation/release of local positive DNA super-coiling in the chromosomal integrated gene, generated by transcription events and the release by gyrases, respectively [9,10].

In (2), a promoter in the ON state proceeds to form a closed complex (P_{cc}), as an RNAP finds the start site. As such, this event depends both on the concentration of freely available RNAPs and on the rate with which one RNAP binds to the start site (k_1) At this stage, there is a significant chance, explicitly represented, that the promoter reverts to the previous stage. After several attempts, the promoter reaches, first, the initial stage of open complex formation ($P_{oc}{}^{i}$). From here onwards the process is nearly irreversible and, in the presence of Mg^{2+}, a fully formed open complex ($P_{oc}{}^{f}$) is created [11]. Subsequently, the RNAP escapes the promoter and proceeds with elongation, which leads to the production of a fully formed RNA. Finally, note that elongation is not considered explicitly, since it is a fast process relative to the previous events and since it does not affect the mean duration of intervals between consecutive RNA productions.

F. τ Plots

The mean time-length between transcription events can be altered by changing the free RNAP concentration, as demonstrated both *in vitro* and *in vivo* [3, 12-14]. Only the duration of the closed complex formation is affected by these changes in free RNAP concentration [3]. As such, from a set of measurements of mean interval durations between consecutive RNA production events in individual cells whose RNAP concentrations differ, it is possible to estimate the rate of RNA production for an infinite RNAP concentration. This rate should correspond to the inverse of the duration of the open complex formation [3].

Here, we use a similar strategy. However, instead of altering intracellular RNAP concentrations, we alter inducer concentrations. As shown below, this produces data that allows estimating the time spent in both the closed and open complex formation. Further, assuming the model of transcription in (1) and (2), confronting this data with the intervals between

consecutive RNA productions informs on the mean time that promoters spend in the OFF state.

III. Results

We first derive, assuming the model described by (1) and (2), the equations supporting our method of dissection of the transcription repression kinetics. From (1) and (2), the mean duration of the intervals between consecutive RNAs, Δt, is:

$$\Delta t = \tau_{OFF} + \tau_{cc} + \tau_{oc} \quad (3)$$

In (3), τ_{OFF} is the mean time spent in P_{OFF} state between two RNA production events (note that this mean time can, and likely does, result from multiple 'passages' through the OFF state between two consecutive RNA production events). Meanwhile, τ_{cc} is the mean time spent in P_{ON} state until forming a successful closed complex (again, the promoter will likely be several times in this state between two consecutive RNA production events). Finally, τ_{oc} is the mean time for a closed complex to successfully form an open complex. As such, it includes the time to change from P_{cc} to P_{oc}^{i} (open complex in initial state) along with the time to change from P_{oc}^{i} to P_{oc}^{f} (open complex in fully formed state). Since the open complex formation is physically nearly irreversible once initiated, usually it only occurs once between two consecutive RNA production events.

Of these, only τ_{OFF} is expected to differ with the inducer concentration, *provided that, e.g. the inducer acts by inactivating the repressor,* as in the case of IPTG [5]. Given this, we define Δt_{ind} as the mean time between transcription events aside from the OFF period, and rewrite (3) as:

$$\Delta t = \tau_{OFF} + \Delta t_{ind} \quad (4)$$

From the above definitions, changing IPTG levels will alter only τ_{OFF} (and thus Δt). I.e. assuming a new condition, differing in [IPTG] (within the range where changes in [IPTG] result in changes in the RNA production rate):

$$\Delta t^{new} = \tau_{OFF}^{new} + \Delta t_{ind} \quad (5)$$

Next, one can plot the mean Δt versus the inverse of the [IPTG], for various IPTG concentrations. On this data, one can do a linear fit to estimate Δt^{inf} (i.e. the expected Δt assuming infinite [IPTG]), as it is given by the height at which the line of the fit crosses the y-axis (see Fig. 2).

Infinite [IPTG] implies that all the repressor molecules in the cell should be inactive. As such, the validity of the inference can be tested, e.g., by measuring the transcription kinetics in cells lacking the ability to produce the repressor.

To implement this strategy, we must first find the induction levels that significantly differ in transcription rates. We thus measured RNA numbers in individual cells at various induction levels of LacO3O1, both to determine its maximum induction level, as well as to find the region of the induction curve where the transcription rate is sensitive to small changes in inducer concentration.

From the results in Figure 1, maximum induction occurs from 50 μM IPTG and beyond, while the region of the induction curve where the RNA production rate is most sensitive to changes in [IPTG] is between 0 and 50 μM.

Thus, we performed microscopy time-series of cells subject, respectively, to 5, 25 and 50 μM IPTG (Methods). From the images, we obtained the mean Δt for each condition (Methods). We also imaged deletion mutants lacking the repression mechanism (i.e. unable to express lacI repressor molecules). Results are shown in Table 1.

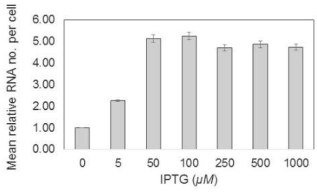

Fig. 1 Mean relative RNA numbers in individual cells. Images taken 2 hours after activating the target gene. The error bars are the standard error of the mean. In all conditions, more than 350 cells were observed.

Table 1 Empirical mean and uncertainty of the intervals between transcription events in individual cells for various induction levels

Condition	No. of cells	No. of intervals	Mean inferred interval and uncertainty (s)
5 μM	360	156	4362 ± 647
25 μM	92	30	2024 ± 709
50 μM	72	54	1922 ± 588
Deletion Mutant	44	23	1805 ± 757

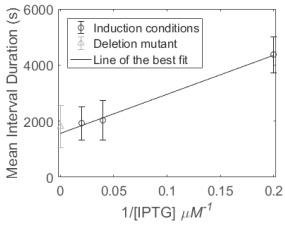

Fig. 2 τ -plot for LacO3O1, showing Δt of each induction level (circles) and standard uncertainties, along with best-fit line by least-squares fit. Also shown is the data from a mutant strain (triangle, not used in the fit).

From Table 1, following the strategy in [3], we obtained a τ-plot (Fig. 2), where the X-axis is the inverse of [IPTG], while the Y-axis is the mean duration of transcription intervals in individual cells. Next, we fitted a line to the data by the least squares method and extrapolated this linear fit to the

'infinite [IPTG]' condition. The intercept of the best fitted line with the Y axis is Δt^{inf}.

From Fig. 2, the best fit line intercepts the y-axis at *1559 s*. Since, under full induction (50 μM IPTG), the mean interval between transcription events is 1839 s *(according to the best fit line in Fig. 2)*, we expect the promoter to spend ~280 s in the OFF state between transcription events (15% of the time), under full induction.

To validate our result, we used a mutant strain lacking the ability to produce the repressor of the promoter of interest. The data from these cells is also in the τ plot and in Table 1, and, visibly, their RNA production rate is in agreement with the estimated rate for 'infinite' induction, from which we find the estimation to be reliable.

IV. Conclusions

We proposed a novel methodology for the estimation of the mean time spent by promoters in the OFF state between transcription events. To our knowledge, this is the first time this is achieved from data from live, individual cells.

At the moment, the method has several requirements that depend on the repression mechanism. E.g., it requires prior knowledge or measurements of how the inverse of the RNA production rate changes with the inverse of the inducer concentration. Here, this change was assumed to be linear within a certain range (which the data did not disprove), which allowed estimating the kinetics of RNA production under infinite induction levels. Also, it requires an inducer or activator that acts by freeing the promoter from an OFF state, rather than by affecting the kinetics of subsequent steps in transcription, in which case the outcome of the method will inform on different parameters than those in the present work.

We applied the methodology to LacO3O1 and its repression system. For validation, we measured the transcription kinetics in a deletion mutant for the repressor of LacO3O1, and found it to be in agreement with our method's estimation.

We believe that, in its present state, this method can already be of use to dissect the kinetics of repression mechanisms of transcription of various promoters of *E. coli*. We expect that several improvements can be further made to the present method, to allow its application to a broader range of repression (as well as induction) mechanisms.

Acknowledgment

Work supported by Academy of Finland (295027 and 305342 ASR), Jane & Aatos Erkko Foundation (610536 ASR), Tampere University of Technology President's Graduate Program (S.S. and R.N.-V.), Finnish Academy of Science and Letters (SO), Doctoral Programme of Computing and Electrical Engineering of TUT (NG) and Erasmus+ program 2919(713)2915/2016/SMS (C.P.). The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

Conflict of Interest

The authors declare that they have no conflict of interest.

References

1. Golding I, Paulsson J, Zawilski S, Cox EC (2005) Real-time kinetics of gene activity in individual bacteria. Cell 123:1025–36.
2. Muthukrishnan A-B et al (2014) *In vivo* Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in Stressed *Escherichia coli* cells. PLoS ONE 9(9): e109005
3. Lloyd-Price J et al. (2016) Dissecting the stochastic transcription initiation process in live *Escherichia coli*. DNA Res. 23:203–14.
4. Baba T et al. (2006) Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol. Syst. Biol.2:1-11.
5. Glascock CB and Weickert MJ (1998) Using chromosomal lacI Q1 to control expression of genes on high-copy-number plasmids in *Escherichia coli*. Gene 223, 221–23.
6. NSM Goncalves et al. (2016) In vivo single-molecule dynamics of transcription of the viral T7 Phi 10 promoter in *E. coli*. In Proc. of the BIOTECHNO 2016, June 26-30, 2016 Lisbon, Portugal.
7. Nevo-Dinur K et al. (2011) Translation-independent localization of mRNA in *E. coli*. Science 331(6020):1081-4.
8. Häkkinen A and Ribeiro AS (2015) Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. Bioinformatics 31 (1): 69-75.
9. Chong S, Chen C, Ge H, and Xie XS (2014) Mechanism of Transcriptional Bursting in Bacteria. Cell 158, 314–326.
10. Liu L and Wang J (1987) Supercoiling of the DNA template during transcription. Proc. Natl. Acad. Sci. U. S. A. 84, 7024–7.
11. DeHaseth PL, Zupancic ML and Record MT (1998) RNA polymerase-promoter interactions: The comings and goings of RNA polymerase. J. Bacteriol. 180, 3019–25.
12. Liang S et al. (1999) Activities of constitutive promoters in *Escherichia coli*. J. Mol. Biol. 292:19–37.
13. McClure WR (1985) Mechanism and control of transcription initiation in prokaryotes. Annu. Rev. Biochem. 54:171–204
14. Ehrenberg M et al. (2013) Medium-dependent control of the bacterial growth rate. Biochimie. Elsevier Masson 95:643–58.

Corresponding Author: Andre S. Ribeiro
Institute: Tampere University of Technology
Street: Korkeakoulunkatu 10, 33720 Tampere
City: Tampere
Country: Finland
Email: andre.ribeiro@tut.fi

PUBLICATION II

Dissecting the *in vivo* dynamics of transcription locking due to positive supercoiling buildup

C.S.D. Palma, V. Kandavalli, M.N.M. Bahrudeen, M. Minoia, V. Chauhan, S. Dash, and A.S. Ribeiro

Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms, 1863(5), 194515, 2020
DOI: 10.1016/j.bbagrm.2020.194515

BBA - Gene Regulatory Mechanisms 1863 (2020) 194515

Contents lists available at ScienceDirect

BBA - Gene Regulatory Mechanisms

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

Dissecting the *in vivo* dynamics of transcription locking due to positive supercoiling buildup

Cristina S.D. Palma, Vinodh Kandavalli, Mohamed N.M. Bahrudeen, Marco Minoia, Vatsala Chauhan, Suchintak Dash, Andre S. Ribeiro*

Laboratory of Biosystem Dynamics, BioMediTech, Faculty of Medicine and Health Technology, Tampere University, 33101 Tampere, Finland

ARTICLE INFO

Keywords:
Single-RNA production dynamics
Positive supercoiling buildup
LineWeaver-Burk plots
Transcription locking kinetics

ABSTRACT

Positive supercoiling buildup (PSB) is a pervasive phenomenon in the transcriptional programs of *Escherichia coli*. After finding a range of Gyrase concentrations where the inverse of the transcription rate of a chromosome-integrated gene changes linearly with the inverse of Gyrase concentration, we apply a LineWeaver-Burk plot to dissect the expected *in vivo* transcription rate in absence of PSB. We validate the estimation by time-lapse microscopy of single-RNA production kinetics of the same gene when single-copy plasmid-borne, shown to be impervious to Gyrase inhibition. Next, we estimate the fraction of time in locked states and number of transcription events prior to locking, which we validate by measurements under Gyrase inhibition. Replacing the gene of interest by one with slower transcription rate decreases the fraction of time in locked states due to PSB. Finally, we combine data from both constructs to infer a range of possible transcription initiation locking kinetics in a chromosomal location, obtainable by tuning the transcription rate. We validate with measurements of transcription activity at different induction levels. This strategy for dissecting transcription initiation locking kinetics due to PSB can contribute to resolve the transcriptional programs of *E. coli* and in the engineering of synthetic genetic circuits.

1. Introduction

Transcription in *Escherichia coli* generates positive supercoiling ahead of the RNAP and negative supercoiling behind it ([11,46,51]; [87,95,99]). Discrete, topologically constrained segments along the DNA cause this process to generate local supercoiling buildup [31,33,41,70,77]. Evidence suggests that this torsional stress can affect gene activity [2,96].

E. coli has (at least) two proteins to resolve torsional stress. Namely, Gyrase removes positive supercoils [9,15,46] while Topoisomerase I removes negative supercoils [9,15,22,35,46,90]. Interestingly, in normal conditions, Topoisomerase I removes the negative supercoils at sufficient speed for R loops to not emerge, which is essential for cell survival [16]. This is made possible by the existence of a direct physical interaction between the RNAP and Topoisomerase I, allowing the latter to remove the negative supercoils, as soon as they form [7]. Contrarily to this, the removal of positive supercoils is not as efficient (being an ATP-dependent reaction likely contributes to this [77]), in the sense that positive supercoiling buildup (PSB) is commonly observed, particularly in highly active operons [17,28]. In support, measurements have shown that Topoisomerase I can relax plasmid DNA ~6 times faster than Topoisomerase IV [97], which has the same catalytic rate as Gyrase [84].

As positive supercoils accumulate, elongation slows down and, eventually, there are transient halts in transcription initiation [9,73]. These halts in initiation tangibly decrease RNA production rates and increase transcriptional noise [9,55,59]. Thus, dissection of the *in vivo* kinetics of transcription locking due to PSB is needed in order to dissect the transcriptional programs of *E. coli*.

A strategy was recently introduced for dissecting the *in vivo* kinetics of rate-limiting steps of active transcription initiation from *in vivo* measurements of individual RNA production events at different RNA polymerase (RNAP) concentrations [48]. It uses a Lineweaver–Burk plot [44] to infer the time-length of events *prior* and *after* commitment to open complex formation [58] from measurements of *in vivo*transcription rates at different RNAP concentrations ([RNAP]) [48]. This is possible due to the independence of the kinetics of the open complex formation from [RNAP], and because there is a range of values of [RNAP] for which the inverse of RNA production rate changes linearly with the inverse of [RNAP] [48].

* Corresponding author at: Tampere University, Arvo Ylpön katu 34, P.O. Box 100, 33014 Tampere, Finland.
E-mail address: andre.sanchesribeiro@tuni.fi (A.S. Ribeiro).

https://doi.org/10.1016/j.bbagrm.2020.194515
Received 14 June 2019; Received in revised form 7 February 2020; Accepted 20 February 2020
Available online 27 February 2020

Similarly, chromosomal RNA production rates (particularly of genes in highly transcribed operons) are expected to differ with Gyrase concentration, due to the existence of discrete topological constraints [9,69,70]. Thus, it should be possible to, from *in vivo* RNA production rates at different Gyrase concentrations, infer the kinetics of *in vivo* transcription locking due to PSB. For this, it must hold true that there is a range of conditions for which the inverse of the RNA production rate changes linearly with the inverse of Gyrase concentration.

Here we verify this hypothesis and then use this strategy to dissect the contribution of transcription initiation locking due to PSB on the kinetics of RNA production of a chromosome-integrated gene. We validate the estimation by time-lapse microscopy of single-RNA production kinetics of the same gene when single-copy plasmid-borne, shown to be impervious to Gyrase inhibition. Based on this, we estimate the fraction of time in locked states and the number of transcription events prior to locking, which we validate by measurements of RNA production under the inhibition of Gyrase activity by the addition of Novobiocin (see Section 2.3). Replacing our gene by a gene with a slower transcription rate, we show that changes in the *basal* transcription rate (expected rate of RNA production in the absence of effects from PSB) affect the contribution of locking due to PSB on *effective* transcription rates (measured rate of RNA production). Finally, we infer a range of possible transcription initiation locking kinetics in a chromosomal location, obtainable by tuning the basal transcription rate, and validate this inference using measurements of transcription activity at different induction levels.

2. Materials and methods

2.1. Strains and plasmids

We engineered two strains from *E. coli* BW25993 (*lacI*^{q} *hsdR514* $\Delta araBAD_{AH33}$ $\Delta rhaBAD_{LD78}$) [10]. In one strain, the target gene $P_{LacO3O1}$-mCherry-MS2-BS is integrated into a single-copy F-plasmid (~11 kbp), pBELOBAC11 (target plasmid). This plasmid is not known to form long-lasting bounds to the cell membrane and is originally responsible for the expression of transient DNA-binding proteins [27,32,60]. In the other strain this plasmid is absent and the same target gene, $P_{LacO3O1}$-mCherry-MS2-BS, is integrated into the lac locus of the genome using Red/ET recombination (Gene Bridges, Heidelberg, Germany) (Supplementary Figs. S1A and S1B). We found no significant differences in the growth rates of the two strains and the original strain.

$P_{LacO3O1}$, inducible by IPTG, was engineered from the *E. coli* native P_{Lac} by removing the O2 repressor binding site downstream of the transcription start site [62]. Thus, strong topological barriers are not expected to form when fully induced [21]. Also, both strains were transformed with the medium-copy reporter plasmid pZA25-GFP [61] (kind gift from Orna Amster-Choder, Hebrew University of Jerusalem, Israel), coding for the reporter MS2-GFP under the control of the P_{BAD} promoter. The strain with the target gene in a single-copy F-plasmid also contains the native Lac promoter in the chromosome. Thus, it has 4 LacI binding sites. The strain with the chromosome-integrated target gene only has 2 LacI binding sites, as the original Lac promoter was replaced by the target promoter, $LacO_3O_1$. However, as both strains overexpress LacI [10], effects of this difference are expected to be negligible.

Both target genes (plasmid and chromosome constructs) code for an RNA with an array of bindings sites (BS) for the modified viral coat protein MS2-GFP [25,67,68]. Due to the multiple BS in the target RNA and the strong binding affinity of each site [25], MS2-GFP tagged RNAs appear as bright spots soon after produced (Supplementary Figs. S2B and S2D). Their maximum fluorescence is reached rapidly (< 1 min) and have long half-lives (Supplementary Section I).

For overexpressing Gyrase, we constructed a plasmid (pZe11 P_{rham} gyrAB-*sfGFP*, with ampicillin resistance) with the *gyrA* and *gyrB* genes under the control of a Rhamnose promoter. These genes were arranged in a polycistronic manner, using their (identical) ribosome-binding site to maintain the physiological stoichiometry of the two subunits. We amplified the sfGFP using the primers: Forward: 5′CATATGAGCAAAG GAGAAGAACTTTT 3′, Reverse: 5′ CGGCCGTTTGTAGAGCTCATCCA TGC 3′ with restriction enzymes and cloned it after the *gyrAB* genes by digestion followed by ligation (Supplementary Fig. S3). We also constructed a plasmid without sfGFP, by digesting with the restriction enzymes *Nde*I and *Nae*I, followed by ligation, which was transferred to *E. coli* BW25993 with the $P_{LacO3O1}$-mCherry-MS2-BS integrated into a single-copy F-plasmid [27] and to *E. coli* BW25993 with the $P_{LacO3O1}$-mCherry-MS2-BS integrated in the chromosome. Finally, in another strain, we replaced the chromosome-integrated $P_{LacO3O1}$ by the native Lac promoter, followed by the same array of binding sites for MS2-GFP.

To access the intracellular levels of Gyrase A proteins, we used a strain with a gyrA gene endogenously tagged with the YFP coding sequence [85]. From the glycerol stock (-80 °C), cells were streaked on the LB agar plates and incubated at 37 °C overnight. From the plate, a single colony was picked, inoculated in an LB medium supplemented with the antibiotics, and incubated at 30 °C overnight with shaking at 250 RPM. Next, cells were diluted into fresh LB medium to an OD of 0.03 (Optical Density, 600 nm; Ultrospec 10, Amersham biosciences, UK) and grown at 37 °C with 250 RPM until it reaches to the mid-exponential phase (OD ~0.4–0.5).

2.2. Nucleoid visualization by DAPI staining

DAPI (4′,6-diamidino-2-phenylindole) stains nucleoids specifically with little or no cytoplasmic labelling. Gyrase induced and un-induced cells were grown at 37 °C and fixed with 3.7% formaldehyde in phosphate buffered saline (PBS, pH 7.4) for 30 min at room temperature, followed by washing with PBS to remove excess formaldehyde. The pellets were suspended in PBS, and DAPI (2 μg/ml) was added to the suspension. After incubating for 20 min in the dark, cells were centrifuged and washed twice with PBS to remove excess DAPI. Cells were then re-suspended in PBS and 3 μl of these cells were placed on a 1% agarose gel pad for microscopy.

2.3. Growth conditions and induction of the reporter and target gene

From a -80 °C glycerol stock, cells were placed in LB medium agar plates with 34 μg/ml Chloramphenicol and 35 μg/ml Kanamycin (Sigma-Aldrich, USA) and incubated overnight at 37 °C (Innova® 40 incubator, New Brunswick Scientific, USA). Cells were cultured in LB medium from single colonies on LB agar plates with the appropriate concentration of antibiotics and incubated overnight at optimal temperature at 250 rpm with aeration. These cultures were diluted to an optical density (OD_{600}) of 0.05 in fresh M9 medium, with a culture volume of 20 ml supplemented with the appropriate antibiotics and 0.4% of Glycerol (Sigma-Aldrich, USA), and incubated for 3 h with a 250 rpm agitation until an OD_{600} of ~ 0.3. Next, to induce MS2-GFP expression, 0.4% of L-Arabinose (Sigma-Aldrich, USA) was added and cells were incubated for another 45 min for sufficient MS2-GFP to accumulate for detecting target RNAs [26]. Next, the target gene was induced by IPTG (Sigma-Aldrich, USA) and cells were incubated for 1 h, prior to image acquisition or RT-PCR. To obtain induction curves of target genes (under the control of $P_{LacO3O1}$ and P_{Lac}), 0, 50, 100, 250, 500 and 1000 μM IPTG was added (Supplementary Fig. S4A). Unless stated otherwise, the target genes are always fully induced by 1000 μM IPTG.

We also performed measurements when inactivating and when overexpressing Gyrase. To inactive gyrase, we follow the protocol above, but when MS2-GFP expression is induced, we further added Novobiocin (100 μg/ml) [22]. Since all strains used here contain the gene *acrA*, Novobiocin at this concentration is not expected to affect cell division rate [50]. We verified this by measuring growth rates by OD_{600} for varying Novobiocin concentration (0, 25, 50, 75, 100, 200,

400, 500 ng/μl). The measurements show that growth rates to not differ significantly at 100 μg/ml or lower (Supplementary Fig. S5A). We further verified that Novobiocin does not affect morphology at these concentrations (see Section 3.6).

To overexpress Gyrase, we use Rhamnose (see previous Section) [93]. We follow the protocol above but, when inducing MS2-GFP expression, we also added Rhamnose. When applicable, Gyrase and RNAP concentrations were measured 1 h after adding Rhamnose. Gyrase overexpression did not affect bacteria growth (Supplementary Fig. S5B) nor morphology (see Section 3.2).

2.4. RT-PCR

One hour after inducing the target gene, cells were fixed by RNAProtect bacteria reagent (Qiagen, Germany), followed by enzymatic lysis with Tris-EDTA Lysozyme (15 mg/ml) buffer (pH 8.3). From the lysates, the RNA content was isolated using RNeasy purification kit (Qiagen) as per the manufacturer instructions. The RNA was then separated by electrophoresis using 1% agarose gel stained with SYBR® Safe DNA Gel Stain (Thermo Scientific, USA). The RNA was intact, with clear bands for the 16S and 23S ribosomal RNA. The RNA yield (~2 μg/μl) and absorbance ratios A260/A280 nm and A260/A230 nm were measured by a Nanovue Plus Spectrophotometer (GE Healthcare Life Sciences, USA). The ratio (2.0–2.1) indicates highly purified RNA. To remove DNA contamination, samples were treated with DNaseI (Thermo Scientific, USA) as per the manufacturer instructions. The cDNA was synthesized from 1 μg of RNA using iScript Reverse Transcription Supermix (Biorad, USA) as per the manufacturer instructions. cDNA samples (10 ng/μl) were mixed with qPCR master mix with iQ SYBR Green supermix (Biorad, USA) with primers (200 nM) for target and reference genes. 16S rRNA was used as reference. Primers set for target mRNA (mCherry) and reference (16S rRNA) genes were: mCherry (Forward: 5' CACCTACAAGGCCAAGAAGC 3′, Reverse: 5′ TGGTGTAGTCCTCGTTGTGG 3′), 16S rRNA (Forward: 5' CGTCAGCTC GTGTTGTGAA 3′, Reverse: 5′ GGACCGCTGGCAACAAAG 3′). To determine fold changes in mRNA Gyrase, cells were grown in M9 media supplemented with different Rhamnose concentrations. For the Gyrase mRNA (GyrA) gene the primer set was: Forward: 5' GGATTATGCGAT GTCGGTTCAT 3′, Reverse: 5′ CTAGCACAGTATCTGGCGGCT 3′. For mRNA sfGFP the primer set used was: Forward: 5′ GGAAAACTACCTG TTCCGTGGC 3′, Reverse: 5' ACATAACCTTCGGGCATGGCAC 3′. Experiments were performed by a Biorad MiniOpticon Real Time PCR System (Biorad, USA). The thermal cycling protocol was 40 cycles of 95 °C for 10 s, 52 °C for 30 s, and 72 °C for 30 s, with the fluorescence being read after each cycle. For each condition, we performed 3 biological replicates. qPCR efficiencies of these reactions were > 95%. No-RT and no-template controls were used to crosscheck non-specific signals and contamination. Cq values from the CFX ManagerTM Software were used to calculate fold changes in the target gene (normalized to the reference gene) and standard error, using Livak's $2^{-\Delta\Delta CT}$ method (40). RT-qPCR results are presented in Table S1.

2.5. Flow cytometry

To measure single cell Gyrase-GFP expression levels, cells were grown as described in Section 2.1. Upon reaching mid exponential phase, cells were diluted 1:1000 into 1 ml PBS vortexed for 10 s and 50,000 cells were tested in each run. Data was collected by an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego USA) using a blue laser (488 nm) for excitation and the fluorescein isothiocyanate detection channel (FITC) (530/30 nm filter) for emission, at a flow rate of 14 μl/min and a core diameter of 7.7 μM. A PMT voltage of 417 was used for FITC. To avoid background signal from particles smaller than bacteria, the detection threshold was set to 5000 in FSC-H analyses. We set the fraction of the cells used in the analysis (α) to 0.55, to remove any undesired data points from debris, cell doublets etc. Reducing α further did not change the results.

2.6. Western blot

Cells were grown as above until reaching an OD_{600} of 0.6. Pelleted cells were lysed with B-PER bacterial protein extraction reagent (Thermo scientific) and proteins were extracted. Protein samples were diluted with 4× laemmli sample loading buffer and boiled for 5 mins at 95 °C. 30 μg of proteins were loaded on the 4–20% TGX stain free pre cast gel (Biorad) and separated by electrophoresis. Proteins were then transferred to PVDF membrane using Trans-Blot Turbo transfer system (Biorad). The membrane was blocked with 5% non-fat milk and incubated with primary RpoC antibody 1:2000 dilutions (Biolegend) overnight at 4 °C and followed by HRP-secondary antibodies 1:5000 dilutions (Sigma Aldrich) for 1 h at room temperature. For band detection, the membrane was treated with a chemiluminescence reagent (Biorad). Images were acquired by the Chemidoc XRS system (Biorad). Band quantification was done using Image lab software (version 5.2.1). For each condition, we performed 3 biological replicates.

2.7. Microscopy and image analysis

Cells were grown as above, and pelleted and re-suspended in ~100 μl of the remaining media. Prior to imaging, cells were placed on a 2% agarose gel pad of M9 medium and kept in between the microscope slide and a coverslip. Cells were visualized by a Nikon Eclipse (Ti-E) inverted microscope with a 100× Apo TIRF (1.49 NA, oil) objective. Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system with a pinhole size of 1.2 AU. In confocal images, the size of a pixel corresponds to 0.062 μm using a scan area resolution of 2048 × 2048 pixels. MS2-GFP-RNA spots and GyrA-YFP regions were visualized by a 488 nm laser and a 514/30 emission filter, while DAPI-stained nucleoids were visualized by a 405 nm laser and a 447/60 emission filter.

Phase contrast images were taken by an external phase contrast system and DS-Fi2 CCD camera (Nikon). Image sizes were 2560 × 1920 pixels, each pixel corresponding to 0.048 μm. Phase contrast and confocal images were taken simultaneously by Nis-Elements software (Nikon).

From phase contrast images, we segmented cells with the software iCellFusion [78] (Supplementary Figs. S2A and S2C). Errors were manually corrected. Next, phase-contrast and corresponding fluorescence images were aligned by the software CellAging [29]. We used CellAging to detect RNA-MS2-GFP fluorescence spots (Supplementary Figs. S2B and S2D) and assess the intensity of each spot. From these, integer-valued RNA numbers were calculated for each spot (Supplementary Section I).

Nucleoid(s) segmentation was performed as in [63], using a 2D Gaussian approximation, followed by manual corrections. Cells whose size is smaller than 500 pixels were excluded from the analysis since, in general, they were not real cells (e.g. only half of the cell appeared in the image). Also removed were cells larger than 1000 pixels, as they were abnormally elongated. In general, this led to removing < 5% of the cell population.

The segmentation of the intracellular regions with significant GyrA-YFP was done using a tailored software, SCIP [56]. Errors were manually corrected. To remove measurement noise, we applied a 2D Gaussian filter to each region [94].

2.8. Models and simulations

We use stochastic models of gene expression to test if arrests during elongation, caused by PSB, disturb significantly the mean RNA production rate (within realistic intervals of parameter values).

These models are at single-cell, and single-molecule level. Specifically, two models are simulated. One is the 'Single-Nucleotide

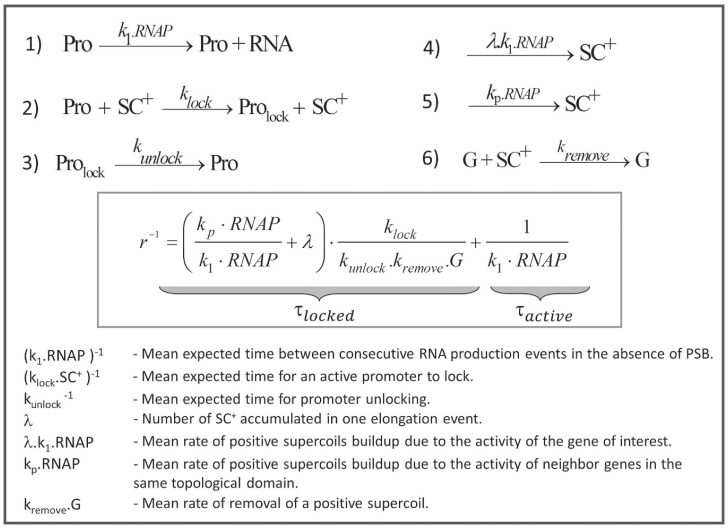

Fig. 1. Minimal (reduced) model of the dynamics of RNA production and transcription locking due to PSB of a chromosome-integrated gene in *E. coli*. The model includes the promoter when active (*Pro*) and when locked due to PSB (*Pro$_{Lock}$*), RNA molecules, RNA polymerases (*RNAP*), Gyrases (*G*), and positive supercoils (*SC$^+$*). Reaction 1 represents transcription at the rate k_1, which is the basal rate of RNA production of an active promoter assuming one RNAP in the cell. Reaction 2 models promoter locking due to PSB, with k_{lock} being the rate at which an active promoter is locked given the presence of one *SC$^+$*. Positive supercoils emerge during transcription of the gene of interest (reaction 4) and/or from the activity of genes in the same topological domain (reaction 5). λ corresponds to a tenth of the number of nucleotides of the gene of interest. Reaction 3 accounts for the unlocking of the promoter at the rate k_{unlock}. Finally, reaction (6) models the removal of positive supercoils by Gyrases. All parameter values are extracted or derived from empirical data , including measurements of a chromosome-integrated $LacO_3O_1$ promoter (k_1 = 0.0014 s^{-1}, k_{lock} = 0.0012 s^{-1} and k_{unlock} = 7 × 10^{-4} s^{-1}) (Supplementary Sections IV and V).

model' (SN Model), since elongation is modelled at the single nucleotide level (Supplementary section II). The other is the "Minimal model', as it is designed from the former, but lacks elongation at the nucleotide level (Fig. 1, Section 3.1).

The time length of each simulation is 10^5 s, found to be long enough for not underestimating the mean length of the time intervals between consecutive RNA production events (which would bias the data with right-censoring) [30]. The simulations have a reading time of 1 s^{-1}. The results shown in the Results section are obtained from 100 runs per condition, as this number suffices to obtain consistent results. Finally, the initials components at the start of simulations are 1 promoter (where the transcription start site is located), 1 Gyrase, and 28 RNAPs. In addition, the SN model has also 4058 nucleotides, in the state "unoccupied" (Supplementary Section II) along which elongation will occur.

The models are implemented in the simulator SGNSim [74] and their dynamics follow the Stochastic Simulation Algorithm [23,24]. In short, the stochastic nature of their dynamics arises from the generation of two random numbers at each step. As described in [24], one of these random numbers determines what is the next reaction (which differs with the propensity of each reaction at that moment), while the other random number determine when the next event will occur (which depends on the total propensity when considering all possible reactions combined). SGNSim makes use of 'Mersenne Twister' to produce these random variables at each step [57].

3. Results

3.1. Expected effects of changing gyrase concentration on the dynamics of transcription

We started by designing a stochastic model of transcription at the single nucleotide level (here named 'SN model'), described in detail in Supplementary Section II and shown in Supplementary Table S2. The model is based on a model proposed in [75] and later used in [53,72], to which we add positive supercoiling buildup/removal. The reactions composing the model should not be interpreted as elementary transitions. Instead, they represent the rates of the rate-limiting steps of the various events. The model dynamics and simulations are described in Materials and Methods, Section 2.8, while its assumption of homogeneous mixing of RNAP and Gyrases is validated in Supplementary Section III.3.

The model consists of the following events (Table S2): transcription starts when an RNAP finds the promoter (reaction S2.1) and unwinds the DNA for reading and escapes the promoter (reaction S2.5). After this, stepwise transcription elongation is initiated, accounting for realistic RNAP footprint in the DNA template, transcriptional pausing, arrests, editing, premature terminations, pyrophosphorolysis and collisions between RNA polymerases [53].

In addition, the model accounts for the phenomenon of production of positive supercoils during elongation [46,51,95], in reaction S2.6 in Supplementary Table S2. As these supercoils accumulate, they enhance the propensity for RNAP arrest (reaction S2.11) [20,51] and transcription initiation locking (reaction S2.2) [9]. The removal of positive supercoils by Gyrase [84] is also modelled explicitly (reaction S2.17). Finally, the model accounts for the potential accumulation of positive supercoils due to transcriptional activity of neighbor genes (reaction S2.4).

This model does not include RNA degradation, as we measured RNA numbers by MS2-GFP tagging, which prevents degradation for a few hours (Supplementary Section I and Figs. S1 and S2) [26,86], thus avoiding this source of noise. Further, for the purposes of this work, we are only interested in RNA production rates, which do not depend on degradation.

Results in Supplementary Sections III.1 and III.2 show that this SN model mimics the effects of PSB on the kinetics of stepwise transcription elongation. Namely, from Supplementary Fig. S6A, one finds that the mean elongation time increases as Gyrase numbers are decreased. Meanwhile, from Supplementary Fig. S6B, one finds that this slowdown of stepwise elongation does not affect the mean rate of RNA production, within the range of parameter values tested, which is expected.

Note that in this model, for the realistic range of parameter values considered, once the system reaches steady state (near constant number of RNAPs on the DNA strand), the mean RNA production rate depends only on the rate with which RNAPs initiate new elongation events (reaction S2.1 in Table S2) and on the rate of abortions of elongation (reaction S2.14 in Table S2), with the latter being near negligible (~4% per transcription initiation event). Only in the unlikely scenario of excessive accumulation of RNAPs in the DNA template that would jam the promoter region, would events in elongation affect the mean RNA production rate.

Therefore, for purposes of estimating the effects of changing Gyrase

concentration on the mean RNA production rate, we instead use a minimal model, where elongation is not explicitly represented. Fig. 1 shows the minimal model that, as the SN model, also has a stochastic dynamics in accordance to the SSA (Materials and Methods, Section 2.8). In detail, reaction 1 models RNA production by an active promoter, *Pro*, and its propensity differs with the basal transcription rate, k_1, and with RNA polymerase numbers. Meanwhile, reaction 2, which models transcription locking, is in all identical to reaction S2.2 of the SN model and, thus, its propensity differs with the number of positive supercoils (Supplementary Section II).

Positive supercoils can be generated via reactions 4 and 5 (also as in the SN model) [1,9,51,70,83,84,89]. The propensity of reaction 4 depends on the basal transcription rate, k_1, of the gene of interest, as described in [17] and in agreement with results from anchored plasmids [9,87] as well as with results reported here. The parameter λ in reaction 4 accounts for the length of the gene of interest (RNAP will take longer to transcribe a longer gene, during which time positive supercoils are produced). As for reaction 5, responsible for the accumulation of PSB due to the transcription activity in the topological domain of the gene of interest, its kinetics differs with the neighboring activity, which can be tuned by k_p and RNAP numbers (Fig. 1).

Once locked, a promoter can become unlocked via reaction 3. The unlocking kinetics can be tuned by the rate constant k_{unlock}. Because the propensity for locking changes linearly with the number of positive supercoils, the propensity for reaction 3 is kept independent from this number. Else, the overall time spent in locked states would change quadratically with the inverse of Gyrase numbers, and not linearly (Fig. 3 provides empirical support for the assumption that this relationship is linear within realistic ranges of parameter values). Finally, reaction 6 represents the removal of positive supercoils by Gyrases. As this takes place, the propensity for reaction 2 decreases, thereby accounting for the expected decrease in the effects of PSB with increasing Gyrase numbers [9].

To verify that the minimal model constitutes a valid approximation of the SN model, we performed simulations for various Gyrase numbers. Visibly, from Supplementary Fig. S6E, the minimal model matches the mean rate of RNA production of the SN model (and the empirical data) as a function of Gyrase numbers. This is expected since, as noted, all its parameter values are the same as in the SN model, except for reaction 4 in Fig. 1, since this reaction needs to account for the number of nucleotides of the gene of interest (which are modelled explicitly in the SN model). This adjustment is done by having the rate of SC^+ production of the gene of interest dependent on its nucleotide length (with λ equaling a tenth of its number of nucleotides, as this is approximately the expected number of SC^+ produced during one elongation event [84]).

We then derived an analytical solution of the minimal model, for the inverse of the mean rate of RNA production (r^{-1}) as a function of Gyrase (inset of Fig. 1). Here, τ_{active} is the mean time between consecutive RNA production events of an unlocked/active promoter, which equals the inverse of $k_1 \times$ RNAP (with RNAP being the number of RNA polymerases). Meanwhile, r is the inverse of the sum of τ_{active} and τ_{locked}, with the latter being the mean time spent in locked states (equation in the larger inset in Fig. 1). From this solution, we find that increasing [G] decreases τ_{locked} [9,46], which increases r. In detail, r^{-1} is expected to change linearly with $[G]^{-1}$ (large inset, Fig. 1). If this holds true, from measurements of r and [G], it should be possible to extrapolate τ_{active}, since τ_{active} should equal r^{-1} for infinite [G]. Further, from τ_{active} and r, it should be possible to estimate τ_{locked}. Finally, note that while k_1 does not affect the mean time for Gyrase to release the gene from a locked state, it does affect the rate of occurrence of locked states.

Interestingly, many plasmids only have weak, transient topological barriers (such as short-term protein-DNA complexes [42]). In particular, aside from when they are anchored to the membrane [3,11,49,71] or have many tandem copies of a DNA-binding site [42], no long-term PSB is expected, since positive and negative supercoils diffuse in opposite directions and annihilate one another [42] (unlike in the chromosome that has topological barriers). As such, it should be possible to simulate the dynamics of plasmid-borne genes using the model in Fig. 1, by setting k_{lock} to null, causing τ_{locked} to be null. Consequently, r^{-1} of a model plasmid-borne gene should equal τ_{active} of the same model gene, when chromosome-integrated. Further, if this holds true, then a plasmid-borne gene can be used as a proxy for the same gene when chromosome-integrated when unaffected by PSB.

3.2. Changing intracellular concentration of gyrases

Above, we hypothesized that r^{-1} should be linear with respect to $[G]^{-1}$ within a given range of Gyrase concentrations (see Fig. 1 and Supplementary Section VI). If true, one should observe a line on a Lineweaver–Burk plot [44] of r^{-1} against $[G]^{-1}$, from which one can extrapolate τ_{active}. From τ_{active} and r, one can then estimate τ_{locked}.

To test this hypothesis, it is necessary to measure r in cells differing in [G]. For this, we inserted a plasmid carrying a copy of the *gyraseA* and *gyraseB* genes under the control of the Rhamnose promoter (pZe11 P_{rham} gyrAB, Materials and Methods). We further added sfGFP, also under the control of the Rhamnose promoter (pZe11 P_{rham} gyrAB-*sfGFP*, Section 2.1 and Supplementary Fig. S4). The region coding for sfGFP allows measuring mRNA coding for Gyrase and the corresponding protein levels produced solely by the plasmid.

We subjected cells to different Rhamnose concentrations until finding a range for which the production rate of the mRNA coding for Gyrase increases linearly with Rhamnose concentration. For this, we performed qPCR using the region of the RNA from the plasmid that is absent in the native RNA coding for Gyrase (i.e. the region coding for sfGFP). In Fig. 2A we find a linear relationship between mRNA fold changes (measured by qPCR) and Rhamnose (0, 0.1, 0.2 and 0.4%). In particular, small deviations from linearity were rejected (p-value > 0.5, see Fig. 2 legend for details).

Next, we verified that cell growth rates were not disturbed in this

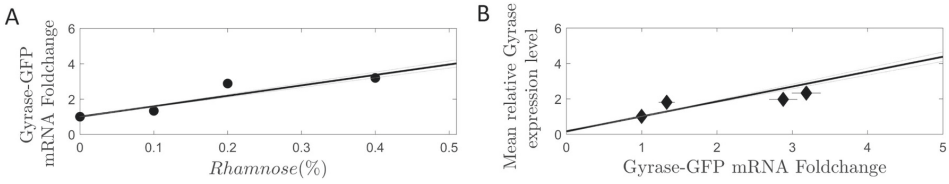

Fig. 2. Gyrase overexpression levels. (A) Fold change of mRNA Gyrase-GFP measured by qPCR, for different concentrations of Rhamnose (0, 0.1, 0.2 and 0.4%), standard error of the mean (vertical error bars) are not visible. Results are relative to the control condition (0% Rhamnose). Also shown is the best-fit line. (B) Calibration line between relative Gyrase-GFP mean expression levels (measured by flow cytometer with the FITC-H detection channel) and mRNA Gyrase-GFP fold change measured by qPCR for 0, 0.1, 0.2 and 0.4% Rhamnose. Gyrase numbers are relative to the 0% Rhamnose condition. Horizontal error bars represent the standard error of the mean. For both figures, we fitted a first order polynomial to the data points by WTLS by minimizing χ^2 [39]. To determine if small deviations from linearity are statistically significant, we performed a likelihood ratio test between the best linear fit and fits by higher order polynomials. In both cases, the test did not reject the linear model (p-values larger than 0.5 and 0.15, for Fig. 2A and B, respectively).

range (Supplementary Fig. S5B). Further, to test if morphology was affected, we measured cell areas in the control condition (165 cells analyzed) and when subject to 0.4% Rhamnose (182 cells analyzed). The cell area was obtained from phase contrast images, using the software iCellFusion (Section 2.7). We performed a 2-sample Kolmogorov-Smirnov test and found that, at the significance level of 0.05, the two distributions cannot be distinguished (p-value of 0.5).

Finally, we verified that, within this range of conditions, the mean relative Gyrase expression level changed linearly with the RNA production rate of the plasmid coding for Gyrase (Fig. 2B), as measured by Flow-cytometry (Section 2.5). In particular, small deviations from linearity were rejected (p-value of 0.15, see Fig. 2 legend for details). We thus conclude that the fold change in Gyrase-GFP protein levels corresponds to the fold change in the mRNA coding for Gyrase-GFP (Supplementary Section VII and Table S3 show the parameters of the calibration line and procedure).

We expect the quantitative relationship between mRNA and protein numbers of the plasmid-borne Gyrase to be the same as in the native Gyrase mRNA and proteins, since we used the native ribosome binding site in the plasmid construct. Thus, we measured by qPCR the fold change of the mRNA produced by both the native and the plasmid-borne Gyrase genes and used the line in Fig. 2B as a calibration line, to estimate the fold change with Rhamnose in Gyrase protein levels (Supplementary Section VII and Table S3).

Finally, we considered that Gyrase overexpression could change the proteome and, eventually, change cellular functioning (e.g. in 1–2 h). To mitigate effects from this eventuality (to avoid unknown changes in the processes represented in Fig. 1), subsequent measurements were conducted 1 h after inducing Gyrase overexpression (Materials and Methods). Given this and the above, we expect that, for 0.2% or lower Rhamnose concentrations (Fig. 2A), changes in RNA production rate in this time window are largely due to changes in concentrations of the components of the reactions in Fig. 1.

3.3. Transcription rate of a chromosome-integrated gene under the control of $P_{LacO3O1}$ in the absence of positive supercoiling buildup

Data in [85] indicates that the expression rate of (at least) three of the RNAP sub-units are, in normal conditions, approximately double the average expression rate of *E. coli* genes. Since several highly expressed genes are supercoiling sensitive [17], it is tangible that Gyrase overexpression may affect [RNAP], which according to the model (Fig. 1), could affect the transcription rate (r) of our gene of interest (Fig. 1). Thus, we first assessed for potential fast changes in [RNAP] when overexpressing Gyrase.

For this, we used the same plasmid as above, with the *gyrA* and *gyrB* genes controlled by a Rhamnose promoter, to overexpress Gyrase (but having removed sfGFP, so as to not affect RNA counting or Gyrase functioning, see Materials and Methods). Next, we measured [RNAP] at the different Rhamnose concentrations (0%, 0.1% and 0.2%) by measuring the RpoC protein by Western blot, 1 h after inducing Gyrase overexpression. From Fig. 3A and B, at the same OD_{600}, the [RNAP] differs by 12% between the two extreme conditions. This difference was found to be statistically significant by a 1-sample 2-tailed t-test, with the null hypothesis that the increase is 12% (p-value of 0.42). In addition, we performed a 2-sample, 2-tailed t-test with the null hypothesis that there is no difference between the conditions, which was rejected (p-value of 0.0008). This is expected to partially explain changes in r due to Gyrase overexpression and, thus, needs to be accounted for when quantifying the direct effects of changing [G] (Supplementary Section VIII).

In addition, it is tangible that overexpression of Gyrase could affect the negative supercoiling state of the chromosome, e.g. by introducing negative supercoils [6,47,80]. This, in turn, could affect DNA supercoiling density and its folding and compaction [36,92], which could alter transcription rates by affecting the time-lengths of open complex formations [52].

Unfortunately, we cannot measure directly the *in vivo* kinetics of open complex formation at a given Gyrase concentration, as this would require measuring the *in vivo* transcription in cells with differing RNAP concentrations [48,81], which would also affect the intracellular Gyrase concentration. Therefore, instead, we estimated indirectly if Gyrase overexpression (between 0% and 0.2% Rhamnose) suffices to alter significantly the chromosome folding and compaction. For this, we assessed if the nucleoid area (with area being a proxy for compaction strength) is altered by Gyrase overexpression, using DAPI staining and image analysis (Sections 2.2 and 2.7).

The mean and standard deviation of the nucleoid area, when and when not overexpressing Gyrase, are shown in Table S4. We performed a 2-sample student t-test for the null hypothesis that the two data sets of absolute nucleoid area come from the same distribution. The test did not reject the null-hypothesis (p-value > 0.01). We thus conclude that, in the range of Gyrase overexpression levels used here, the nucleoid size was not significantly affected. As such, we do not expect the indirect effects of Gyrase overexpression on DNA supercoiling density to significantly affect the kinetics of open complex formation.

Given this, we again used the plasmid with the *gyrA* and *gyrB* genes controlled by a Rhamnose promoter (without sfGFP) to study the effects of Gyrase overexpression on transcription initiation locking due to PSB of a chromosome-integrated gene under the control of $P_{LacO3O1}$. This promoter was used as its dynamics has been previously studied when plasmid-borne, including using single-RNA MS2-GFP tagging [34,54,64,66,81].

We first measured the absolute mean r^{-1} in the control condition (Materials and Methods, Section 2.3) by microscopy measurements of integer valued RNA numbers in single cells at different time moments (Supplementary Section IX). The absolute mean r^{-1} in the control condition was found to equal 1476 s, with a standard error of 145 s.

Using the value of r^{-1} in the control condition, we scaled the relative qPCR values to obtain the values of r^{-1} in conditions where Gyrase is overexpressed. Results from qPCR are shown in Supplementary Table S1. In these, the gradually increasing expression of Gyrase did not affect significantly the expression of the 16S rRNA gene. This is expected, since 16S rRNA is a stable component of ribosomes and, thus, should not change significantly between conditions when growth rates are not affected significantly [88] (Supplementary Fig. S5B). Namely, even if the small changes between conditions were considered significant, there is no monotonic change with increasing Rhamnose concentration. As such, 16S rRNA is used as the reference gene.

Next, from Supplementary Table S1 and the microscopy data in the control condition, we obtained absolute rates of RNA production in each condition (black circles in Fig. 3C). Finally, we fitted a line by weighted total least squares (WTLS) [39] (black line in Fig. 3C) to estimate τ_{active} (where the line intersects the Y-axis), when not accounting for changes in [RNAP]. We performed a likelihood ratio test between the best linear fit and fits of higher order polynomials which showed that the linear model best fits the data (p-value > 0.9).

Next, from the [RNAP] in each condition (Fig. 3B) and the model fitting (Supplementary Section IV), we estimated the effects of changes in [RNAP] (Supplementary Section VIII and Supplementary Table S5). Supplementary Figs. S7A and S7B show the Z surfaces of the best fitting models and empirical results. From the R^2 values (legend of Supplementary Fig. S7), one finds that the model well-fits the empirical data. We thus used this model to estimate the weight of the changes in [RNAP] (Fig. 3B) on r^{-1} and then quantified the changes in r^{-1} due to changes in [G] alone. Results are shown in the blue circles in Fig. 3C. Next, we fitted a line (blue line in Fig. 3C) to these data points using WTLS, from which we estimated r^{-1} for infinite [G] (i.e. τ_{active}) to be 749 ± 247 s.

Finally, we determined if the small deviations from linearity are statistically significant by performing a likelihood ratio test between

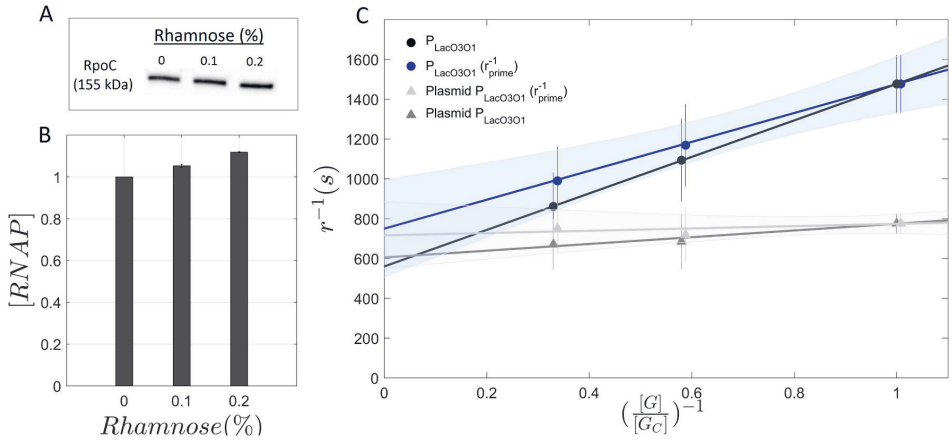

Fig. 3. Effects of Gyrase overexpression in the RNA production rate of $LacO_3O_1$ when chromosome-integrated and when plasmid-borne. (A) Replicate of Western Blot image of RpoC subunit for cells subjected to 0, 0.1% and 0.2% Rhamnose. (B) Bar chart of [RNAP] fold change with Gyrase overexpression, relative to the control condition (0% Rhamnose). In all conditions, OD_{600} was 0.6. (C) LineWeaver-Burk plot of the inverse of the RNA production rate (r^{-1}), for different Gyrase concentrations (black circles), relative to the control ($[G]/[G_C] = 1$) of the chromosome-integrated construct. Also shown is the standard error of the mean (vertical error bars), along with the best-fit line (black line). Further shown are the RNA production rates after correcting for the weight of the changes in [RNAP] (r^{-1}_{prime}), when overexpressing Gyrase (blue circles) and the correspondent best-fit line (blue solid line) and its standard error of the mean (light blue area) obtained by Monte Carlo simulations (5000 iterations). Blue circles are 0.008 units deviated to the right, for figure legibility. The equations of the black and blue lines are $r^{-1} = (917 \pm 329) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (559 \pm 246)$ and $r^{-1} = (726 \pm 329) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (749 \pm 247)$, respectively. Finally, the dark grey triangles are the values of r^{-1} for the plasmid-borne construct, when subject to the same levels of Gyrase overexpression while the light grey triangles correspond to r^{-1} after correcting for the weight of the changes in [RNAP] on r^{-1} (dark grey triangles). Light grey triangles are 0.008 units deviated to the right, for figure legibility. Also shown are the respective best-fit lines and its standard errors of the mean (light grey area) obtained by Monte Carlo simulations (5000 iterations). The equation of the dark grey line is $r^{-1} = (168 \pm 184) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (605 \pm 167)$. The equation of the light grey line is $r^{-1} = (58 \pm 184) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (715 \pm 167)$. Data from 368 cells (chromosome-integrated gene) and 476 cells (plasmid-borne gene). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the best linear fit and fits by higher order polynomials (by WTLS by minimizing χ^2) [39]. The test did not reject the linear model (p-value > 0.9), from which we conclude that r^{-1} decreases linearly with $[G]^{-1}$.

Several phenomena could have forced this plot to be non-linear. E.g., if the ratio between free and total Gyrase concentrations would increase as Gyrase is overexpressed, the plot would exhibit negative curvature (see Section VI in Supplementary). Meanwhile, if the resolution of supercoils in the control condition was near-saturation, overexpressing Gyrase would result in positive curvature. We therefore interpret the observed linearity as evidence that these changes in r^{-1} are largely due to changes in [RNAP] and [G] as assumed by the model in Fig. 1, rather than due to unknown factors.

3.4. Transcription kinetics of $P_{LacO3O1}$ when single-copy plasmid-borne

To validate the estimation of τ_{active}, we integrated the same gene under the control of $P_{LacO3O1}$ into a single-copy plasmid (Materials and Methods). We expect this to reduce the effects of PSB on the activity of the gene of interest to a minimum. I.e., the value of r^{-1} of the single copy plasmid-borne gene should approximate the estimated τ_{active} of the chromosome-integrated gene. If this holds true, adding Novobiocin, which inhibits Gyrase activity [9,15,22,91], should not disturb significantly its activity.

To test this, we performed time-lapse microscopy measurements of RNA numbers in cells subject to 100 μg/ml Novobiocin (Materials and Methods). Images were taken every 15 min, starting 30 min after introducing 1 mM IPTG in the media to ensure full induction of the target gene [86]. We also performed measurements where Novobiocin was not added.

From Supplementary Fig. S4B, the RNA production rate of the plasmid-borne gene is not affected by the addition of Novobiocin, as expected if PSB is absent. Meanwhile, in the absence of Novobiocin, we observe the same behavior but higher r, which is consistent with the cells subject to Novobiocin having lesser number of active RNAP and/or σ factors [8,13,18,76], etc. Further, both behaviors are significantly different from cells with the chromosome-integrated construct subject to Novobiocin, where a clear blocking of the RNA production is observed shortly after adding Novobiocin (Fig. 4, blue line). We conclude that the gene in the single-copy plasmid is not directly affected by Novobiocin, suggesting that it is impervious to the effects from PSB.

In support, according to the model (Fig. 1), for equal mean RNA production rate, the kinetics of RNA production from a gene unaffected by PSB (such as when on a single-copy plasmid) should be less noisy than otherwise (e.g. when chromosome-integrated) [64]. Lesser noise should reduce cell-to-cell variability in RNA numbers. To test this, we compared the squared coefficient of variation of RNA numbers in single cells, CV^2(RNA), in conditions where the two constructs exhibit the same mean RNA numbers per cell (50 μM IPTG for the plasmid-borne gene and 1000 μM IPTG for the chromosome-integrated gene, Supplementary Fig. S4A). The CV^2(RNA) in cells with the chromosome-integrated construct is found to be much higher than in cells with the single-copy plasmid-borne gene (3.18 and 1.58, respectively), in agreement with the model prediction, even though the plasmid-borne gene is being partially affected by LacI repression, which adds variability in RNA numbers [48].

Finally, we verified that the RNA production rate of the single-copy plasmid construct equals the inverse of τ_{active} of the chromosome construct. For this, we performed microscopy measurements of the integer valued RNA numbers in cells with the plasmid construct and estimated r^{-1} to be 775 ± 50 s (dark grey triangle in Fig. 3C, for the control condition). This result cannot be distinguished, in a statistical sense,

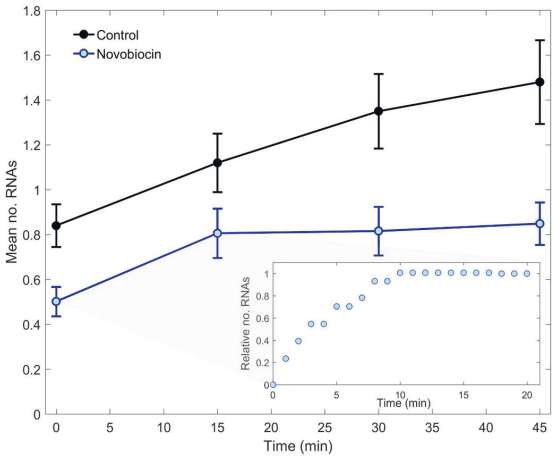

Fig. 4. RNA production over time. Mean integer-valued RNA numbers in individual cells with the chromosome-integrated $P_{LacO3O1}$ when subjected to 100 μg/ml Novobiocin (blue line) and in the control condition (black line, absent of Novobiocin). Measurements performed by microscopy, with single RNA tagging by MS2-GFP. For each time point, new cells were taken from the original culture. On average, 200 cells were used per condition. Error bars represent the standard error of the mean. Finally, the inset shows the number of RNA production events per cell relative to the total number of RNAs produced during the measurement time. Data collected at the single RNA level, from time-lapse microscopy measurements with images taken once per minute. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

from the estimate of τ_{active} for the chromosome construct assuming infinite [G] (749 ± 247 s) (in agreement with the model predictions, Fig. 1).

In this regard, in Supplementary Section X, we estimated the minimum PSB effects that would be detectable, provided the same degree of sensitivity in the measurements of r^{-1}. We found that there needs to exist a fold change between two conditions of, at least, 1.6. However, we estimate that tripling the number of data points collected allows a reduction of this number to 1.2.

Further, we also performed qPCR measurements of the plasmid construct when subject to the same levels of Gyrase overexpression. Results are shown in Fig. 3C. Next, we fitted a line (dark grey line in Fig. 3C) to the data points. From this, we find that the change in RNA production rate of the plasmid gene with gyrase overexpression is ~5 times weaker than in the chromosome-integrated construct. Next, using WTLS [39] we tested if the small deviations from linearity are statistically significant. The test did not reject the linear model (p-value > 0.8). Subsequently, as before, we corrected the data points to account for the changes in RNAP concentrations. Results in Fig. 3 (light grey line) show that the corrected line is nearly horizontal and, as expected, cannot be distinguished from a horizontal line, in a statistically sense, using the same test as above. We conclude that the RNA production kinetics from the plasmid construct is nearly non-responsive to Gyrase overexpression.

In this regard, note that estimation of τ_{active} when accounting for changes in [RNAP] (blue line in Fig. 3C) fits the measurements better (light grey line in Fig. 3C), then when not accounting for [RNAP] changes (black line in Fig. 3C). From comparing the blue and black circles in each condition, we also find that, e.g., for maximum Gyrase (0.2% Rhamnose), the increase in [RNAP] accounts for 31% of the decrease in r^{-1} relative to the control, with the remaining 69% being due to increased [G] (and/or other, unknown factors). Similarly, in the extrapolated condition of infinite [G], the increase in [RNAP], compared to the control condition, accounts for 41% of the decrease in r^{-1}, with the remaining 59% being due to increased [G].

3.5. Mean time spent in locked states and average number of transcription events between consecutive locking events

Since Fig. 3 shows that r^{-1} changes linearly with $[G]^{-1}$, we used the Lineweaver-Burk equation [44] to estimate the mean time spent in locked states, τ_{locked}, as follows:

$$\tau_{locked} = \frac{[G]_2 (r_2^{-1} - r_1^{-1})}{([G]_1 - [G]_2)} \quad (1)$$

From (1), given the control and the condition where relative $[G]^{-1}$ is 0.33 (0% and 0.2% Rhamnose, respectively), we infer τ_{locked} to be 735 s, with a standard error of the mean (SEM) of 341 s (obtained by the Delta Method [5]). Using the other pair of conditions (0% and 0.1% Rhamnose) we obtain the same result, in a statistical sense. As the mean time interval between transcription events is 1476 s, we estimate transcription initiation locking due to PSB to account for ~50% of this interval.

Meanwhile, to estimate the mean number of transcription events between consecutive locked states, N, consider that, according to the model:

$$N = \frac{\tau_{escape}}{\tau_{locked}} \quad (2)$$

To solve for N, we used the value of τ_{locked} obtained above, and τ_{escape} obtained from measurements in [9], which reported that the average DNA binding time of Gyrase is ~333 s while the unbind time is $\sim 10^3$ s [9]. Since Gyrase is expected to resolve multiple positive supercoils during this time [1,84], we assumed that the sum of these times (~1333 s) is an upper bound of the time for a locked gene to escape PSB (i.e. τ_{escape}). Introducing the estimated values of τ_{locked} and τ_{escape} in eq. 2, we find that N equals ~1.8 ± 0.84.

3.6. Kinetics of transcription initiation locking in the presence of a gyrase inhibitor

To validate the above estimations, we performed time series measurements at the single-RNA level in cells carrying the chromosome-integrated $P_{LacO3O1}$ subject to Novobiocin, a Gyrase inhibitor [22]. Assuming that, when Novobiocin first enters the cytoplasm, $P_{LacO3O1}$ activity is not subject to PSB, then the mean number of RNAs produced until transcription ceases should correspond to the mean number of transcription events between consecutive locking events. As it is not likely that the gene of interest is absent of effects from PSB in all cells, the empirical result should correspond to a lower bound estimate. Interestingly, from the same experiment, it should also be possible to measure τ_{active} (Fig. 1) from the time for RNA production to cease in all cells.

First, we tested whether Novobiocin, at the concentrations used here, affects cell morphology. For this, as above, we measured cell areas in the control condition (165 cells analyzed) and when subject to 100 ng /ml Novobiocin (180 cells analyzed), and then performed a 2-sample Kolmogorov-Smirnov test. We found that, at the significance level of 0.05, the two distributions cannot be distinguished (p-value of 0.13).

Next, we measured integer-valued number of RNAs in individual cells over time, every 15 min, 45 min after inducing the target gene (with IPTG) and adding Novobiocin (Gyrase inhibitor), so as to account for the mean time taken by cells to intake IPTG [64,86] and because only at this moment did we observe any tangible reduction in transcription activity (inset in Fig. 4). RNAs were detected by MS2-GFP tagging, preventing RNA degradation (Materials and Methods). We also performed a control experiment, where Novobiocin was not introduced.

Results in Fig. 4 show that when and only when adding Novobiocin,

the RNA production ceases. In the presence of Novobiocin, on average we observed 0.8 ± 0.11 RNAs per cell after 15 min. Considering mean cell division times (Fig. S5), we estimated the mean number of RNAs produced per cell for 15 min to be ~1.04 ± 0.14. This agrees (statistically) with the above estimation of N (~1.8 ± 0.84). It also agrees with past estimations that, in live cells, transcription initiation locking can occur after less than 5 transcription events [9].

We also extracted the time for transcription events to cease after introducing Novobiocin. For this, we performed additional time-lapse microscopy (1 min interval between images). The number of RNAs produced in individual cells during the observation time were obtained as in (66) and verified by visual inspection. Results in the inset of Fig. 4 show that transcription activity started to be reduced at minute 1 and that no RNA was produced after 10 min, which can be used as a lower bound for τ_{active} (see above). This agrees with the previous estimation of τ_{active} (~12 ± 4 min) from Fig. 3.

3.7. Effects of PSB differ with the basal transcription rate

Previous works reported evidence that a gene's activity affects its own PSB when the gene is on a circular template tethered to a surface [9,87]. We hypothesized that the same occurs on a chromosome-integrated gene, due to discrete topological constraints. This follows from the reasoning that, if the expected time interval between consecutive transcription events becomes longer, while [G] is kept constant, there is more time for Gyrase to resolve transcription initiation locking due to PSB in between transcription events. The model in Fig. 1 accounts for this, as the responsiveness of r^{-1} to changes in [G] should decrease with k_1. To test this, we replaced $P_{LacO3O1}$ by a native Lac promoter (P_{Lac}). We chose this promoter because it has similar sequence and repression-activation mechanism (Methods), which could affect PSB, and because it exhibits slower RNA production when fully induced (Supplementary Fig. S4A). By being in the same location, we expect the contribution to PSB from the activity of neighboring genes to be the same.

First, we obtained an induction curve of P_{Lac} (Supplementary Fig. S4A). Visibly, under maximum induction, P_{Lac} has a slower transcription rate than $P_{LacO3O1}$ (less ~62% MS2-GFP tagged RNAs per cell). In detail, r^{-1} (P_{Lac}) equals 2704 ± 493 s (obtained as described in Supplementary Section IX).

Next, we measured by qPCR the transcription rate for various [G] (as in Fig. 3). Results were scaled by r^{-1} in the control condition (Fig. 5A, black diamonds). Afterwards, we fitted a line by WTLS (black line in Fig. 5A) and corrected its slope by accounting for changes in [RNAP] (Supplementary Section VIII). Finally, we fitted a (green) line by WTLS to the corrected data points (green diamonds in Fig. 5A). From the best fitting (green) line in Fig. 5A we find that, for maximum [G] (0.2% Rhamnose), the increase in [RNAP] explains 28% of the increase in r, with the remaining 72% being due to increased [G] and/or unknown factors.

To assess if the effects of PSB differ with the promoter strength, we plotted r^{-1} against $([G]/[G_C])^{-1}$ for both constructs (P_{Lac} and $P_{LacO3O1}$). Results in Fig. 5B show that r^{-1} decreases faster with [G] for $P_{LacO3O1}$ (in agreement with the model). We thus conclude that changing [G] has smaller effects in the effective transcription rate of the lesser active promoter (P_{Lac}).

3.8. Inference of the parameter values of the model that best fit the empirical data and prediction of τ_{locked} as a function of the basal transcription rate

We searched for parameter values for the model (Fig. 1) that best match the empirical data of both $P_{LacO3O1}$ and P_{Lac}, assuming that they differ only in the basal transcription rate (k_1). We found that the model fits the empirical data with a mean squared error of 0.0004 and R^2 values larger than 0.95 (Supplementary Figs. S7A and S7B).

From the fitting, we obtained the parameter values (α, β_1, β_2, and η, Supplementary Section V) and inferred the duty cycles of transcription initiation locking due to neighboring and 'self-produced' PSB, for each [G] (Table S6). From Table S6, the lower are [G] and τ_{active}, the longer will the gene remain locked and the higher is its OFF/ON duty cycle ratio.

In addition, we used the inferred values of α, β_1, β_2, and η, to extrapolate τ_{locked} relative to r^{-1} as a function of $\frac{[G]}{[G_C]}$ and $\frac{1}{k_1 \times [RNAP]}$. The inferred surface is shown in Fig. 6.

If the differences in the locking dynamics due to PSB of $P_{LacO3O1}$ and P_{Lac} are solely due to the difference in their values of k_1, as hypothesized, this surface should fit other empirical values of $\frac{1}{k_1 \times [RNAP]}$ and $\frac{\tau_{Locked}}{r^{-1}}$ obtained when changing k_1 (e.g. by tuning their induction strength).

We thus performed qPCR measurements when inducing $P_{LacO3O1}$ with 0 and 50 μM IPTG (from Supplementary Fig. S4A, note that, at these concentrations, the number of RNAs produced differs significantly from maximum induction). The results from qPCR measurements, added to Fig. 6, fit well the predicted surface, suggesting that combining data from a promoter(s) of differing basal transcription rates in the same location in the DNA one can predict a state space of possible kinetics of transcription initiation locking of genes differing in k_1 in a given chromosomal location.

4. Discussion and conclusions

Past studies have shown that DNA topology and gene expression mutually affect one another [4,14,37,38].

We found that, for a certain range of Gyrase concentrations, the inverse of the transcription rate of a chromosome-integrated gene controlled by $LacO_3O_1$ changes linearly with the inverse of Gyrase concentration, while not perturbing cell growth or morphology. Given this, we developed and validated a method that uses a LineWeaver-Burk plot to dissect, from single-cell, single-RNA data, key kinetic parameters of transcription initiation locking due to PSB. Namely, we dissected the rate of occurrence of these locks and their weight on the effective RNA production rate. Next, we compared with a promoter at the same chromosomal location and similar in structure and regulation but differing in strength. From this, we inferred a range of potential kinetics of transcription initiation locking in a given topological domain that can be achieved by tuning the basal transcriptional rate of the gene of interest. Relevantly, the method was sensitive to detect PSB effects causing a minimum of 1.6 fold changes in transcription rates. Further, we estimate that simple enhancements (e.g. increasing the number of data points used for the LineWeaver-Burk plot from 3 to 10) reduced this to 1.2 fold changes. Other improvements (e.g. higher precision in data collection) should further enhance the sensitivity, which should suffice to, e.g., dissect the effects of interference between closely spaced promoters (Supplementary Section X).

To an extent, the interpretation of the empirical data relies on the models and, thus, it is necessary to assess their reliability, i.e. the robustness of their predictions. In this regard, we observed that, first, the models accurately estimated how much of the change in r^{-1}, following Gyrase overexpression, is due to changes in RNAP numbers (Fig. 3). In detail, the dynamics of the chromosome integrated gene, when corrected for RNAP changes (with this correction relying on the model), only differs from the plasmid dynamics by 3.5% (not statistically significant) while, prior to considering the model, it differed by 28%. Second, the model predicted the mean time to lock the promoter due to PSB (τ_{active} ~12 min) from qPCR and population level microscopy data. This estimation was validated by direct measurements using time-lapse microscopy data at the single cell level (one image per minute). In detail, estimated and real data differed solely by ~15% that, when accounting for the measurement error, is also not statistically significant (Figs. 3 and 4). Further, the model accurately predicted (in a statistical sense) how much the basal transcription rate affects the

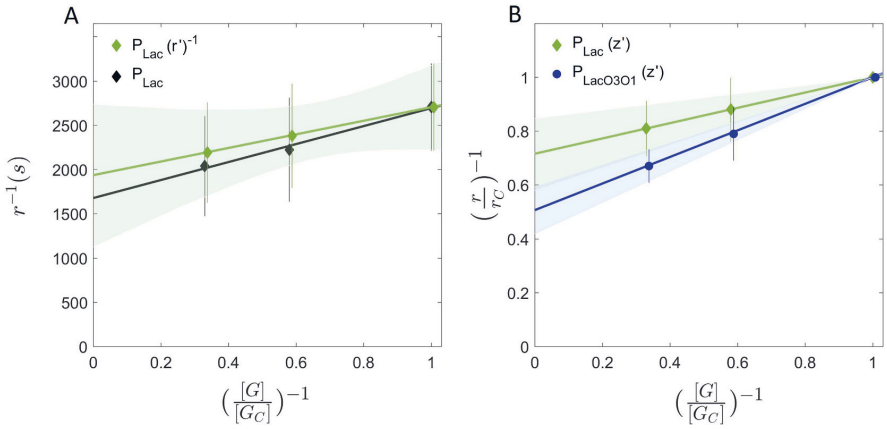

Fig. 5. LineWeaver-Burk plots for P_{Lac} and $P_{LacO3O1}$. (A) LineWeaver-Burk plot of the inverse of the RNA production rate of the chromosome-integrated Lac gene for different Gyrase concentrations (black diamonds), relative to the control (0% Rhamnose). Also shown are the standard error of the mean (vertical error bars), along with the best-fit line (black line). Further shown are the inverse of the RNA production rates corrected for the increased RNAP concentration when overexpressing Gyrase (r'^{-1}), and the correspondent best-fit line (green line) and its standard error of the mean (light green area) obtained by Monte Carlo simulations (10,000 iterations). The green diamonds are 0.008 units deviated to the right, for figure legibility. The line equations are $r^{-1} = (768 \pm 1096) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (1936 \pm 802)$ and $r^{-1} = (1016 \pm 1096) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (1677 \pm 802)$ for the green and black lines, respectively. RNA production rates were obtained by qPCR and microscopy. (B) LineWeaver-Burk plot of the inverse of the fold change in RNA production rate of the chromosome-integrated gene under the control of $LacO_3O_1$ (blue circles) and of the chromosome-integrated under the control of Lac gene (green diamonds) against the inverse of the Gyrase concentrations (0, 0.1 and 0.2% Rhamnose induction), measured by qPCR, relative to the control condition (0% Rhamnose). Vertical error bars represent the standard error of the mean. In addition, shown are the best-fit lines and their standard errors of the mean (green and light blue areas), obtained by Monte Carlo simulations (500 iterations). Both lines (blue and green) were corrected for the effects of the RNAP increase in the RNA production rate when overexpressing Gyrase. z' stands for the ratio $\left(\frac{r}{r_C}\right)^{-1}$ after the correction. The blue circles are 0.008 units deviated to the right, for legibility. The line equations are $\left(\frac{r}{r_C}\right)^{-1} = (0.28 \pm 0.13) \times \left(\frac{[G]}{[G_C]}\right)^{-1} + (0.72 \pm 0.13)$ and for the green and blue lines, respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

fraction of time spent by the promoter in locked state (Fig. 6). Finally, the estimations of k_1 and k_{unlock} using the models agree with past estimations (respectively in [9,48]).

Overall, the results suggest that the weight of PSB on the effective RNA production rate of a gene depends not only on the mean activity of the DNA loop that the gene belongs to, but also on the basal transcription activity of the observed gene. This dependence was found to be sufficiently strong to require the introduction of this phenomenon in the model, if one is to predict the effects of changing Gyrase levels on the dynamics of transcription (reaction 4 in Fig. 1). This is because the fraction of time spent in locked states depends not only on the rate of accumulation of positive supercoils, but also on how much time Gyrases have to resolve enough supercoils (in between consecutive transcription events) to avoid reaching a supercoiling density that suffices for promoter locking.

Given that increasing the *basal* transcription rates enhances the influence of PSB on the *effective* transcription rate, we hypothesize that, at least in some genes, increasing the basal transcription rate may come at the cost of increased transcriptional noise due to PSB, even if lowering the noise from basal transcription dynamics. We thus expect that the relationship between basal transcription rate and PSB needs to be directly accounted for in models of prokaryotic gene expression. As such, when reducing the SN model (Supplementary Table S2) to a minimal model (Fig. 1), one of the critical components kept from the SN model was reaction 4 (Fig. 1), as it is responsible for the production positive supercoils at a rate that differs with the basal transcription rate of the gene interest.

This was required even though, similar to past models [4,9], there is also reaction 5, which introduces positive supercoils from 'external' sources, at a rate that differs with the average transcriptional activity of all genes in the same DNA loop (or topological domain) [17,43] and DNA replication [84]. Interestingly, the existence of this dependency suggests that it should be possible to, some extent, regulate the robustness of chromosome-integrated synthetic circuits to PSB, by tuning its own transcription rates, as well as placing it in a topological domain with desired mean activity.

In this regard, since increasing the basal transcription rate enhances the effects of PSB, is there an effective upper limit on the transcription rate? If so, this could potentially explain (at least partially) why some genes exist in multi-copy form. Such form would allow crossing this limit, while also supporting more stable expression levels.

Meanwhile, the combination of the results from two different constructs suggest that it may be possible to map a state space of transcription initiation locking of the topological domains of *E. coli*. However, since domain barriers are not likely to be at fixed sites [45,70,92], it may be necessary to set constructs in various regions of the DNA and measure not only the mean, but also the variability of the propensity for transcription locking as a function of DNA location. Using several constructs, differing in features (e.g. in regulatory mechanisms), should allow accounting for changes in parameters, other than the basal transcription rate. Namely, while here we mapped a 1-dimensional space by tuning the basal transcription rate, changing other variables would facilitate mapping a multi-dimensional state-space of transcription initiation locking kinetics. We expect such mapping to be of use in dissecting global transcription programs of *E. coli*, as well as for implementing chromosome-integrated synthetic circuits with predictable kinetics.

Our methodology may also assist in quantifying effects of environmental shifts (e.g. temperature) on the kinetics of transcription initiation locking. One could then explore whether *E. coli* uses this

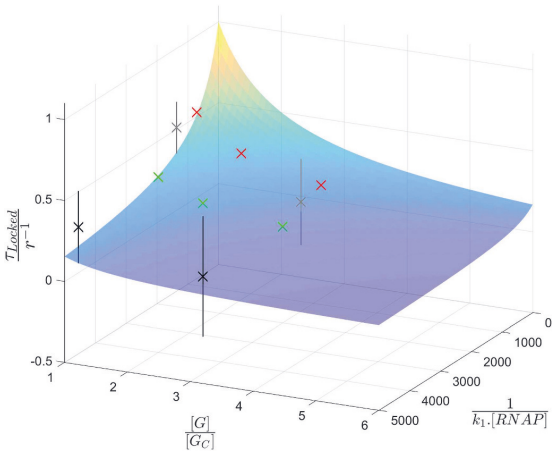

Fig. 6. Expected time in locked states relative to the expected interval between consecutive RNA production events as a function of Gyrase concentration and of the inverse of the basal transcription rate ($k_1{}^{-1}$), with $\tau_{active} = \frac{1}{k_1.RNAP}$. The surface is the model prediction of the relative τ_{Locked} as a function of τ_{active} and of the Gyrase concentration relative to the control. Red crosses are the empirical data for the $LacO_3O_1$ promoter under full induction (1000 μM IPTG), green crosses are the empirical data for the native Lac promoter under full induction (1000 μM IPTG), grey crosses are the empirical data for the $LacO_3O_1$ promoter under 50 μM IPTG induction, and black crosses are the empirical data for the $LacO_3O_1$ promoter uninduced (0 μM IPTG). The vertical bars are the standard error of the mean. All error bars intersect the surface. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

phenomenon to adapt to fluctuating environments. This hypothesis is supported by recent observations [64] that cold-shock genes have atypical supercoiling-sensitivity (for unknown reasons). I.e., genes with long-term responses to cold-shocks appear to be impervious to supercoiling, while genes with short-term responses have more-than-expected-by-chance sensitivity to supercoiling. Our methodology may assist in dissecting the responsible mechanisms, e.g. by measuring τ_{locked} and τ_{active} of these genes following mutations, etc.

We expect our methodology to be compatible with other techniques. E.g., it is potentially valuable to combine it with measurements of local DNA supercoiling density, such as trimethylpsoralen intercalation [40], to quantify the relationship between this density and the effects of PSB on transcription. Similarly, it may be valuable to combine it with the method in [48] to dissect the kinetics of rate limiting steps of active transcription initiation. For chromosome-integrated genes, we expect that only by using both methods will be possible to estimate the times spent prior to open complex formation, since models suggest that this state of activity will differ with the kinetics of promoter locking due to PSB [59], due to the expected competition between the formations of closed complexes and locked states.

Further, our methodology should be applicable using other techniques, such as RNA FISH (Fluorescence in situ hybridization) [79] and RNA aptamer-fluorogen systems [12,19,65,82,98].

Finally, our results derived from a first attempt at dissecting the *in vivo* dynamics of locking of transcription initiation using a Lineweaver-Burk plot. Many questions remain unanswered and require further study. It may turn out that fluctuations in Gyrase concentration have non-uniform effects at the genome-wide level, due to the dependency on basal transcription rates and mean rates of topological domains. Potentially, this could be used by cells as means to activate specific gene cohorts (e.g. of genes sharing the same topological domain), involved in responsive transcriptional programs. It could also be used to change the state of small genetic circuits responsible for triggering response programs to fluctuations in supercoiling density (e.g. fluctuations in supercoiling densities may alter the stable state of a, e.g., genetic switch with genes in different topological domains). If this holds true, the 'optimal' level of Gyrase may differ with the environment and/or internal cell state, depending on whether a given gene cohort (supercoiling density dependent) should be active or not.

In conclusion, the methods and results here presented are expected to support near-future research on the role of Gyrase on the global dynamics of gene regulatory networks.

Funding

This work was supported by the Finnish Academy of Science and Letters [to C.P.]; Pirkanmaa Regional Fund [to V.K.]; Tampere University Graduate Program (Finland) [to V.C. and M.B.]; EDUFI Fellowship [TM-19-11105 to S.D.]; Academy of Finland [295027 to A.S.R.]; and Jane and Aatos Erkko Foundation [610536 to A.S.R.]. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Author's statement

C.P. and A.S.R. conceived the study. C.P. performed data analysis. C.P., M.B., and A.S.R. performed modelling. V.K.K., M.M., V.C., and S.D. performed measurements. C.P., V.K., and A.S.R. drafted the manuscript, which was revised by all authors. The authors declare no competing interests.

Transparency document

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

Declaration of competing interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbagrm.2020.194515.

References

[1] R.E. Ashley, A. Dittmore, S.A. McPherson, C.L. Turnbough, K.C. Neuman, N. Osheroff, Activities of gyrase and topoisomerase IV on positively supercoiled DNA, Nucleic Acids Res. 45 (16) (2017) 9611–9624, https://doi.org/10.1093/nar/gkx649.
[2] N. Blot, R. Mavathur, M. Geertz, A. Travers, G. Muskhelishvili, Homeostatic regulation of supercoiling sensitivity coordinates transcription of the bacterial genome, EMBO Rep. 7 (7) (2006) 710–715, https://doi.org/10.1038/sj.embor.7400729.
[3] J.D. Boeke, P. Model, A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane, Proc. Natl. Acad. Sci. U. S. A. 79 (17) (1982) 5200, https://doi.org/10.1073/pnas.79.17.5200.
[4] C.H. Bohrer, E. Roberts, A biophysical model of supercoiling dependent transcription predicts a structural aspect to gene regulation, BMC Biophys. 9 (1) (2015) 2, https://doi.org/10.1186/s13628-016-0027-0.
[5] G. Casella, R.L. Berger, Statistical Inference, Thomson Learning, 2002.
[6] J.J. Champoux, DNA topoisomerases: structure, function, and mechanism, Annu. Rev. Biochem. 70 (1) (2001) 369–413, https://doi.org/10.1146/annurev.biochem.70.1.369.
[7] B. Cheng, C.-X. Zhu, C. Ji, A. Ahumada, Y.-C. Tse-Dinh, Direct interaction between *Escherichia coli* RNA polymerase and the zinc ribbon domains of DNA topoisomerase I, J. Biol. Chem. 278 (33) (2003) 30705–30710, https://doi.org/10.1074/jbc.M303403200.
[8] B.-K. Cho, D. Kim, E.M. Knight, K. Zengler, B.O. Palsson, Genome-scale reconstruction of the sigma factor network in *Escherichia coli*: topology and functional states, BMC Biol. 12 (1) (2014) 4, https://doi.org/10.1186/1741-7007-12-4.
[9] S. Chong, C. Chen, H. Ge, X.S. Xie, Mechanism of transcriptional bursting in Bacteria, Cell 158 (2) (2014) 314–326, https://doi.org/10.1016/j.cell.2014.05.038.
[10] K.A. Datsenko, B.L. Wanner, One-step inactivation of chromosomal genes in

Escherichia coli K-12 using PCR products, Proc. Natl. Acad. Sci. 97 (12) (2000) 6640–6645, https://doi.org/10.1073/pnas.120163297.
[11] S. Deng, R.A. Stein, N.P. Higgins, Organization of supercoil domains and their reorganization by transcription, Mol. Microbiol. 57 (6) (2005) 1511–1521, https://doi.org/10.1111/j.1365-2958.2005.04796.x.
[12] E.V. Dolgosheina, S.C.Y. Jeng, S.S.S. Panchapakesan, R. Cojocaru, P.S.K. Chen, P.D. Wilson, N. Hawkins, P.A. Wiggins, P.J. Unrau, RNA mango aptamer-fluorophore: a bright, high-affinity complex for RNA labeling and tracking, ACS Chem. Biol. 9 (10) (2014) 2412–2420, https://doi.org/10.1021/cb500499x.
[13] T. Dong, H.E. Schellhorn, Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933, BMC Genomics 10 (1) (2009) 349, https://doi.org/10.1186/1471-2164-10-349.
[14] C.J. Dorman, M.J. Dorman, DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression, Biophys. Rev. 8 (S1) (2016) 89–100, https://doi.org/10.1007/s12551-016-0238-2.
[15] K. Drlica, Control of bacterial DNA supercoiling, Mol. Microbiol. 6 (4) (1992) 425–433, https://doi.org/10.1111/j.1365-2958.1992.tb01486.x.
[16] M. Drolet, Growth inhibition mediated by excess negative supercoiling: the interplay between transcription elongation, R-loop formation and DNA topology, Mol. Microbiol. 59 (2006) 723–730, https://doi.org/10.1111/j.1365-2958.2005.05006.x.
[17] D. El Hanafi, L. Bossi, Activation and silencing of leu-500 promoter by transcription-induced DNA supercoiling in the Salmonella chromosome, Mol. Microbiol. 37 (3) (2000) 583–594, https://doi.org/10.1046/j.1365-2958.2000.02015.x.
[18] A. Farewell, K. Kvint, T. Nyström, Negative regulation by RpoS: a case of sigma factor competition, Mol. Microbiol. 29 (4) (1998) 1039–1051, https://doi.org/10.1046/j.1365-2958.1998.00990.x.
[19] G.S. Filonov, J.D. Moon, N. Svensen, S.R. Jaffrey, Broccoli: rapid selection of an RNA mimic of green fluorescent protein by fluorescence-based selection and directed evolution, J. Am. Chem. Soc. 136 (46) (2014) 16299–16308, https://doi.org/10.1021/ja508478x.
[20] K. Fujita, M. Iwaki, T. Yanagida, Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA, Nat. Commun. 7 (2016) 13788, , https://doi.org/10.1038/ncomms13788.
[21] G. Fulcrand, S. Dages, X. Zhi, P. Chapagain, B.S. Gerstman, D. Dunlap, F. Leng, DNA supercoiling, a critical signal regulating the basal expression of the lac operon in *Escherichia coli*, Sci. Rep. 6 (1) (2016) 19243, , https://doi.org/10.1038/srep19243.
[22] M. Gellert, M.H. O'Dea, T. Itoh, J. Tomizawa, Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase, Proc. Natl. Acad. Sci. U. S. A. 73 (12) (1976) 4474–4478, https://doi.org/10.1073/pnas.73.12.4474.
[23] D.T. Gillespie, A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (4) (1976) 403–434, https://doi.org/10.1016/0021-9991(76)90041-3.
[24] D.T. Gillespie, Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (1977) 2340–2361, https://doi.org/10.1021/j100540a008.
[25] I. Golding, E.C. Cox, RNA dynamics in live *Escherichia coli* cells, Proc. Natl. Acad. Sci. U. S. A. 101 (31) (2004) 11310–11315, https://doi.org/10.1073/pnas.0404443101.
[26] I. Golding, J. Paulsson, S.M. Zawilski, E.C. Cox, Real-time kinetics of gene activity in individual bacteria, Cell 123 (6) (2005) 1025–1036, https://doi.org/10.1016/j.cell.2005.09.031.
[27] N.S.M. Goncalves, S.M.D. Oliveira, V. Kandavalli, J.M. Fonseca, A.S. Ribeiro, Temperature dependence of leakiness of transcription repression mechanisms of *E. coli*, Lect. Notes Comput. Sci 9859 (2016) 341–342.
[28] P. Guptasarma, Cooperative relaxation of supercoils and periodic transcriptional initiation within polymerase batteries, BioEssays 18 (4) (1996) 325–332, https://doi.org/10.1002/bies.950180411.
[29] A. Hakkinen, A.-B. Muthukrishnan, A. Mora, J.M. Fonseca, A.S. Ribeiro, CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*, Bioinformatics 29 (13) (2013) 1708–1709, https://doi.org/10.1093/bioinformatics/btt194.
[30] A. Häkkinen, A.S. Ribeiro, Characterizing rate limiting steps in transcription from RNA production times in live cells, Bioinformatics 32 (9) (2016) 1346–1352, https://doi.org/10.1093/bioinformatics/btv744.
[31] C.D. Hardy, N.R. Cozzarelli, Alteration of *Escherichia coli* topoisomerase IV to novobiocin resistance. *Antimicrob*, Agents Chemother. 47 (3) (2003) 941–947, https://doi.org/10.1128/aac.47.3.941-947.2003.
[32] Y. Hayakawa, T. Murotsu, K. Matsubara, Mini-F protein that binds to a unique region for partition of mini-F plasmid DNA, J. Bacteriol. 163 (1) (1985) 349–354.
[33] N.P. Higgins, Species-specific supercoil dynamics of the bacterial nucleoid, Biophys. Rev. 8 (S1) (2016) 113–121, https://doi.org/10.1007/s12551-016-0207-9.
[34] V.K. Kandavalli, H. Tran, A.S. Ribeiro, Effects of σ factor competition are promoter initiation kinetics dependent, Biochim. Biophys. Acta - Gene Regul. Mech. 1859 (10) (2016) 1281–1288, https://doi.org/10.1016/j.bbagrm.2016.07.011.
[35] K. Kirkegaard, J.C. Wang, Bacterial DNA topoisomerase I can relax positively supercoiled DNA containing a single-stranded loop, J. Mol. Biol. 185 (3) (1985) 625–637, https://doi.org/10.1016/0022-2836(85)90075-0.
[36] N. Kleckner, J.K. Fisher, M. Stouf, M.A. White, D. Bates, G. Witz, The bacterial nucleoid: nature, dynamics and sister segregation, Curr. Opin. Microbiol. 22 (2014) 127–137, https://doi.org/10.1016/j.mib.2014.10.001.
[37] M.V. Kotlajich, D.R. Hron, B.A. Boudreau, Z. Sun, Y.L. Lyubchenko, R. Landick, Bridged filaments of histone-like nucleoid structuring protein pause RNA polymerase and aid termination in bacteria, Elife 4 (2015), https://doi.org/10.7554/eLife.04970.
[38] F. Kouzine, S. Sanford, Z. Elisha-Feil, D. Levens, The functional response of upstream DNA to dynamic supercoiling *in vivo*, Nat. Struct. Mol. Biol. 15 (2) (2008) 146–154, https://doi.org/10.1038/nsmb.1372.
[39] M. Krystek, M. Anton, A weighted total least-squares algorithm for fitting a straight line, Meas. Sci. Technol. 18 (11) (2007) 3438–3442, https://doi.org/10.1088/0957-0233/18/11/025.
[40] A. Lal, A. Dhar, A. Trostel, F. Kouzine, A.S.N. Seshasayee, S. Adhya, Genome scale patterns of supercoiling in a bacterial chromosome, Nat. Commun. 7 (1) (2016) 11055, , https://doi.org/10.1038/ncomms11055.
[41] T.B.K. Le, M.V. Imakaev, L.A. Mirny, M.T. Laub, High-resolution mapping of the spatial organization of a bacterial chromosome, Science 342 (6159) (2013) 731–734, https://doi.org/10.1126/science.1242059.
[42] F. Leng, B. Chen, D.D. Dunlap, Dividing a supercoiled DNA molecule into two independent topological domains, Proc. Natl. Acad. Sci. 108 (50) (2011) 19973–19978, https://doi.org/10.1073/pnas.1109854108.
[43] D.M. Lilley, C.F. Higgins, Local DNA topology and gene expression: the case of the leu-500 promoter, Mol. Microbiol. 5 (4) (1991) 779–783, https://doi.org/10.1111/j.1365-2958.1991.tb00749.x.
[44] H. Lineweaver, D. Burk, The determination of enzyme dissociation constants, J. Am. Chem. Soc. 56 (3) (1934) 658–666, https://doi.org/10.1021/ja01318a036.
[45] V.S. Lioy, A. Cournac, M. Marbouty, S. Duigou, J. Mozziconacci, O. Espéli, F. Boccard, R. Koszul, Multiscale structuring of the *E. coli* chromosome by nucleoid-associated and condensin proteins, Cell 172 (4) (2018) 771–783.e18, https://doi.org/10.1016/j.cell.2017.12.027.
[46] L.F. Liu, J.C. Wang, Supercoiling of the DNA template during transcription, Proc. Natl. Acad. Sci. U. S. A. 84 (20) (1987) 7024–7027, https://doi.org/10.1073/pnas.84.20.7024.
[47] Y. Liu, A.M. Berrido, Z.-C. Hua, Y.-C. Tse-Dinh, F. Leng, Biochemical and biophysical properties of positively supercoiled DNA, Biophys. Chem. 230 (2017) 68–73, https://doi.org/10.1016/j.bpc.2017.08.008.
[48] J. Lloyd-Price, S. Startceva, V. Kandavalli, J.G. Chandraseelan, N. Goncalves, S.M.D. Oliveira, A. Häkkinen, A.S. Ribeiro, Dissecting the stochastic transcription initiation process in live *Escherichia coli*, DNA Res. 23 (3) (2016) 203–214, https://doi.org/10.1093/dnares/dsw009.
[49] A.S. Lynch, J.C. Wang, Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: a mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I, J. Bacteriol. 175 (6) (1993) 1645, https://doi.org/10.1128/jb.175.6.1645-1655.1993.
[50] D. Ma, D.N. Cook, M. Alberti, N.G. Pon, H. Nikaido, J.E. Hearst, Genes acrA and acrB encode a stress-induced efflux system of *Escherichia coli*, Mol. Microbiol. 16 (1) (1995) 45–55, https://doi.org/10.1111/j.1365-2958.1995.tb02390.x.
[51] J. Ma, L. Bai, M.D. Wang, Transcription under torsion, Science 340 (6140) (2013) 1580–1583, https://doi.org/10.1126/science.1235441.
[52] J. Ma, M.D. Wang, DNA supercoiling during transcription, Biophys. Rev. 8 (Suppl. 1) (2016) 75–87, https://doi.org/10.1007/s12551-016-0215-9.
[53] J. Mäkelä, J. Lloyd-Price, O. Yli-Harja, A.S. Ribeiro, Stochastic sequence-level model of coupled transcription and translation in prokaryotes, BMC Bioinformatics 12 (1) (2011) 121, https://doi.org/10.1186/1471-2105-12-121.
[54] J. Mäkelä, V. Kandavalli, A.S. Ribeiro, Rate-limiting steps in transcription dictate sensitivity to variability in cellular components, Sci. Rep. 7 (1) (2017) 10588, , https://doi.org/10.1038/s41598-017-11257-2.
[55] H. Mannerstrom, O. Yli-Harja, A.S. Ribeiro, Inference of kinetic parameters of delayed stochastic models of gene expression using a markov chain approximation, EURASIP J. Bioinform. Syst. Biol. 2011 (1) (2011) 572876, , https://doi.org/10.1155/2011/572876.
[56] L. Martins, R. Neeli-Venkata, S.M.D. Oliveira, A. Häkkinen, A.S. Ribeiro, J.M. Fonseca, SCIP: a single-cell image processor toolbox, Bioinformatics (Oxford, England) 34 (24) (2018) 4318–4320, https://doi.org/10.1093/bioinformatics/bty505.
[57] M. Matsumoto, T. Nishimura, Mersenne Twister, ACM Trans. on Modeling and Comp. Simulation 8 (1998) 3–30, https://doi.org/10.1145/272991.272995.
[58] W.R. McClure, Mechanism and control of transcription initiation in prokaryotes, Annu. Rev. Biochem. 54 (1) (1985) 171–204, https://doi.org/10.1146/annurev.bi.54.070185.001131.
[59] N. Mitarai, I.B. Dodd, M.T. Crooks, K. Sneppen, The generation of promoter-mediated transcriptional noise in Bacteria, PLoS Comput. Biol. 4 (7) (2008) e1000109, , https://doi.org/10.1371/journal.pcbi.1000109.
[60] H. Mori, A. Kondo, A. Ohshima, T. Ogura, S. Hiraga, Structure and function of the F plasmid genes essential for partitioning, J. Mol. Biol. 192 (1) (1986) 1–15, https://doi.org/10.1016/0022-2836(86)90459-6.
[61] K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, O. Amster-Choder, Translation-independent localization of mRNA in *E. coli*, Science 331 (6020) (2011) 1081–1084, https://doi.org/10.1126/science.1195691.
[62] S. Oehler, M. Amouyal, P. Kolkhof, B. von Wilcken-Bergmann, B. Müller-Hill, Quality and position of the three lac operators of *E. coli* define efficiency of repression, EMBO J. 13 (14) (1994) 3348–3355.
[63] S.M.D. Oliveira, R. Neeli-Venkata, N.S.M. Goncalves, J.A. Santinha, L. Martins, H. Tran, J. Mäkelä, A. Gupta, M. Barandas, A. Häkkinen, J. Lloyd-Price, J.M. Fonseca, A.S. Ribeiro, Increased cytoplasm viscosity hampers aggregate polar segregation in Escherichia coli, Mol. Microbiol. 99 (4) (2016) 686–699, https://doi.org/10.1111/mmi.13257.
[64] S.M.D. Oliveira, N.S.M. Goncalves, V.K. Kandavalli, L. Martins, R. Neeli-Venkata, J. Reyelt, J.M. Fonseca, J. Lloyd-Price, H. Kranz, A.S. Ribeiro, Chromosome and plasmid-borne PLacO3O1 promoters differ in sensitivity to critically low temperatures, Sci. Rep. 9 (1) (2019) 4486, https://doi.org/10.1038/s41598-019-39618-z.
[65] J.S. Paige, K.Y. Wu, S.R. Jaffrey, RNA mimics of green fluorescent protein, Science 333 (6042) (2011) 642–646, https://doi.org/10.1126/science.1207339.

[66] C.S.D. Palma, S. Startceva, R. Neeli-Venkata, M. Zare, N.S.M. Goncalves, J.M. Fonseca, S.M.D. Oliveira, A.S. Ribeiro, A strategy for dissecting the kinetics of transcription repression mechanisms, Proceedings of the European Medical and Biolgical Engineering Conference (EMBEC), June 11–15, Tampere, Finland, 65 Springer, Singapore, 2017, pp. 1097–1100, , https://doi.org/10.1007/978-981-10-5122-7_274 Also published in: IFMBE Proceedings.
[67] D.S. Peabody, The RNA binding site of bacteriophage MS2 coat protein, EMBO J. 12 (2) (1993) 595.
[68] D.S. Peabody, Role of the coat protein-RNA interaction in the life cycle of bacteriophage MS2, Mol. Gen. Genet. MGG 254 (4) (1997) 358–364, https://doi.org/10.1007/s004380050427.
[69] B.J. Peter, J. Arsuaga, A.M. Breier, A.B. Khodursky, P.O. Brown, N.R. Cozzarelli, Genomic transcriptional response to loss of chromosomal supercoiling in *Escherichia coli*, Genome Biol. 5 (11) (2004) R87, https://doi.org/10.1186/gb-2004-5-11-r87.
[70] L. Postow, C.D. Hardy, J. Arsuaga, N.R. Cozzarelli, Topological domain structure of the *Escherichia coli* chromosome, Genes Dev. 18 (14) (2004) 1766–1779, https://doi.org/10.1101/gad.1207504.
[71] G.J. Pruss, K. Drlica, Topoisomerase I mutants: the gene on pBR322 that encodes resistance to tetracycline affects plasmid DNA supercoiling, Proc. Natl. Acad. Sci. U. S. A. 83 (23) (1986) 8952–8956, https://doi.org/10.1073/pnas.83.23.8952.
[72] T. Rajala, A. Häkkinen, S. Healy, O. Yli-Harja, A.S. Ribeiro, Effects of transcriptional pausing on gene expression dynamics, PLoS Comput. Biol. 6 (3) (2010) 29–30, https://doi.org/10.1371/journal.pcbi.1000704.
[73] A. Revyakin, R.H. Ebright, T.R. Strick, Promoter unwinding and promoter clearance by RNA polymerase: detection by single-molecule DNA nanomanipulation, Proc. Natl. Acad. Sci. 101 (14) (2004) 4776–4780, https://doi.org/10.1073/pnas.0307241101.
[74] A.S. Ribeiro, J. Lloyd-Price, SGN Sim, a stochastic genetic networks simulator, Bioinformatics 23 (6) (2007) 777–779, https://doi.org/10.1093/bioinformatics/btm004.
[75] A.S. Ribeiro, O.-P. Smolander, T. Rajala, A. Häkkinen, O. Yli-Harja, Delayed stochastic model of transcription at the single nucleotide level, J. Comput. Biol. 16 (4) (2009) 539–553, https://doi.org/10.1089/cmb.2008.0153.
[76] P.E. Rouvière, A. De Las Peñas, J. Mecsas, C.Z. Lu, K.E. Rudd, C.A. Gross, rpoE, the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*, EMBO J. 14 (5) (1995) 1032–1042.
[77] N. Rovinskiy, A.A. Agbleke, O. Chesnokova, Z. Pang, N.P. Higgins, Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome, PLoS Genet. 8 (8) (2012) e1002845, , https://doi.org/10.1371/journal.pgen.1002845.
[78] J. Santinha, L. Martins, A. Häkkinen, J. Lloyd-Price, S.M.D. Oliveira, A. Gupta, T. Annila, A. Mora, A.S. Ribeiro, J.R. Fonseca, iCellFusion: Tool for Fusion and Analysis of Live-Cell Images from Time-Lapse Multimodal Microscopy, (2016), https://doi.org/10.4018/978-1-4666-8811-7.ch004.
[79] R.H. Singer, D.C. Ward, Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog, Proc. Natl. Acad. Sci. U. S. A. 79 (23) (1982) 7331–7335, https://doi.org/10.1073/pnas.79.23.7331.
[80] C. Sissi, M. Palumbo, In front of and behind the replication fork: bacterial type IIA topoisomerases, Cell. Mol. Life Sci. 67 (12) (2010) 2001–2024, https://doi.org/10.1007/s00018-010-0299-5.
[81] S. Startceva, V.K. Kandavalli, A. Visa, A.S. Ribeiro, Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression, Biochim. Biophys. Acta - Gene Regul. Mech. 1862 (2) (2019) 119–128, https://doi.org/10.1016/j.bbagrm.2018.12.005.
[82] R.L. Strack, M.D. Disney, S.R. Jaffrey, A superfolding Spinach2 reveals the dynamic nature of trinucleotide repeat–containing RNA, Nat. Methods 10 (12) (2013) 1219–1224, https://doi.org/10.1038/nmeth.2701.
[83] M. Stracy, C. Lesterlin, F. Garza de Leon, S. Uphoff, P. Zawadzki, A.N. Kapanidis, Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid, Proc. Natl. Acad. Sci. 112 (32) (2015) E4390–E4399, https://doi.org/10.1073/pnas.1507592112.
[84] M. Stracy, A.J.M. Wollman, E. Kaja, J. Gapinski, J.-E. Lee, V.A. Leek, S.J. McKie, L.A. Mitchenall, A. Maxwell, D.J. Sherratt, M.C. Leake, P. Zawadzki, Single-molecule imaging of DNA gyrase activity in living *Escherichia coli*, Nucleic Acids Res. 47 (1) (2019) 210–220, https://doi.org/10.1093/nar/gky1143.
[85] Y. Taniguchi, P.J. Choi, G.W. Li, H. Chen, M. Babu, J. Hearn, A. Emili, X.S. Xie, Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells, Sci. (New York, NY) 329 (5991) (2010) 533–538, https://doi.org/10.1126/science.1188308.
[86] H. Tran, S.M.D. Oliveira, N. Goncalves, A.S. Ribeiro, Kinetics of the cellular intake of a gene expression inducer at high concentrations, Mol. BioSyst. 11 (9) (2015) 2579–2587, https://doi.org/10.1039/C5MB00244C.
[87] Y.P. Tsao, H.Y. Wu, L.F. Liu, Transcription-driven supercoiling of DNA: direct biochemical evidence from in vitro studies, Cell 56 (1) (1989) 111–118, https://doi.org/10.1016/0092-8674(89)90989-6.
[88] T. Větrovský, P. Baldrian, The variability of the 16S rRNA gene in bacterial genomes and its consequences for bacterial community analyses, PLoS One 8 (2) (2013) e57923, , https://doi.org/10.1371/journal.pone.0057923.
[89] S.M. Vos, E.M. Tretter, B.H. Schmidt, J.M. Berger, All tangled up: how cells direct, manage and exploit topoisomerase function, Nat. Rev. Mol. Cell Biol. 12 (12) (2011) 827–841, https://doi.org/10.1038/nrm3228.
[90] J.C. Wang, DNA topoisomerases, Annu. Rev. Biochem. 54 (1) (1985) 665–697, https://doi.org/10.1146/annurev.bi.54.070185.003313.
[91] James C. Wang, DNA Topoisomerases, Annu. Rev. Biochem. 65 (1) (1996) 635–692, https://doi.org/10.1146/annurev.bi.65.070196.003223.
[92] X. Wang, P.M. Llopis, D.Z. Rudner, Organization and segregation of bacterial chromosomes, Nat. Rev. Genet. 14 (3) (2013) 191–203, https://doi.org/10.1038/nrg3375.
[93] A. Wegerer, T. Sun, J. Altenbuchner, Optimization of an *E. coli* L-rhamnose-inducible expression vector: test of various genetic module combinations, BMC Biotechnol. 8 (1) (2008) 2, https://doi.org/10.1186/1472-6750-8-2.
[94] A. Wheeler, Digital Microscopy, in: A. Wheeler, R. Henriques (Eds.), Standard and Super-Resolution Bioimaging Data Analysis, 2017, https://doi.org/10.1002/9781119096948.ch1.
[95] H.Y. Wu, S.H. Shyy, J.C. Wang, L.F. Liu, Transcription generates positively and negatively supercoiled domains in the template, Cell 53 (3) (1988) 433–440, https://doi.org/10.1016/0092-8674(88)90163-8.
[96] E. Yeung, A.J. Dy, K.B. Martin, A.H. Ng, D. Del Vecchio, J.L. Beck, J.J. Collins, R.M. Murray, Biophysical constraints arising from compositional context in synthetic gene networks, Cell Syst 5 (1) (2017) 11–24.e12, https://doi.org/10.1016/j.cels.2017.06.001.
[97] E.L. Zechiedrich, A.B. Khodursky, S. Bachellier, R. Schneider, D. Chen, D.M.J. Lilley, N.R. Cozzarelli, Roles of topoisomerases in maintaining steady-state DNA supercoiling in Escherichia coli, J. Biol. Chem. 275 (11) (2000) 8103–8113, https://doi.org/10.1074/jbc.275.11.8103.
[98] J. Zhang, J. Fei, B.J. Leslie, K.Y. Han, T.E. Kuhlman, T. Ha, Tandem spinach array for mRNA imaging in living bacterial cells, Sci. Rep. 5 (1) (2015) 17295, , https://doi.org/10.1038/srep17295.
[99] R. Samul, F. Leng, Transcription-coupled Hypernegative Supercoiling of Plasmid DNA by T7 RNA Polymerase in Escherichia coli Topoisomerase I-Deficient Strains, Journal of Molecular Biology 374 (4) (2007) 925–935, https://doi.org/10.1016/j.jmb.2007.10.011.

Supplementary Material

Dissecting the *in vivo* dynamics of transcription locking due to positive supercoiling buildup

Cristina S.D. Palma[1], Vinodh Kandavalli[1], Mohamed N.M. Bahrudeen[1], Marco Minoia[1], Vatsala Chauhan[1], Suchintak Dash[1], and Andre S. Ribeiro[1*]

[1] Laboratory of Biosystem Dynamics, BioMediTech, Faculty of Medicine and Health Technology, Tampere University, Tampere University, 33101 Tampere, Finland.

* To whom correspondence should be addressed. Email: andre.sanchesribeiro@tuni.fi

Present Address: Andre S. Ribeiro, BioMediTech Institute, Tampere University, Arvo Ylpön katu 34, P.O.Box 100, 33014 Tampere, Finland.

SUPPLEMENTARY METHODS

I. RNA quantification from fluorescent spots, spots lifetime, and spots full tagging time

Integer-valued number of MS2-GFP-tagged mRNA molecules in individual cells are obtained from microscopy images as in e.g. (Häkkinen et al., 2013; Oliveira et al., 2016). Shortly, MS2-GFP tagged RNA spots are segmented by Kernel Density Estimation (KDE). Example Figures S2A and S2E show cellular backgrounds generated by unbound MS2-GFP proteins in cells carrying the plasmid-borne and the chromosome integrated target genes, respectively. Meanwhile, Figures S2B and S2F show these cells when with tagged RNAs along with the results of the spot detection methods (Häkkinen and Ribeiro, 2015). These spots are visible to the Human eye and, as seen, detectable by image analysis (Santinha et al., 2016), since their fluorescence is much higher than in near-neighbour pixels (Figures S2D and S2H). In addition, the variability in fluorescence intensity of pixels without spots is much smaller than the mean difference in intensity between pixels with and without spots (Figures S2D and S2H), which lowers the risk of detecting 'false' spots and removing 'true' spots. Consequently, the background fluorescence intensity (average over all pixels not containing 'RNA-spots') can safely be subtracted from the intensity of each fluorescent RNA-spot.

From the resulting RNA-spot fluorescence intensities in arbitrary units (a.u.), we estimate the intensity of individual MS2-GFP tagged RNAs as in (Golding et al., 2005). From histograms of intensities of RNA-spots, we find the intensity of the first "peak" of the histogram (which should correspond to the intensity of one tagged RNA). Next, for each spot, we round its intensity value to the nearest integer, to obtain its integer-valued number of RNA molecules.

We also considered the possibility that some spots correspond to incomplete RNAS due to e.g. arrests during transcription elongation. This could lead to overestimation of RNA numbers. To determine whether this could occur, let us assume, as an example, that incomplete RNAs have, on average, 50% the total intensity of a completely tagged RNA. From the histogram of spots intensities for RNAs produced by the plasmid-borne gene, we estimated than only 1.6 % of the detected spots have less than 50% of the mean intensity of 1 RNA. Similar values were obtained for the

chromosome-integrated gene. This implies that small, forming RNAs introduce little to no error in RNA counting. Further, not all 'weak spots' will be forming RNAs (e.g. a few could be out-of-focus tagged RNAs). Thus, the value of 1.6 % could be considered to be an upper bound for such fraction of forming RNAs that are erroneously counted as fully formed RNAs. From this, we conclude that little to no error is added to the RNA numbers per cell due to this.

As in (Tran et al., 2015), we measured MS2-GFP tagged RNAs decrease in fluorescence intensity during time-lapse microscopy (Figure S9). We then estimated the mean half-life of tagged RNAs by fitting the intensity of each MS2-GFP tagged RNA over time with a decaying exponential function. We found the mean half-time of spots fluorescence to be longer than the measurement period (>150 min.), as in (Häkkinen and Ribeiro, 2015; Tran et al., 2015). In addition, 'bleaching' of MS2 tagged RNAs was not observed. Finally, we observed that target RNA molecules become fully tagged by MS2-GFP in <1 minute (Tran et al., 2015). As such, tagging times are not considered as influencing RNA counting.

To determine if MS2-GFP proteins formed clusters in the absence of target RNA, we analyzed cells with the reporter system (responsible for producing MS2-GFP proteins) but lacking the target system (coding for target RNAs). In these cells, the number of 'fake' spots detected by the image analysis algorithm was approximately 100 times smaller than in cells with a target system. As such, the influence of 'fake spots' or abnormal MS2-GFP clusters is considered to be negligible. Finally, visual inspection of the images showed that all fake spots were due to failures in the image analysis, rather than the presence of MS2-GFP clusters visible to the Human eye.

II. Single-nucleotide model of transcription subject to the effects of PSB

We use a stochastic model of transcription with stepwise elongation at the single nucleotide level based on past models (Rajala et al., 2010; Ribeiro et al., 2009; Mäkelä et al., 2011). All reactions are described in Supplementary Table S2. Parameter values are extracted or derived from empirical data. The main improvement of the model, compared to past similar models, is the introduction of a dynamic PSB phenomenon at the single-nucleotide level.

Positive supercoiling is generated by the RNAP activity on both the neighbour genes (Kouzine et al., 2013; Naughton et al., 2013; Teves et al., 2014; Lilley et al., 1991; Rhee et al., 1999) as well as on the gene of interest (Chong et al., 2014). We assume a constant (stochastic) rate of accumulation of positive supercoils due to the activity of neighbour genes. This rate is expected to differ with the neighbours' location (e.g. whether they are, or not, in the same transcriptional unit), distance from the gene of interest, direction of transcription activity, and external factors, such as environmental perturbations (Weinstein-Fisher et al., 2000; Cheung et al., 2003).

The effects of PSB in the model are: i) elongation arrests at the nucleotide level (resulting in short pauses, which slowdown elongation); and ii) transcription initiation locking, which causes longer transcription activity breakdowns whose resolution requires Gyrase intervention (Chong et al., 2014). The propensities of these events are dynamic, in that they differ with the global level of PSB in the region of the DNA where the gene of interest is located (Ma et al., 2014 and Chong et al., 2014).

To model this process at the single nucleotide level, we first introduce a reaction for transcription initiation, where a promoter is found by an RNAP (reaction S2.1), followed by promoter escape (reaction S2.5), which initiates stepwise elongation (reaction S2.6). As soon as the promoter becomes unoccupied, a new transcription initiation event can occur.

As the RNAP percolates the DNA, following each elongation step from one nucleotide to the next (reaction S2.6), an activation step (reaction S2.7) needs to occur for the RNAP to further progress to the subsequent nucleotide. However, the following events compete with activation: pausing (reaction S2.8), arrest (reactions S2.11), editing (reaction S2.12), premature termination (reaction S2.14) and pyrophosphorolysis (reaction S2.15). All these events, except premature termination, are modelled as reversible, due to the ability of the transcription machinery to resolve them. Finally, pyrophosphorolysis results in moving one step backwards, implying that it does not require resolution.

The model also allows for pauses and pause escapes to occur due to collisions between RNAPs (reactions S2.9 and S2.10, respectively). Further, misincorporation can occur at the end of the transcription process (reaction S2.13). Provided no misincorporation or premature termination, elongation is completed (reaction S2.16) and an RNA is produced and the RNAP is released (reaction S2.16).

In addition to all these events, we also model a dynamic process of accumulation of positive supercoils, which has a direct impact on RNA production (Travers et al., 2005; Lesne et al., 2018). Specifically, PSB causes short arrests to the moving RNAP (accounted for in reaction S2.11), which increase in frequency with increasing PSB (Ma et al., 2013; Fujita et al., 2016). Overall, this progressively decreases the rate of elongation. In addition, for high enough PSB, it can halt transcription initiation (accounted for in reaction S2.2) (Ma et al., 2013; Chong et al., 2014). Further, supercoils are not static, i.e., they can diffuse through the DNA. In some cases, they can reach regions located thousands of base pairs away from the point of origin. Evidence for this include, e.g., the observation of “topological promoter coupling” (Kouzine et al., 2013; Naughton et al., 2013; Teves et al., 2014; Lilley et al., 1991; Rhee et al., 1999), when supercoils produced in the activity of one gene reach the transcription start site of another gene.

Given the above, in the model, positive supercoils can accumulate from two sources: i) RNAP activity on neighbour genes (reaction S2.4) and, ii) RNAP activity on the gene of interest (reaction S2.6) via the production of positive supercoils (SC^+). The number of such SC^+ units allows quantifying the level of PSB in the region of the gene of interest at any given moment. In detail, the RNAP needs to percolate ~ 10 nucleotides for one positive supercoil to accumulate (Stracy et al., 2019; Rovinskiy et al., 2012). This is implemented in reaction S2.6, where the creation of a SC^+ requiring the percolation by the RNAP of 10 nucleotides (i.e. only in 1 out of 10 nucleotides will a supercoil be created). Finally, for simplicity, the model does not record the location of positive supercoils, only their total amount in the region of the gene of interest.

Also modelled is the process of SC^+ removal by the direct action of Gyrases (reaction S2.17) (Gellert et al., 1976; Chong et al., 2014; Stracy et al., 2019).

As a side note, it is physically possible for transcription initiation to halt because the RNAP becomes unable to bind to the promoter (Mitarai et al.,2008), as well as because the RNAP becomes

unable to unwind the promoter once bound (Revyakin et al., 2014). For simplicity, we model only the former phenomenon. Further, it is noted that *E. coli* has mechanisms to handle the effects of PSB in the kinetics of elongation other than the intervention of Gyrases. For example, GreB allows the RNAP to transcribe more efficiently through supercoiled regions of the DNA, by limiting backtracking (Ma et al., 2019). The detailed phenomena are not explicitly modelled here but, in most cases, their effects are indirectly accounted for in the rates of RNAP arresting, etc.

Finally, several phenomena may be more PSB-dependent than currently represented in the model (e.g. pausing, pyrophosphorolysis, etc. (Ma et al., 2019)). Due to the present lack of knowledge of the quantitative relationship between PBS levels and the rates of these events, we opted for modelling them as independent phenomena, using the currently available empirical parameter values in optimal growth conditions. Nevertheless, as noted, events in elongation (aside from misincorporation and pyrophosphorolysis) are not expected to affect the mean RNA production rate.

III. Dynamics of the single-nucleotide model of transcription when subject to the effects of PSB

III.1 Dynamics of stepwise transcription elongation

We first test whether the model in Supplementary Table S2 can mimic the effects of PSB in the kinetics of transcription elongation. For this, we performed simulations at various relative Gyrase concentrations ([G]/[G_C]), with [G_C] being the concentration of Gyrases in the control condition. From these simulations, we extracted the time-length of multiple stepwise transcription elongation events ($\Delta t_{elongation}$). Results in Figure S6A show that as [G] increases, both the mean and the standard error of the mean of $\Delta t_{elongation}$ decrease. This can be explained by the results in Figures S6C and S6D, which show that both the mean number of SC^+ in the DNA, as well as the mean rate of arrests during elongation decrease for increasing Gyrase.

Finally, also from Figure S6A, as expected, $\Delta t_{elongation}$ converges to a minimum value once the number of Gyrases approaches values that suffice to remove positive supercoils as fast as they appear, not allowing their accumulation.

III.2 Effects of elongation slowdown on the dynamics of RNA production

Next, we show that slowdown of stepwise elongation rates due to PSB does not affect the mean time interval between consecutive RNA production events (therefore not affecting the mean RNA production rate). For this, we simulated the model described in Table S2, but without reactions S2.2 to S2.4, so that one can change the number of Gyrases without affecting transcription initiation rates. This allows testing whether the effects of PSB on elongation (alone) alters the mean RNA production rate.

Using this model, we performed simulations for various values of [G] to obtain the mean rate of RNA production (r) as a function of [G]. Figure S6B shows that r is not significantly affected by [G], within realistic intervals of these parameters' values. This entails that the mean RNA production rate is independent from the effects of PSB on elongation. This is expected, provided that the events in

elongation do not affect the rate of transcription initiation and have negligible effects on the fraction of RNAPs that complete elongation, once initiated.

In this regard, only if the arrests due to PSB (reaction S2.11, Table S2) were long enough that the number of accumulated RNAPs in the DNA strand became so high that new transcription events would not be allowed to initiate due to promoter occupancy.

III.3 RNAP and Gyrase fluctuations within the nucleoid region

Both the SN model and the minimal model assume homogeneous mixing of free Gyrases and RNAP inside each cell (or, more precisely, inside the nucleoid region containing the DNA). There are two reactions in Supplementary Table S2 whose kinetics could be affected, in case this assumption does not hold true, specifically, reaction S2.1 (by which RNAPs bind to the promoter) and reaction S2.17 (by which Gyrases remove positive supercoils).

The assumption of homogenously distributed free RNAP in the nucleoid region is supported by live-cell super-resolution microscopy data (Stracy et al., 2015). Further, it has been showed that, within certain ranges, the total RNAP concentration can be used as a proxy for free RNAP concentration when estimating mean RNA production rates (Lloyd-Price et al, 2016), which would not be expected if the free RNAP had significant spatial fluctuations. It is also noteworthy that time intervals between consecutive transcription events are relatively long (700-1500 s, Figure 3) when compared to the diffusion rate of the RNAP (Bratton et al., 2011), supporting the assumption of a well-stirred system, which allows for the assumption of stochastic rate constants (Gillespie, 1977).

Meanwhile, to test the validity of the model assumption of 'homogeneous mixing of Gyrases in the nucleoid region', we measured their spatial heterogeneity by microscopy (example Figure S8A), using an *E. coli* strain where the gyrA gene is endogenously tagged with the YFP coding sequence (Taniguchi et al., 2010) (Materials and Methods, Section 2.1). From the data, we found that almost all Gyrases are located in the cell region(s) where the nucleoid(s) locate (example Figure S8A).

To estimate the spatial heterogeneity of Gyrases in those regions, we first applied a 2-dimensional Gaussian filter (Materials and Methods, Section 2.7), to remove measurement noise (Wheeler, 2017). Supplementary Figure S8B shows the 'raw' and 'filtered' distributions of pixel intensities (arbitrary units). Next, we multiplied each data point of the filtered distribution by a constant (equal to the ratio of the mean over the variance of the filtered distribution), which results in a Poisson distribution (named 'scaled distribution'). Figure S8C shows both this scaled distribution (inset) as well as its probability density function (pdf). The variance of the scaled distribution is expected to be a good proxy for the spatial (and thus temporal) variability in Gyrase numbers within the nucleoid region.

Subsequently, starting from the minimal model in Figure 1, we introduced the additional reactions in Table S7, so as to test if the *in silico* results are significantly affected by inserting in the model this degree of variability in Gyrase numbers over time. Since the reactions in Table S7 are first-order processes of production (reaction S7.1) and degradation (reaction S7.2) of Gyrases, they ensure that the number of Gyrases at any given time follows a Poisson distribution as the empirical data suggests (since the propensity of each event is constant and independent of the occurrence of the former

event). We then tuned the rate constants (*kp_G* and *kd_G*) so that the *in silico* distribution best fitted the empirical distribution (Figure S8C).

We then used the best fitted model, obtained an *in silico* Δt distribution, and compared to the same distribution, obtained prior to introducing the temporal variability in Gyrase's numbers. A 2-sample t-test did not reject the null-hypothesis that the two distributions cannot be statistical distinguished (p-value > 0.05), from which we conclude that the heterogeneity in Gyrases is not sufficiently high to affect the dynamics of the model.

Overall, we conclude that the additional process (Table S7), added to account for the measured heterogeneity in Gyrase numbers, does not change the RNA production kinetics sufficiently (in a statistical sense) for the single-cell distributions of RNA numbers to differ significantly. Given this and the above, for simplicity, we assume homogenous spatial distributions of RNAP and Gyrases in the region(s) occupied by the nucleoid(s).

IV. Model fitting

Assuming the model in Figure 1, we inferred the parameters that best fit the empirical data as follows. From the model, the inverse of the RNA production rate equals:

$$r_x^{-1} = \left[\frac{k_{lock}}{k_{unlock}\cdot k_{rem}}\cdot\frac{k_p\cdot[R_x]}{k_1\cdot[R_x]}\frac{1}{[G_x]}\right]+\left[\frac{k_{lock}\cdot\lambda}{k_{unlock}\cdot k_{rem}}\cdot\frac{1}{[G_x]}\right]+\frac{1}{k_1.[R_x]} \quad \text{(iv.1)}$$

The variables, r_x^{-1}, $[R_x]$ and $[G_x]$ refer to, respectively, the inverse of the RNA production rate, the concentration of RNAP and the concentration of Gyrases, for a condition 'x' of Gyrase overexpression. For the reference, control condition ('ref'), equation (iv.1) becomes:

$$r_{ref}^{-1} = \left[\frac{k_{lock}}{k_{unlock}\cdot k_{rem}}\cdot\frac{k_p\cdot[R_{ref}]}{k_1\cdot[R_{ref}]}\frac{1}{[G_{ref}]}\right]+\left[\frac{k_{lock}\cdot\lambda}{k_{unlock}\cdot k_{rem}}\cdot\frac{1}{[G_{ref}]}\right]+\frac{1}{k_1.[R_{ref}]} \quad \text{(iv.2)}$$

Next, we assume: $w=\frac{[R_x]}{[R_{ref}]}$ (iv.2.1), $y=\frac{[G_x]}{[G_{ref}]}$ (iv.2.2) and $z=\frac{r_x^{-1}}{r_{ref}^{-1}}$ (iv.2.3). Thus:

$$r_{ref}^{-1}\cdot z = \left[\frac{k_{lock}}{k_{unlock}\cdot k_{rem}}\cdot\frac{k_p\cdot[R_{ref}]}{k_1\cdot[R_{ref}]}\frac{1}{[G_{ref}]}\cdot\frac{1}{y}\right]+\left[\frac{k_{lock}\cdot\lambda}{k_{unlock}\cdot k_{rem}}\cdot\frac{1}{[G_{ref}]}\cdot\frac{1}{y}\right]+\left[\frac{1}{k_1.[R_{ref}]}\cdot\frac{1}{w}\right] \quad \text{(iv.3)}$$

From (iv.1), (iv.2.3), and (iv.3):

$$z = \left[\frac{\frac{k_{lock}}{k_{unlock}\cdot k_{rem}}\cdot\frac{k_p\cdot[R_{ref}]}{k_1\cdot[R_{ref}]}\frac{1}{[G_{ref}]}}{r_{ref}^{-1}}\right]\cdot\frac{1}{y}+\left[\frac{\frac{k_{lock}\cdot\lambda}{k_{unlock}\cdot k_{rem}}\cdot\frac{1}{[G_{ref}]}}{r_{ref}^{-1}}\right]\cdot\frac{1}{y}+\left[\frac{\frac{1}{k_1.[R_{ref}]}}{r_{ref}^{-1}}\right]\cdot\frac{1}{w} \quad \text{(iv.4)}$$

Assuming, $\alpha = \frac{k_{lock} \cdot k_p \cdot [R_{ref}]}{k_{unlock} \cdot k_{rem} \cdot [G_{ref}]}$ (iv.4.1), $\beta = \left[\frac{1}{\frac{k_1 \cdot [R_{ref}]}{r_{ref}^{-1}}} \right]$ (iv.4.2), $\eta = \frac{\frac{k_{lock} \cdot \lambda}{k_{unlock} \cdot k_{rem}} \cdot \frac{1}{[G_{ref}]}}{r_{ref}^{-1}}$ (iv.4.3),

equation (iv.4) becomes:

$$z = \alpha\beta \cdot \frac{1}{y} + \eta \cdot \frac{1}{y} + \beta \cdot \frac{1}{w} \quad \text{(iv.5)}$$

Since $LacO_3O_1$ and the native Lac promoter are located in the same position in the chromosome, we assume that they have the same propensity to become locked due to PSB due to variables other than their own transcription rate. Thus, it is imposed that k_{unlock}, k_{lock}, and k_p (Figure 1) do not differ between them. Positive supercoils' removal should also not differ. As such:

$$z_{LacO3O1} = \alpha.\beta_1.\frac{1}{y} + \beta_1.\frac{1}{w} + \eta.\frac{1}{y} \quad \text{(iv.6)}$$

$$z_{Lac} = \alpha.\beta_2.\frac{1}{y} + \beta_2.\frac{1}{w} + \eta.\mathrm{X}.\frac{1}{y} \quad \text{(iv.7)}$$

where $\beta_1 = \frac{\frac{1}{k_{1\,LacO3O1} \cdot [R_{ref}]}}{r_{ref\,LacO3O1}^{-1}}$ (iv.7.1) , $\beta_2 = \frac{\frac{1}{k_{1\,Lac} \cdot [R_{ref}]}}{r_{ref\,Lac}^{-1}}$ (iv.7.2), $\eta = \frac{\frac{k_{lock} \cdot \lambda}{k_{unlock} \cdot k_{rem}} \cdot \frac{1}{[G_{ref}]}}{r_{ref\,LacO3O1}^{-1}}$ (iv.7.3), and

$\mathrm{X} = \frac{r_{ref\,LacO3O1}^{-1}}{r_{ref\,Lac}^{-1}}$ (iv.7.4)

Given the empirical data in Table S5, from (iv.6) and (iv.7) we estimated α, β_1, β_2 and η by imposing the condition $\alpha, \beta_1, \beta_2, \eta \geq 0$. Namely, we searched for the set of solutions that minimizes the mean squared error (equation iv.8):

$$MSE = \frac{\sum_{i=1}^{N_1} \left(z_i^{LacO3O1} - \hat{z}_i^{LacO3O1}(\alpha, \beta_1, \lambda) \right)^2 + \sum_{i=1}^{N_2} \left(z_i^{Lac} - \hat{z}_i^{Lac}(\alpha, \beta_2, \lambda, X) \right)^2}{N_1 + N_2} \quad \text{(iv.8)}$$

The best fitting solution found was: $\alpha = 0.04$, and $\eta = 0.48$ for both promoters. Meanwhile, $\beta_1 = 0.5$ for $P_{LacO3O1}$, and $\beta_2 = 0.71$ for P_{Lac}.

Figures S7A and S7B show the resulting *z* surfaces for $P_{LacO3O1}$ and P_{Lac}. Both models best fit with a mean squared error of 0.004. To determine the goodness of fit of the surfaces we calculated R^2 values. Both surfaces had $R^2 > 0.95$, from which we conclude that the model well-fits the empirical data.

V. Inference of rate constants of $LacO_3O_1$

First, we infer the expected rate of transcription initiation events, in the absence of PSB. From the measurements (Figure 3, main manuscript) and model fitting (Section S.IV), we found that for $P_{LacO3O1}$ one has: r_{ref}^{-1} = 1476 s and β = 0.5. Given that, and the definition of β (equation iv.4.2), we find that $k_1.[R_{ref}]$ = 0.0014 s^{-1}. This is in agreement with the rate of initiation estimated in (Lloyd-Price 2016).

Next, we infer the expected rate of SC^+ production due to the transcription activity of neighboring genes. From equations (iv.4.1) and (iv.4.3), along with the values of k_{unlock} (Section 3.5, main manuscript), α, η, r_{ref}^{-1} (Supplementary Section IV) and λ (equaling a tenth of the number of nucleotides of the elongation region of $LacO_3O_1$) one has:

$$0.04 = \frac{k_{lock} \cdot k_p \cdot [R_{ref}]}{7\times10^{-4} \cdot k_{rem}.[G_{ref}]} \quad \text{(v.1)}$$

and

$$0.48 = \frac{\frac{k_{lock} \cdot 406}{7\times10^{-4} \cdot k_{rem}} \cdot \frac{1}{[G_{ref}]}}{1476} \quad \text{(v.2)}$$

Dividing equation (v.1) with equation (v.2) one has:

$$\frac{0.04}{0.48} = \frac{k_p.[R_{ref}].(1476)}{406} \quad \text{(v.3)}$$

From (v.3), one finds that $k_p \cdot [R_{ref}]$ equals 0.023. From this and equation (v.1), one obtains:

$$0.04 = \frac{0.023 \cdot k_{lock}}{7\times10^{-4} \cdot k_{rem}.[G_{ref}]} \quad \text{(v.4)}$$

From (v.4) one can write:

$$\frac{k_{lock}}{k_{rem}.[G_{ref}]} = 1.2\times10^{-3} \quad \text{(v.5)}$$

Next, we infer the expected rate of SC^+ removal. According to (Stracy et al, 2019) Gyrase dwell times are of 2 seconds to remove ~2 supercoils. Based on this, given that there is approximately one Gyrase molecule per DNA loop (Chong et al., 2014), we set: $k_{rem} \cdot [G_{ref}]$ = 1 s^{-1}. From this, along with equation (v.5), one estimates $k_{lock} = 1.2\times10^{-3}$ s^{-1}.

.

VI. Using the concentration of RNAP and of Gyrases as proxies for the concentrations of free RNAP and free Gyrases, respectively.

According to the model in Figure 1, one expects the transcription rate of a given gene to depend on the concentration of RNAP. However, at any given time, several RNA polymerases may not be available, if already committed to transcription. In detail, at any given time, a significant fraction of RNA polymerases are not available for new transcription events (estimations suggest that, at any given moment, ~48% of all RNAPs are bound to the DNA, interacting with promoter regions or involved transcription elongation (Stracy et al., 2015). Further, this task can take up to 75 s (Vogel and Jensen, 1994). As such, more accurately, and in accordance with the single nucleotide (SN) model, the transcription rate of a gene depends on the concentration of RNA polymerases that are *free* for transcription. Specifically, from reaction 1 in Figure 1, the inverse of the transcription rate of a gene should change linearly with the inverse of the free RNAP concentration.

Since one cannot easily measure the fraction of RNAP that is free for transcription at a given moment, for our estimations, we use the total concentration of RNA polymerases, [RNAP], as a proxy. This is possible because, within the range of conditions of the measurements, it was empirically verified in (Lloyd-Price et al., 2016) that the Lineweaver–Burk plot of the inverse of [RNAP] against the inverse of the transcription rate shows a straight line. This linear relationship was established by WTLS by minimizing χ^2 (Krystek and Anton, 2007), and then confirmed by showing that small deviations from linearity were not statistically significant, using likelihood ratio tests between the best linear fit and fits by higher order polynomials. In no case did the test reject the linear model (p-values above 0.1). This result was subsequently confirmed in (Kandavalli et al., 2016; Mäkelä et al., 2017; Oliveira et al., 2019; Startceva et al., 2019). Since this is strong evidence that the ratio between free and total RNAP concentrations is constant for the range of conditions tested here, we use the total RNAP concentration as a proxy for free RNAP concentration, relative to the control.

Similarly, from the equation in the large inset in Figure 1, one can also expect a straight line on a plot of the inverse of the free Gyrases concentration against the inverse of the transcription rate of the gene of interest, if [G] is a good proxy for the concentration of freely diffusing Gyrases [G^{free}]. In this regard, only ~49% of the Gyrases are expected to be free for resolving new transcription-generated positive supercoils, while the remaining ones are transiently maintaining steady state levels of negative supercoiling (~49%) and resolving replication-generated supercoiling (~2%) (Stracy et al., 2019). Further, resolving positive supercoils can take up to 1 s per supercoil (Stracy et al., 2019).

To test if [G] is a good proxy for [G^{free}], we measured RNA production rates in cells subject to various Gyrase concentrations (obtained by overexpressing Gyrase, see Materials and Methods in main manuscript). In Figure 3C (black line), we show the inverse of the transcription rate as a function of $[G]^{-1}$. Next, we fitted a line by WTLS and determined if small deviations from linearity are statistically significant by a likelihood ratio test between the best linear fit and fits by higher order polynomials (by WTLS by minimizing χ^2) (Krystek and Anton, 2007). The test did not reject the linear model (p-value > 0.99), from which we conclude that r^{-1} decreases linearly with $[G]^{-1}$. As such, in what pertains the extrapolation of τ_{active} and τ_{locked} from Lineweaver–Burk plots shown in the main manuscript, [G] is used as a proxy for [G^{free}] in the range of conditions considered (differing in the

concentration of Rhamnose, which is responsible for Gyrase, overexpression, see Materials and Methods in main manuscript).

Due to the above, in the main manuscript, for simplicity, we refer to the total concentrations of Gyrase and RNAP rather than to the concentrations of *free* Gyrase and RNAP molecules.

VII. Estimation of the quantitative relationship between the concentrations of Rhamnose and active gyrases

To obtain the fold change (F) in protein levels for a given mRNA fold change (due to adding Rhamnose), we use the calibration line ($Y_{Gyr.}$ = (0.85 +/- 0.06).X + (0.15+/- 0.07)) in Figure 2B**.** Let Y_{Gyr1} be the relative protein numbers corresponding to a Gyrase mRNA fold change of 1 and Y_{Gyr2} be the relative protein numbers for any given fold change. If b is the y-axes intersection (which equals 0.15, Figure 2B) one has:

$$F = \frac{Y_{Gyr_1} - b}{Y_{Gyr_2} - b} \qquad \text{(vii.1)}$$

Using this method, we found that fold changes in mRNA numbers resulted in the same fold changes in protein numbers (Table S3).

VIII. Dissection of the effects of RNAP overexpression in the RNA production rate of the target gene when overexpressing Gyrase.

To dissect the effects of RNAP overexpression from the direct effects of Gyrase overexpression on the RNA production rate of target gene, we estimated r^{-1} (equation iv.1) assuming that [G] has no effects on [RNAP] (i.e. considering that [R] = [R_{ref}]). Thus, we re-write equations iv.6 and iv.7 as:

$$z_x^{'} = \alpha.\beta_x.\frac{1}{y} + \beta_x + \eta.\frac{1}{y} \qquad \text{(viii.1)}$$

In (viii.1), $z_x^{'}$ represents z for promoter x assuming that [RNAP] is unaffected by Gyrase overexpression. The results for each condition are shown in Table S5.

IX. Extraction of RNA production rates from microscopy images

To estimate RNA production rates from the number of RNAs in individual cells obtained from microscopy images at two time points (Häkkinen and Ribeiro, 2015; Zimmer et al., 2016), we account for RNA dilution due to cell division, but not for RNA degradation, since the binding by multiple MS2-GFP molecules makes the tagged RNAs virtually immortal (Golding et al., 2005; Tran et al., 2015) and their fluorescence intensity constant for the duration of the measurements (Tran et al., 2015).

The rate of RNA dilution (k_d) due to cell division can be estimated from the numbers of cell division events between the start (t_0) and end (t_f) of the measurements, along with the mean RNA numbers at the start and end of the measurement period. In detail, let M_0 be the mean number of RNAs per cell at

moment t_0, and M be this number at moment t. It follows that, accounting for RNA dilution due to cell division, the rate of RNA production per cell (r) during that period of time equals:

$$r = \frac{k_d}{\ln 2} \cdot \frac{M - M_0 e^{-k_d \cdot t}}{1 - e^{-k_d \cdot t}} \quad \text{(ix.1)}$$

From the values of t, M, M_0, and k_d extracted from empirical data, one can then obtain the mean of r^{-1}, and the standard error of the mean using the Delta Method (Casella et al., 2002).

X. Minimum supercoiling buildup effects that can be detected

It is possible to estimate the sensitivity of the method used in detecting effects of PSB on the transcription rate of the gene of interest. From the predicted bounds of the line fitting (obtained from the standard errors of the mean of each empirical data point), we estimated the minimum, detectable difference in transcription rates in the control condition assuming the same number of measurements as in Figure 3.

For this, using the blue line in Figure 3 as a starting point, we kept its y-intercept unchanged, and incrementally reduced its slope by 0.01 and calculated the resulting r^{-1} values for each of the 3 data points. The standard error of each data point was kept unchanged (which use it as an upper bound for their expected standard error of the mean). Next, we performed a 2-sample t-test with unequal variance to find if this slope differs from the slope of the grey line. We continued this procedure until this test could not find a difference.

Using this method, we found that the smallest slope below which the p-value is above the significance level (alpha=0.1) equals 512.5. This entails that the minimum fold change in transcription rate due to PSB that could be detected equals 1.6.

This value could be reduced by several procedures, such as collecting more data points per condition. For estimation, assume that the number of data points is increased from 3 to 10, selected within the range of 0.2 to 1 (inverse relative Gyrase concentration). From the best fitting line (blue line in Figure 3) and its predicted bounds, we estimate the mean and standard error for each data point (the mean is obtained from the blue line, with the error is set to be equal to the predicted bound, i.e. the width of the shadow area, in the same position in the x-axis). Next, using WTLS (Krystek and Anton, 2007) one can estimate a new best fit line using the estimated data points as above. From the resulting line, as above, we estimated the minimum fold change to equal 1.19. For improving this estimation, we repeated the process of collecting and processing 10 data points 1000 times, by randomly sampling 1000 data points from a normal distribution with the same mean and standard error for each data point. On average, the expected minimum fold change equalled 1.2. Other possible means of further reducing this minimum fold change include increasing the accuracy of the measurements of each data point (e.g. measuring more cells by microscopy, etc.).

Interestingly, this sensitivity is expected to suffice to detect effects of specific closely space promoter configurations. For example, using the same RNA detection technique, (Häkkinen et al., 2019) recently reported that the promoters tetA and $LacO_3O_1$, when in separate constructs, have a

mean RNA production rate of 1/800 RNA/s (~1/670 for tetA and ~1/1100 for $LacO_3O_1$), while in a tandem formation ($LacO_3O_1$ followed by TetA) the rate equals 1/700 RNA/s. Thus, the fold change between them equals 1.2, which is within the range of detectable fold changes.

SUPPLEMENTARY FIGURES

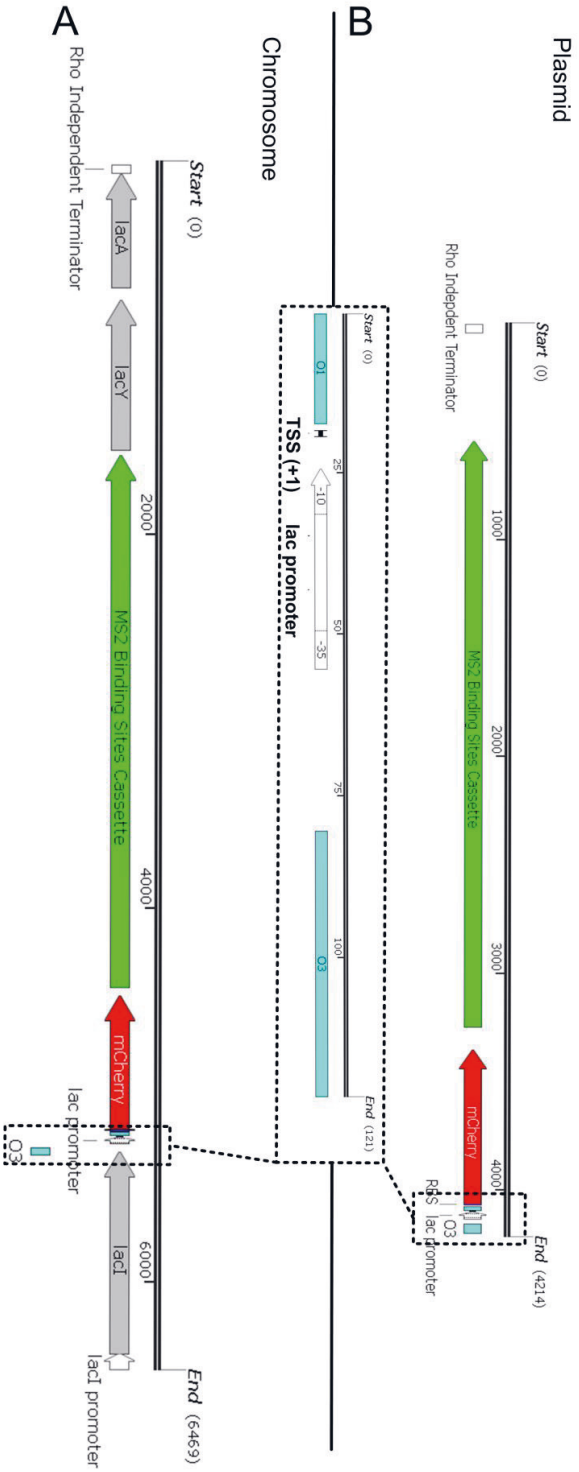

Figure S1. Genetic constructs. **(A)** Chromosome-integrated gene. mCherry-MS2-BS cassette under the control of $P_{LacO3O1}$ in the lac locus of *E. coli* strain BW25993 (ΔlacZ:mCherry-MS2-BSs), followed by the native lacY and lacA genes, and the native Rho-independent transcription termination site. **(B)** Single-copy plasmid-borne gene. mCherry-MS2-BS cassette under the control of $P_{LacO3O1}$ in a single-copy F-Plasmid in *E. coli* strain BW25993, followed by a Rho-independent transcription termination site. Constructs were confirmed by sequencing. Since the plasmid carrying the target gene does not code for lacY and lacA, and the cells carrying this plasmid also contain the original lacY and lacA genes in the chromosome, the two strains express lacY and lacA proteins similarly and, thus, do not differ significantly in the dynamics of IPTG intake. (Inset) The inset image in between the two constructs shows $P_{LacO3O1}$ with functional domains, which is identical in both constructs. It is in this region that the operator site O3 locates, followed by the RNA polymerase binding regions (positions -10 and -35), the transcription start site (TSS, position +1), and the operator site O1. Finally, the plasmid construct has a terminator upstream of the TSS that is 27 nucleotides long and is located 9 nucleotides downstream of the CmR gene (not represented in the figure). Thus, it is similar to the chromosome-integrated construct, where there is an upstream transcriptional terminator provided by the lacI gene. Related to Section 2.1, in main manuscript.

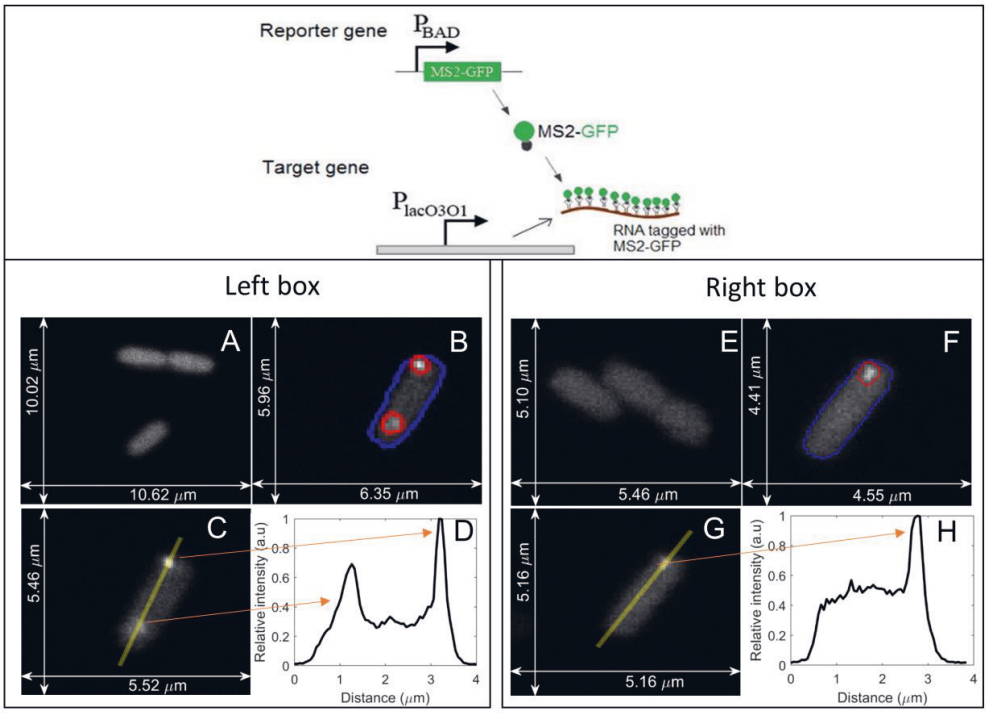

Figure S2. RNA quantification in individual cells by RNA tagging with MS2-GFP, due to which RNAs appear as fluorescent spots **(Top)** Schematic representation of the single-RNA MS2-GFP tagging detection system. Cells produce multiple MS2-GFP reporter proteins, under the control of P_{BAD}, while the production of RNAs target for MS2-GFP is under the control of $P_{LacO3O1}$. **(Left box)** Cells with a plasmid-borne promoter producing the target RNA. **(A)** Example microscopy image of cells carrying the reporter gene coding for MS2-GFP, prior to the production of target RNAs. The cells fluorescence is due to the large amount of MS2-GFP proteins. **(B)** Example microscopy image of cells carrying the reporter gene coding for MS2-GFP, after the production of target RNAs. The RNAs tagged with MS2-GFP are visible as bright spots. Blue line and red circles are the results of cell and RNA spot segmentation, respectively (Materials and Methods). **(C)** Example image of a cell along with a, manually introduced yellow line (using imageJ (Abramoff et al., 2004)) in order to obtain a fluorescence intensity profile along the major axis. **(D)** Pixel intensity (in arbitrary units) along the yellow line shown on (C). The peaks correspond to the regions where the two spots (tagged MS2-GFP RNAs) are located. **(Right box)** Cells with a chromosome-integrated promoter producing RNA target for MS2-GFP. Images from (E) to (H) have the same information as (A) to (D), respectively, but are obtained using cells with a chromosome-integrated promoter responsible for the production of the RNA target for MS2-GFP. Related to Supplementary Section I.

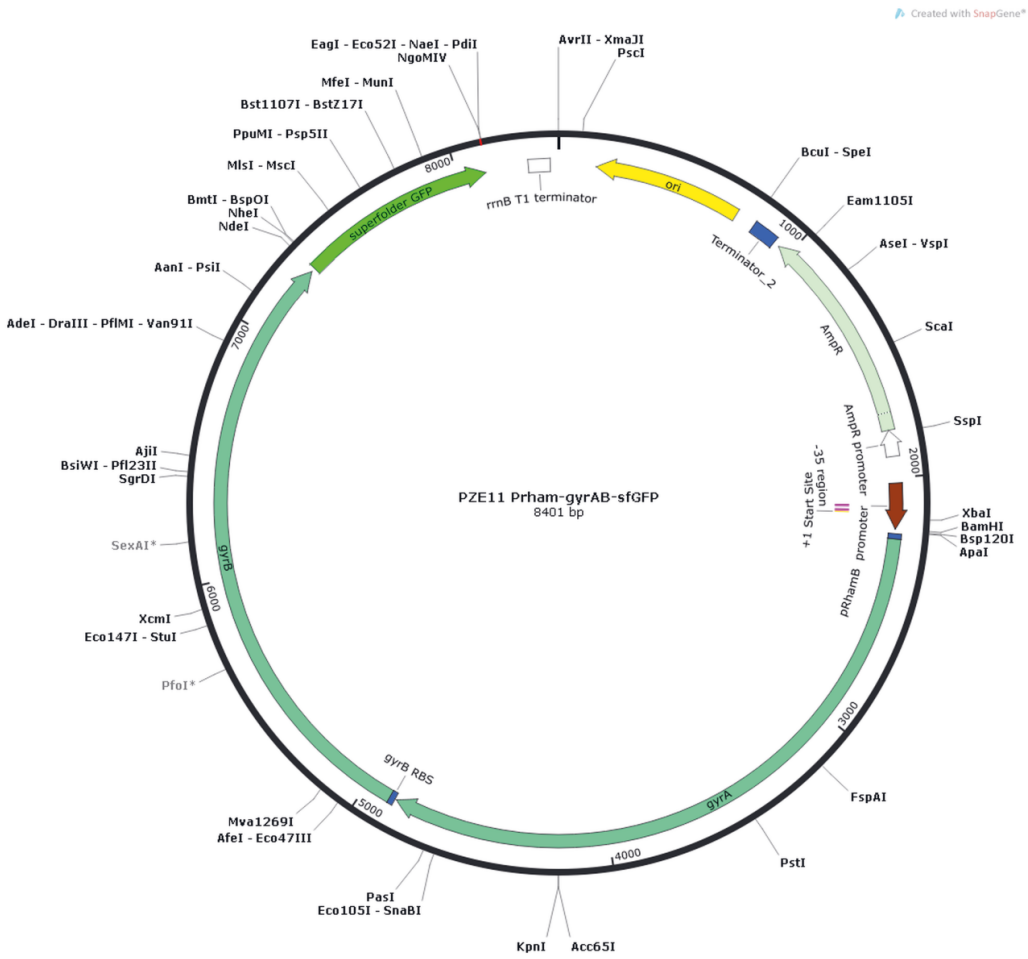

Figure S3. Schematic representation of the plasmid responsible for Gyrase overexpression. This plasmid was constructed by placing the gyraseA and gyraseB under the control of the PRhamB promoter, which is inducible by Rhamnose, and was transformed into BW25993 cells. Adapted from SnapGene® 1.5.2. Related to Section 2.1, in main manuscript.

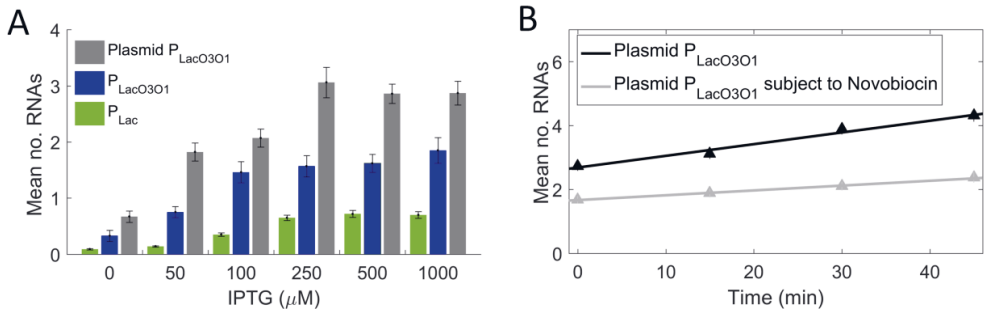

Figure S4. Mean number of RNAs produced by the target genes in individual cells, as measured by microscopy and RNA tagging by MS2-GFP. **(A)** Induction curves. Mean and Standard Error of the Mean (SEM) of the number of target RNA molecules in individual cells as a function of the induction strength. Results are shown for a chromosome-integrated gene controlled by the $P_{LacO3O1}$ promoter (blue bars), for a single-copy plasmid-borne gene controlled by the $P_{LacO3O1}$ promoter (grey bars) and for a chromosome-integrated gene controlled by the native P_{Lac} promoter (green bars), 1 hour after induction of the target gene. More than 100 cells were analyzed per condition. **(B)** Mean integer-valued RNA numbers produced over time in individual cells, each carrying the single-copy plasmid-borne gene under the control of $P_{LacO3O1}$, when subject to 100 μg/ml Novobiocin (grey triangles) and in the control condition (no Novobiocin, black triangles). For each time point, new cells were taken from the original culture. Best linear fits were calculated by WTLS (Krystek and Anton, 2007). The equation of the grey line is Y = (0.015 ± 0.004).X + (1.67 ± 0.11) and of the black line is Y = (0.037 ± 0.005).X + (2.69 ± 0.13). The errors are obtained by the standard error of the mean. Related to Figures 3 and 5. More than 100 cells were analyzed per condition.

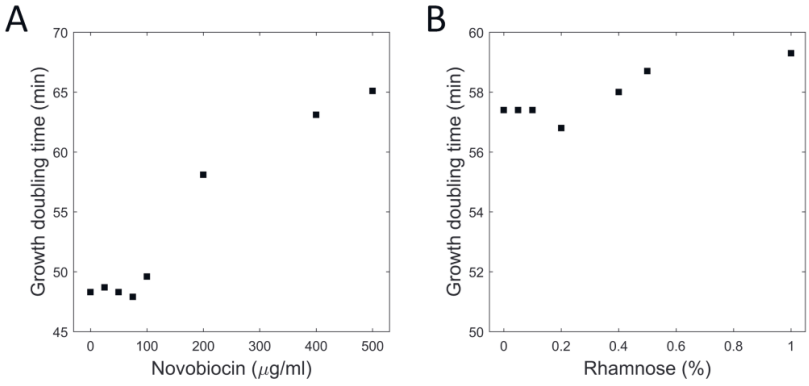

Figure S5**.** Cell growth doubling times. Doubling times versus the concentration of **(A)** Novobiocin and **(B)** Rhamnose. Doubling times were measured from the initial and final OD_{600} and, from the time interval in between (100 minutes). Related to Figures 3 and 4.

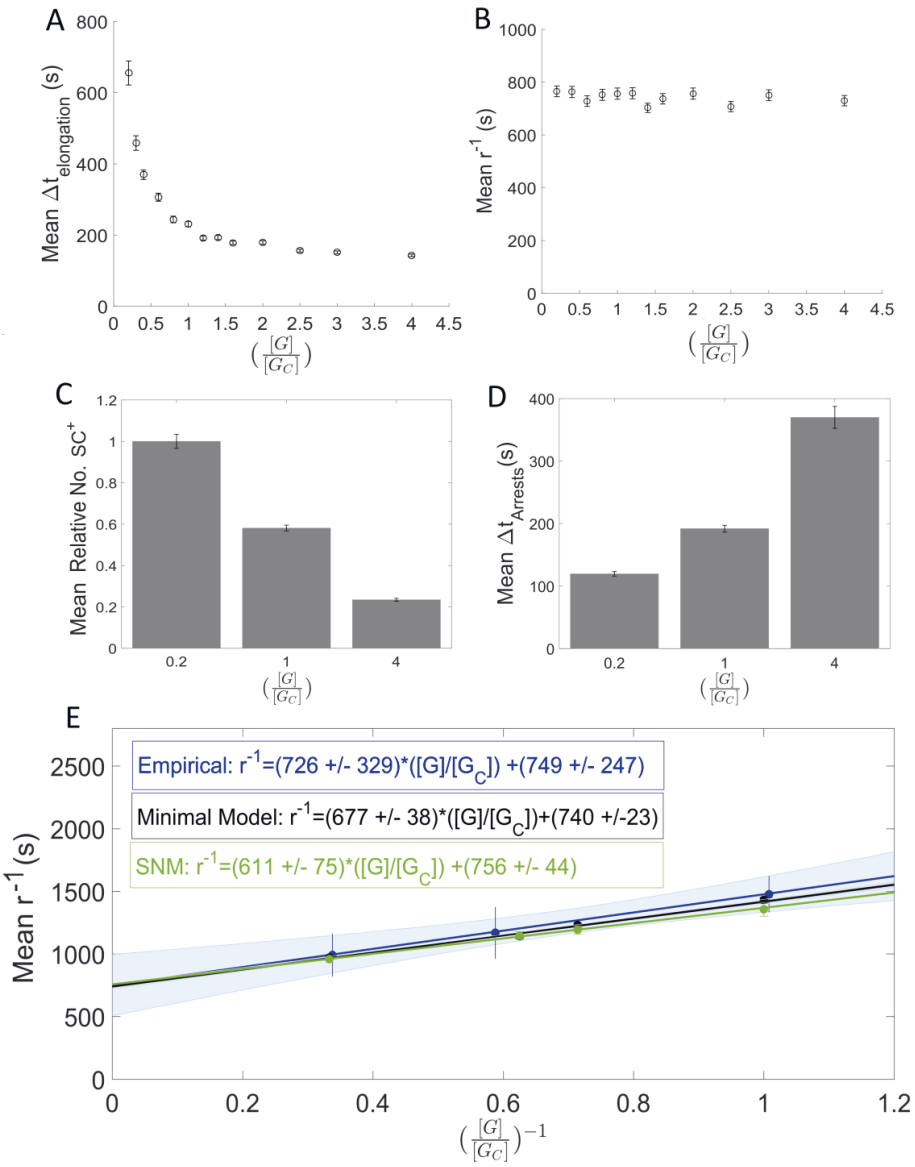

Figure S6. *In silico* results on the effects of Gyrase. **(A)** Mean $\Delta t_{elongation}$ (± SEM) as a function of Gyrase concentration (model in Table S2, with [G_c] being the concentration in the control condition). Relative Gyrase concentrations were set to [0.2, 0.3, 0.4, 0.6, 0.8, 1.0, 1.2, 1.4, 1.6, 2.0, 2.5, 3.0, 4.0]. **(B)** Mean Δt (± SEM) as a function of Gyrase (model in Table S2). **(C)** Mean relative SC^+ levels for different relative Gyrase concentrations [0.2, 1.0, 4.0]. **(D)** Mean time intervals (Δt) between consecutive arrests for different relative Gyrase concentrations [0.2, 1.0, 4.0]. As Gyrase concentration increases, the propensity of arrests during transcription elongation decreases. Related with reaction 6 in Table S2. **(E)** Effects of Gyrase overexpression on the dynamics of RNA production in the SN model (Table S2), the minimal model (Figure 1, main manuscript) and live cells (Figure 3,

main manuscript). In both the minimal and the SN model, the stochastic rate constants were set to k_1 = 0.0014 s^{-1}, k_{lock} = 0.0012 s^{-1}, k_{unlock} = 7x10^{-4} s^{-1}, k_{remove} = 1 s^{-1} and k_p=0.023. All rate constants were extracted or derived from empirical data (Supplementary Section VIII and Table S2). The reactions composing the SN model are shown in Table S2. The reactions of the minimal model are shown in Figure 1 and Section 3.1 of the main manuscript.

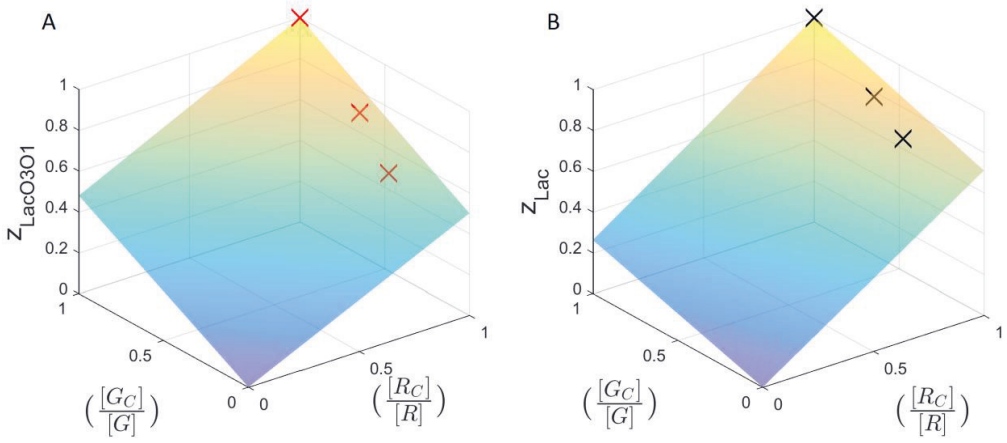

Figure S7. Z surfaces of the best fitting models. **(A)** Z surface for $P_{LacO3O1}$ (α = 0.04, β_1 = 0.5 and η = 0.48) as a function of $\frac{[R_C]}{[R]}$ and $\frac{[G_C]}{[G]}$. The red crosses mark the empirical data points for the $P_{LacO3O1}$ promoter. **(B)** Z surface for P_{Lac} (α = 0.04, β_2 = 0.71 and η = 0.48) as a function of $\frac{[R_C]}{[R]}$ and $\frac{[G_C]}{[G]}$. The black crosses mark the empirical data points for the P_{Lac} promoter. For both (A) and (B), the model fits the empirical data with a mean squared error of 0.0004. To estimate the goodness of fit of the surfaces, we calculated R^2 values. Both surfaces had R^2 > 0.95. Related to Figure 1 and Supplementary Section IV.

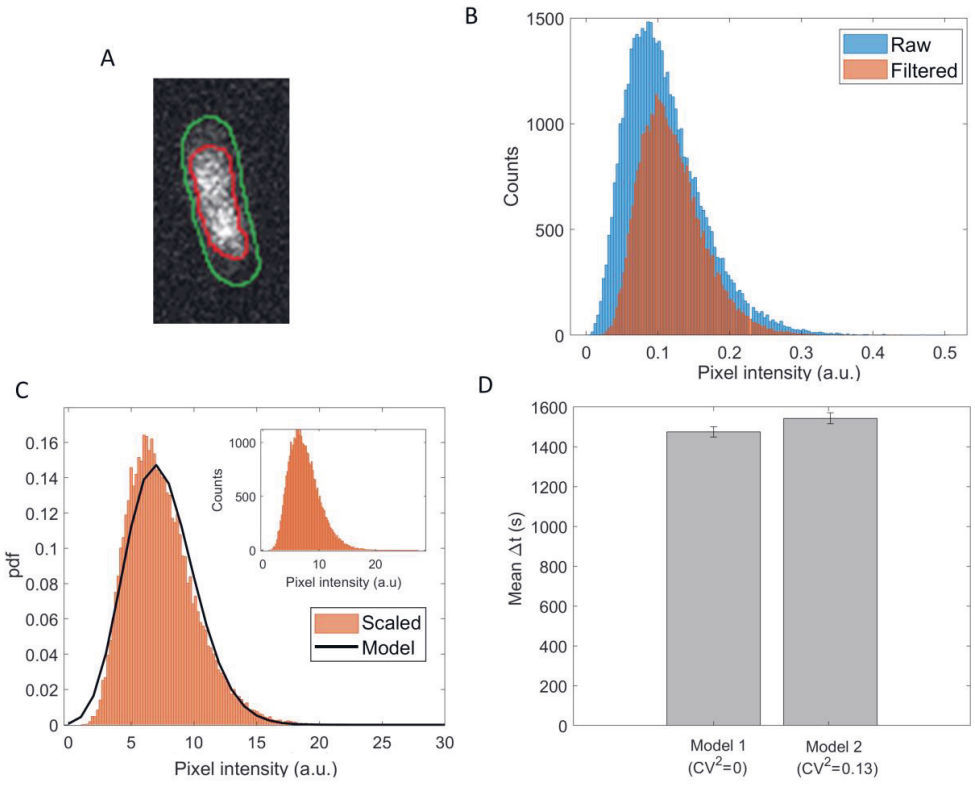

Figure S8. Microscopy measurements of pixel fluorescence intensities of *E. coli* cells expressing the gyrA gene endogenously tagged with a YFP coding sequence (data from 91 cells). **(A)** Example image of a cell with segmented borders (green line, based on phase contrast images) and segmented region with GyrA-YFP (red line). **(B)** Empirical pixel intensity distribution (in arbitrary units (a.u.)) prior ('Raw') and after applying a 2D Gaussian filtering ('Filtered'). The Raw distribution has a mean of 0.11 and a standard deviation of 0.054, while the Filtered distribution has a mean of 0.12 and a standard deviation of 0.044. **(C)** (inset) Empirical pixel intensity distribution multiplied by a constant (equal to the ratio between the mean and the variance of the filtered distribution in B). This distribution has a mean of 7.43 and a variance of 7.43 and can be well approximated by a Poisson (scaled) distribution. The scaled distribution is used to model the variability in Gyrase numbers over time (Table S7). Probability density function (PDF) of the scaled distribution and PDF of the *in silico* distribution of Gyrase numbers over time, fitted by a Poisson (scaled) distribution (black line). **(D)** Mean of the Δt distribution of intervals between RNA production events with (model 2) and without (model 1) variability in Gyrase numbers. Error bars correspond to the standard error of the mean. The imposed CV^2 is set (equal to) from the CV^2 of the scaled empirical distribution in (C). A two-sample t-test between the results from models 1 and 2 did not reject the null hypothesis that the two Δt distributions cannot be statistical distinguished (p-value > 0.05). Related to Supplementary Section III.3.

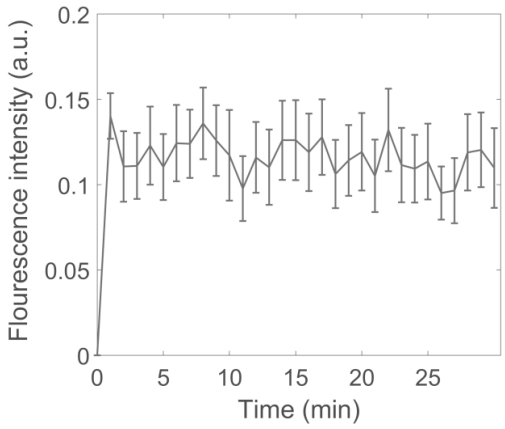

Figure S9. Mean fluorescence intensity of MS2-GFP tagged RNA molecules over time, since first appearing. 76 tagged RNAs were tracked, all from cells with only one tagged RNA. Also shown is the standard error of the mean (vertical bars). Related to Supplementary Section I.

SUPPLEMENTARY TABLES

Table S1. qPCR of the target genes under the control of $P_{LacO3O1}$ and P_{Lac}. Shown are the normalized average of the C_T value of the target gene by the respective mean C_T value of the respective reference condition ($\Delta C_{T(tested)}$), normalization of the $\Delta C_{T(tested)}$ value of the target gene to the $\Delta C_{T(tested)}$ value of the calibrator gene (0 % Rhamnose, $\Delta\Delta C_T$) and, finally, the fold change between the target and calibrator gene ($2^{-\Delta\Delta Ct}$). For each promoter, the samples were identically isolated, prepared and handled, and thus contain identical amounts of cDNA in each well. Related to Figures 3 and 5.

		mCherry			16SrRNA			$\Delta C_{T(tested)}$	$\Delta\Delta C_T$	$2^{-\Delta\Delta Ct}$
	Rhamnose (%)	1	2	3	1	2	3			
$LacO_3O_1$	**0.0**	26.63	26.49	26.26	13.52	13.33	13.28	13.08	0	1
	0.1	26.28	26.03	26.19	13.27	13.65	13.63	12.65	-0.43	0.74
	0.2	25.62	25.48	25.56	13.38	13.13	13.23	12.31	-0.78	0.58
Plasmid-Borne $LacO_3O_1$	**0.0**	23.17	23.18	23.04	7.79	7.81	7.36	15.48	0	1
	0.1	23.20	23.30	23.04	8.31	7.75	7.59	15.30	0.18	0.88
	0.2	23.18	23.18	23.28	8.34	7.72	7.78	15.27	-0.21	0.86
Lac	**0.0**	23.90	23.68	23.60	8.57	8.61	8.58	15.14	0	1
	0.1	22.71	22.76	23.27	8.15	8.09	7.94	14.86	-0.28	0.82
	0.2	23.39	23.23	22.87	8.60	8.34	8.36	14.73	-0.41	0.75

Table S2. Model of transcription at the single nucleotide (SN) level. Shown are the chemical reactions representing the various processes, and the stochastic rate constants (in s^{-1}) used to model transcription initiation, elongation and parallel and competing events at the nucleotide level (on example nucleotide *n*), and termination and RNA production. References from which rate constants were extracted are reported in column "Ref.". Pro stands for the promoter region, *RNAP* for the RNA polymerase, and *RNAP.Pro* for the promoter region when occupied by an *RNAP*. A_n, O_n and U_n stand for the n^{th} nucleotide when active, occupied, and unoccupied, respectively. Ranges of nucleotides are denoted as in *U*[start,end], which denotes a particular set of consecutive, unoccupied nucleotides from indexes start to end. O_{n_p}, $O_{n_{ar}}$ and $O_{n_{correcting}}$ are used to represent a paused, arrested, or error correcting *RNAP* at position n, respectively. On the template, each RNAP occupies (2*Δ*+1) nucleotides, where *Δ* = 12. These nucleotides cannot be occupied by any other *RNAP* at the same time.

Event	Reaction	Rate constant	Ref.
Reaction S2.1 Initiation	$\text{Pro} + \text{RNAP} \xrightarrow{k_1} \text{RNAP.Pro}$	$k_1 = 0.0014$	Section V In accordance with (Lloyd-Price et al., 2016)
Reaction S2.2 Promoter locking	$\text{Pro} + \text{SC}^+ \xrightarrow{k_{lock}} \text{Pro}_{\text{lock}} + \text{SC}^+$	$k_{lock} = 0.0012$	Section V Based on Empirical results and (Stracy et al., 2019)
Reaction S2.3 Promoter unlocking	$\text{Pro}_{\text{lock}} \xrightarrow{k_{unlock}} \text{Pro}$	$k_{unlock} = 7\text{x}10^{-4}$	(Chong et al., 2014)
Reaction S2.4 External Positive supercoils production	$\xrightarrow{k_p \cdot RNAP} \text{SC}^+$	$k_p = 0.023$	Section V

Reaction S2.5 Promoter escape	$\mathrm{RNAP.Pro} + \mathrm{U}_{[1,(\Delta+1)]} \xrightarrow{k_m} \mathrm{O}_1 + \mathrm{Pro}$	k_m = 150	(Phroskin et al., 2010)
Reaction S2.6 Elongation	$\mathrm{A}_n + \mathrm{U}_{n+\Delta+1} \xrightarrow{k_m} \mathrm{O}_{n+1} + \mathrm{U}_{n-\Delta} + \left\{\mathrm{SC}^+, \text{if } n = 10 \times k, k \in N\right\}$	k_m = 150	(Vogel and Jensen, 1994)
Reaction S2.7 Activation	$\mathrm{O}_{n+1} \xrightarrow{k_{act}} \mathrm{A}_{n+1}$	k_{act} = 150, n>10, k_{act} = 30, n<=10,	(Vogel and Jensen, 1994; Phroskin et al., 2010)
Reaction S2.8 Pausing	$O_n \underset{1/\tau_p}{\overset{k_{pause}}{\rightleftarrows}} O_{n_p}$	k_{pause} = 0.55 τ_p = 3	(Greive and von Hippel, 2005; Rajala et al., 2010; Landick, 2009)
Reaction S2.9 Pause release due to collision	$\mathrm{O}_{n_p} + \mathrm{A}_{n-2\Delta-1} \xrightarrow{k_{m1}} \mathrm{O}_n + \mathrm{A}_{n-2\Delta-1}$	k_{m1} = 120	(Epshtein and Nudler, 2003)
Reaction S2.10 Pause induced by collision	$\mathrm{O}_{n_p} + \mathrm{A}_{n-2\Delta-1} \xrightarrow{k_{m2}} \mathrm{O}_{n_p} + \mathrm{O}_{n-2\Delta-1_p}$	k_{m2} = 30	(Epshtein and Nudler, 2003)
Reaction S2.11 RNAP arrest due to supercoiling	$\mathrm{O}_n + \mathrm{SC}^+ \underset{1/d_{ar}}{\overset{k_{ar}}{\rightleftarrows}} \mathrm{O}_{n_{ar}} + \mathrm{SC}^+$	k_{ar} =0.03, d_{ar} = 100	(Fujita et al., 2016; Greive and von Hippel, 2005)

Reaction S2.12 Editing	$O_n \underset{1/d_{ed}}{\overset{k_{ed}}{\rightleftarrows}} O_{n_{correcting}}$	k_{ed} = 0.008, d_{ed} = 5	(Greive and von Hippel, 2005)
Reaction S2.13 Misincorporation	$A_{n_{last}} \xrightarrow{k_{mis}} \mathrm{RNA}_{error} + \mathrm{RNAP} + \mathrm{U}_{n_{[last,last-\Delta]}}$	k_{mis}= 0.05	(Greive and von Hippel, 2005)
Reaction S2.14 Premature termination	$\mathrm{O_n} \xrightarrow{k_{pre}} \mathrm{RNAP} + \mathrm{U}_{[(n-\Delta),(n+\Delta)]}$	k_{pre}= 0.00019	(Lewin, 2008)
Reaction S2.15 Pyrophosphor-olysis	$\mathrm{O_n} + \mathrm{U_{n-\Delta-1}} \xrightarrow{k_{pyr}} \mathrm{O_{n-1}} + \mathrm{U_{n+\Delta}}$	k_{pyr} = 0.75	(Erie et al., 1993)
Reaction S2.16 Completion	$\mathrm{A}_{n_{last}} \xrightarrow{k_f} \mathrm{R} + \mathrm{RNAP} + \mathrm{U}_{n_{[last,last-\Delta]}}$	k_f = 2	(Greive et al., 2008)
Reaction S2.17 PSB removal	$\mathrm{Gyr} + \mathrm{SC}^{+} \xrightarrow{k_{remove}} \mathrm{Gyr}$	k_{remove} = 1	(Stracy et al., 2019; Rovinskiy et al., 2012 ; Chong et al., 2014)

Table S3. Gyrase mRNA fold changes (measured by qPCR) for different concentrations of Rhamnose (0, 0.1 and, 0.2 %) and the corresponding Gyrase protein levels relative to the control condition (0 % Rhamnose), obtained from the calibration line (Figure 2B). Also shown are the fold changes in Gyrase protein levels (calculated using equation vii.1). Finally, it is shown the respective standard errors of the mean, calculated using the Delta method (11). Related to Figure 2.

Rhamnose (%)	**Gyrase mRNA fold change**	**Relative Gyrase Protein level (from calibration line Figure 2B)**	**Fold change in Protein levels (Equation vii.1)**
0.0 %	1.00	1± 0.09	1 .00 ± 0.15
0.1 %	1.71 ± 0.20	1.60 ± 0.21	1.71 ± 0.31
0.2 %	2.97 ± 0.32	2.67 ± 0.33	2.98 ± 0.53

Table S4. Average (μ) and standard deviation (σ) of the absolute nucleoid area (pixel) in the control condition (0 % Rhamnose) and when subject to Gyrase overexpression (0.2 % Rhamnose). A 2-sample student t-test for the null hypothesis that the two data sets are from the same distribution was not rejected (p-value > 0.01). The sample sizes are 224 and 237 cells for 0 % and 0.2 % Rhamnose conditions, respectively. Related to Section 3.3 in main manuscript.

	0 % Rhamnose	**0.2 % Rhamnose**
μ	421	449
σ	105	130

Table S5. Dissection of the effects of RNAP overexpression on the transcription rate of $P_{LacO3O1}$ and P_{Lac} when overexpressing Gyrase. Shown are RNAP fold changes, ω, relative to the control (measured by Western blot) and the inverse of the fold change in RNA production rate, z (measured by qPCR). Also shown are the expected fold change in RNA production rate in the absence of change in [RNAP] (z', obtained from equation viii.1), and $(r')^{-1}$, the inverse of the RNA production rate of the target gene in the absence of indirect effects of Gyrase overexpression on the RNAP concentration. Related to Figures 3 and 5.

	$P_{LacO3O1}$		**P_{Lac}**	
	0.1 % Rhamnose	**0.2 % Rhamnose**	**0.1 % Rhamnose**	**0.2 % Rhamnose**
r_{ref}^{-1} (s)	1476	1476	2704	2704
r^{-1} (s)	1094	861	2223	2037
ω	1.05	1.12	1.05	1.12
y	1.71	2.98	1.71	2.98
z	0.74	0.58	0.82	0.75
z'	0.79	0.67	0.88	0.81
$(r')^{-1}$ (s)	1168	990	2380	2190

Table S6. Estimation of *OFF/ON* duty cycle ratios for $P_{LacO3O1}$ and for $P_{Lac,}$ based on model fitting, for different levels of Gyrase overexpression (0, 0.1 and, 0.2% Rhamnose). Shown are the total *OFF/ON* duty cycle ratio (OFF_{total}/ON), the *OFF/ON* duty cycle ratio due to neighbouring activity ($OFF_{neighboring}/ON$) and, the *OFF/ON* duty cycle ratio due to RNA polymerase activity (OFF_{self}/ON). Related to Figure 1 and Section 3.8 in main manuscript.

	$P_{LacO3O1}$			P_{Lac}		
Duty cycle	**0 %**	**0.1 %**	**0.2 %**	**0 %**	**0.1 %**	**0.2 %**
$OFF_{total}/ON = \frac{\left(\frac{k_p.RNAP}{k_1.RNAP}+\lambda\right).\frac{k_{lock}}{k_{unlock}\cdot k_{rem}\cdot G}}{\frac{1}{k_1.RNAP}}$	1.00	0.61	0.37	0.72	0.44	0.28
$OFF_{neighboring}/ON = \frac{k_{lock}\cdot k_p\cdot RNAP}{k_{unlock}\cdot k_{rem}\cdot G}$	0.04	0.02	0.01	0.04	0.02	0.01
$OFF_{self}/ON = \frac{\frac{k_{lock}\cdot\lambda}{k_{unlock}\cdot k_{rem}\cdot G}}{\frac{1}{k_1.RNAP}}$	0.96	0.59	0.36	0.68	0.42	0.27

Table S7. Reactions added to the minimal model to introduce variability in Gyrase numbers. Related with Figure S8. The empirical distribution in Figure S8C is modelled by reactions (S7.1) and (S7.2), provided accurate fitting of the rate constants k_{p_G} and k_{d_G} (their ratio should equal the mean of the empirical distribution). Meanwhile, k_{remove} in reaction (S7.3) was tuned so that 1 Gyrase resolves approximately 1 positive supercoil (SC^+) per second (Stracy et al., 2019) (related to Supplementary Section III.3).

Event	**Reaction**	**Rate constant (s^{-1})**
Gyrase production (S7.1)	$\xrightarrow{k_{p\_G}} Gyr$	k_{p_G} =7.43
Gyrase degradation (S7.2)	$Gyr \xrightarrow{k_{d\_G}}$	k_{d_G} =1
PSB removal (S7.3)	$Gyr + SC^+ \xrightarrow{k_{remove}} Gyr$	k_{remove} = 0.13

SUPPORTING REFERENCES

Abramoff, M.D., Magalhaes, P.J., Ram, S.J. (2004). Image Processing with ImageJ. Biophotonics International 11, 36–42.

Bratton, B. P., Mooney, R. A., Weisshaar, J. C. (2011). Spatial distribution and diffusive motion of RNA polymerase in live *Escherichia coli*. *Journal of Bacteriology*, *193*(19), 5138–5146. doi:10.1128/JB.00198-11

Casella, G., and Berger, R. L. (2002). *Statistical inference*. Thomson Learning.

Cheung, K. J., Badarinarayana, V., Selinger, D. W., Janse, D., & Church, G. M. (2003). A Microarray-Based Antibiotic Screen Identifies a Regulatory Role for Supercoiling in the Osmotic Stress Response of Escherichia coli. Genome Research, 13(2), 206–215. doi:10.1101/gr.401003

Chong, S., Chen, C., Ge, H., and Xie, X. S. (2014). Mechanism of Transcriptional Bursting in Bacteria. *Cell*, *158*(2), 314–326. doi:10.1016/j.cell.2014.05.038

Epshtein, V., Nudler, E. (2003) Cooperation between RNA polymerase molecules in transcription elongation. Science 300: 801-5. doi:10.1126/science.1083219

Erie, D.A., Hajiseyedjavadi, O., Young, M.C., von Hippel, P.H. (1993) Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. Science, 262: 867-873. doi: 10.1126/science.8235608

Fujita, K., Iwaki, M., & Yanagida, T. (2016). Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. Nature Communications, 7, 13788. doi:10.1038/ncomms13788

Gellert, M., O'Dea, M. H., Itoh, T., and Tomizawa, J. (1976). Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase. *Proc. Natl. Acad. Sci. U. S. A.*, *73*(12), 4474–4478. doi: 10.1073/pnas.73.12.4474

Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4), 403–434. doi:10.1016/0021-9991(76)90041-3

Golding, I., Paulsson, J., Zawilski, S. M., and Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, *123*(6), 1025–1036. doi:10.1016/j.cell.2005.09.031

Greive, S.J., von Hippel, P.H.(2005) Thinking quantitatively about transcriptional regulation. Nat Rev Mol Cell Biol, 6: 221-232. doi: 10.1038/nrm1588

Greive SJ, Weitzel SE, Goodarzi JP, Main LJ, Pasman Z, et al. (2008) Monitoring RNA transcription in

real time by using surface plasmon resonance. Proc. Natl. Acad. Sci. USA 105: 3315-20. doi: 10.1073/pnas.0712074105

Häkkinen, 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. doi:10.1093/bioinformatics/btt194

Häkkinen, A. and Ribeiro, A. S. (2015). Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. *Bioinformatics*, *31*(1), 69–75. doi:10.1093/bioinformatics/btu592

Häkkinen, A., Oliveira, S. M. D., Neeli-Venkata, R., & Ribeiro, A. S. (2019). Transcription closed and open complex formation coordinate expression of genes with a shared promoter region. BioRxiv, 842484. doi:10.1101/842484

Kandavalli, V. K., Tran, H., and Ribeiro, A. S. (2016). Effects of σ factor competition are promoter initiation kinetics dependent. *Biochim. Biophys. Acta - Gene Regul. Mech.*, *1859*(10), 1281–1288. doi:10.1016/j.bbagrm.2016.07.011

Kouzine, F., Gupta, A., Baranello, L., Wojtowicz, D., Ben-Aissa, K., Liu, J., Przytycka, T.M., Levens, D. (2013). Transcription-dependent dynamic supercoiling is a short-range genomic force. *Nat Struct Mol Biol* 20:396–403.doi:10.1038/nsmb.2517

Krystek, M. and Anton, M. (2007). A weighted total least-squares algorithm for fitting a straight line. *Meas. Sci. Technol.*, *18*(11), 3438–3442. doi:10.1088/0957-0233/18/11/025

Landick R. (2009) Transcriptional pausing without backtracking. *Proc Natl Acad Sci USA*, 106(22): 8797-8798. doi: 10.1073/pnas.0904373106

Lesne, A., Victor, J.-M., Bertrand, E., Basyuk, E., Barbi, M. (2018). The Role of Supercoiling in the Motor Activity of RNA Polymerases. In Methods in molecular biology (Clifton, N.J.) 1805, 215–232. doi:10.1007/978-1-4939-8556-2_11

Lewin B (2008) Genes IX, 256-299. Jones and Bartlett Publishers, USA

Lilley, D. M., & Higgins, C. F. (1991). Local DNA topology and gene expression: the case of the leu-500 promoter. Mol. Microbiol., 5(4), 779–783. doi: 10.1111/j.1365-2958.1991.tb00749.x

Lloyd-Price, J., Startceva, S., Kandavalli, V., Chandraseelan, J. G., Goncalves, N., Oliveira, S. M. D., A. Häkkinen, Ribeiro, A. S. (2016). Dissecting the stochastic transcription initiation process in live *Escherichia coli*. *DNA Res.*, *23*(3), 203–214. doi:10.1093/dnares/dsw009

Ma, J., Bai, L., and Wang, M. D. (2013). Transcription Under Torsion. Science., 340(6140), 1580–1583. doi:10.1126/science.1235441

Ma, J., & Wang, M. (2014). Interplay between DNA supercoiling and transcription elongation. Transcription, 5(3), e28636. doi:10.4161/trns.28636

Ma, J., Tan, C., Gao, X., Fulbright, R. M., Roberts, J. W., Wang, M. D., & Wang, M. D. (2019). Transcription factor regulation of RNA polymerase's torque generation capacity. *Proc.Nat. Aca.of Sci. U.S.A.*, 116(7), 2583–2588. doi:10.1073/pnas.1807031116

Mäkelä, J., Lloyd-Price, J., Yli-Harja, O., & Ribeiro, A. S. (2011). Stochastic sequence-level model of coupled transcription and translation in prokaryotes. BMC Bioinformatics, 12(1), 121. doi:10.1186/1471-2105-12-121

Mäkelä, J., Kandavalli, V., and Ribeiro, A. S. (2017). Rate-limiting steps in transcription dictate sensitivity to variability in cellular components. *Sci. Rep.*, *7*(1), 10588. doi:10.1038/s41598-017-11257-2

Mitarai, N., Dodd, I. B., Crooks, M. T., and Sneppen, K. (2008). The Generation of Promoter-Mediated Transcriptional Noise in Bacteria. *PLoS Comput. Biol.*, *4*(7), e1000109. doi:10.1371/journal.pcbi.1000109

Muthukrishnan, A.-B., Martikainen, A., Neeli-Venkata, R., and Ribeiro, A. S. (2014). *In vivo* transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells. *PLoS One*, *9*(9), e109005. doi:10.1371/journal.pone.0109005

Naughton, C., Avlonitis, N., Corless, S., Prendergast, J. G., Mati, I. K., Eijk, P. P., Cockroft,S., Bradley.M., Ylstra, B., Gilbert, N. (2013). Transcription forms and remodels supercoiling domains unfolding large-scale chromatin structures. Nat Struct Mol Biol 20(3), 387–395. doi:10.1038/nsmb.2509

Oliveira, S.M.D., Goncalves, N.S.M., Kandavalli, V.K., Martins, L., Neeli-Venkata, R., Reyelt, J., Fonseca,J.M., Lloyd-price,J., Kranz,H. and Ribeiro,A.S. (2019). Chromosome and plasmid-borne PLacO3O1 promoters differ in sensitivity to critically low temperatures. *Sci. Rep.*, 9(1), 4486, doi:10.1038/s41598-019-39618-z

Oliveira, S.M.D., Neeli-Venkata, R., Goncalves, N.S.M., Santinha, J.A., Martins, L., Tran, H., Mäkelä, J., Gupta, A., Barandas, M., Häkkinen, A., Lloyd-Price, J., Fonseca, J.M. and Ribeiro, A.S. (2016). Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*. *Mol. Microbiol.*, 99(4), 686–699, doi:10.1111/mmi.13257

Phroskin S, Rachid Rahmouni A, Mironov A, Nudler E (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science, 328(5977): 504-508

Rajala, T., Häkkinen, A., Healy, S., Yli-Harja, O., & Ribeiro, A. S. (2010). Effects of transcriptional pausing on gene expression dynamics. PLoS Computational Biology, 6(3), e1000704. doi:10.1371/journal.pcbi.1000704

Revyakin, A., Ebright, R. H., & Strick, T. R. (2004). Promoter unwinding and promoter clearance by RNA polymerase: Detection by single-molecule DNA nanomanipulation. Proceedings of the National Academy of Sciences, 101(14), 4776–4780. doi:10.1073/pnas.0307241101

Ribeiro, A. S., Smolander, O.-P., Rajala, T., Häkkinen, A., & Yli-Harja, O. (2009). Delayed Stochastic Model of Transcription at the Single Nucleotide Level. Journal of Computational Biology, 16(4), 539–553. doi: 10.1089/cmb.2008.0153

Rhee, K. Y., Opel, M., Ito, E., Hung, S. p, Arfin, S. M., & Hatfield, G. W. (1999). 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, 96(25), 14294–14299. doi: 10.1073/pnas.96.25.14294

Rovinskiy, N., Agbleke, A. A., Chesnokova, O., Pang, Z., and Higgins, N. P. (2012). Rates of Gyrase Supercoiling and Transcription Elongation Control Supercoil Density in a Bacterial Chromosome. *PLoS Genet.*, *8*(8), e1002845. doi:10.1371/journal.pgen.1002845

Santinha, J., Martins, L., Häkkinen, A., Lloyd-Price, J., Oliveira, S. M. D., Gupta, A., Annila, T., Mora, A., Ribeiro, A.S. and Fonseca, J. R. (2016). *iCellFusion: Tool for Fusion and Analysis of Live-Cell Images from Time-Lapse Multimodal Microscopy*. doi:10.4018/978-1-4666-8811-7.ch004

Stracy, M., Lesterlin, C., Garza de Leon, F., Uphoff, S., Zawadzki, P., and Kapanidis, A. N. (2015). Live-cell superresolution microscopy reveals the organization of RNA polymerase in the bacterial nucleoid. *Proc. Natl. Acad. Sci.*, *112*(32), E4390–E4399. doi:10.1073/pnas.1507592112

Stracy, M., Wollman, A. J. M., Kaja, E., Gapinski, J., Lee, J.-E., Leek, V. A., McKie, S.J., Mitchenall, L.A., Maxwell,A., Sherratt, D.J., Leake, M,C., and Zawadzki, P. (2019). Single-molecule imaging of DNA gyrase activity in living *Escherichia coli*. *Nucleic Acids Res.*, *47*(1), 210–220. doi:10.1093/nar/gky1143

Startceva, S., Kandavalli, V. K., Visa, A., and Ribeiro, A. S. (2019). Regulation of asymmetries in the kinetics and protein numbers of bacterial gene expression. *Biochim. Biophys. Acta - Gene Regul. Mech.*, *1862*(2), 119–128. doi:10.1016/j.bbagrm.2018.12.005

Taniguchi, Y., Choi, P. J., Li, G. W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S. (2010). Quantifying *E. coli* Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. *Sci. (New York, NY)*, *329*(5991), 533–538. doi:10.1126/science.1188308

Teves, S. S., Henikoff, S. (2014). Transcription-generated torsional stress destabilizes nucleosomes.

Nat Struct Mol Biol 21(1), 88–94. doi:10.1038/nsmb.2723

Tran, H., Oliveira, S. M. D., Goncalves, N., and Ribeiro, A. S. (2015). Kinetics of the cellular intake of a gene expression inducer at high concentrations. *Mol. Biosyst.*, *11*(9), 2579–2587. doi:10.1039/C5MB00244C

Travers, A., & Muskhelishvili, G. (2005). DNA supercoiling — a global transcriptional regulator for enterobacterial growth? Nature Reviews Microbiology, 3(2), 157–169. doi:10.1038/nrmicro1088

Vogel, U. and Jensen, K. F. (1994). The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J. Bacteriol.*, *176*(10), 2807–2813. doi:10.1128/jb.176.10.2807-2813.1994

Weinstein-Fischer, D., Elgrably-weiss, M., & Altuvia, S. (2002). *Escherichia coli* response to hydrogen peroxide: a role for DNA supercoiling, Topoisomerase I and Fis. Molecular Microbiology, 35(6), 1413–1420. doi:10.1046/j.1365-2958.2000.01805.x

Wheeler, A. (2017). Digital Microscopy. In Standard and Super-Resolution Bioimaging Data Analysis (eds A. Wheeler and R. Henriques). doi:10.1002/9781119096948.ch1

Zimmer, C., Häkkinen, A., and Ribeiro, A. S. (2016). Estimation of kinetic parameters of transcription from temporal single-RNA measurements. *Math. Biosci.*, *271*, 146–153. doi:10.1016/j.mbs.2015.10.001

PUBLICATION III

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

V. Chauhan*, M.N.M. Bahrudeen*, C.S.D. Palma, I. Baptista, B.L.B. Almeida, S. Dash, V. Kandavalli, and A.S. Ribeiro. *Equal contributions.

PLOS Computational Biology, 18(1), 2022
DOI:10.1371/journal.pcbi.1009824

PLOS COMPUTATIONAL BIOLOGY

RESEARCH ARTICLE

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

Vatsala Chauhan[1☯], **Mohamed N. M. Bahrudeen**[1☯], **Cristina S. D. Palma**[1], **Ines S. C. Baptista**[1], **Bilena L. B. Almeida**[1], **Suchintak Dash**[1], **Vinodh Kandavalli**[2], **Andre S. Ribeiro**[1]*

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

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

Data Availability Statement: A data package was deposited in Dryad (Ref. [59] in main manuscript) under the DOI: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/CYS3FjMMhrs8aqPq4lLFsGunynao-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 Erkko 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'

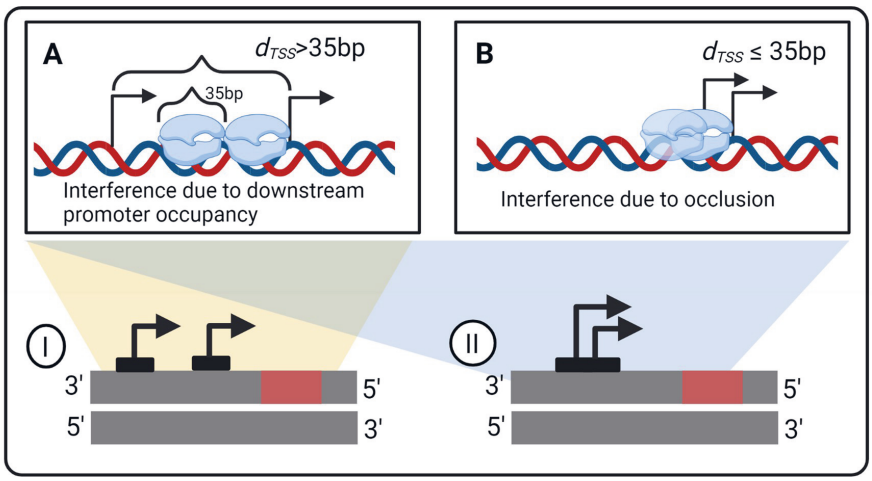

Fig 1. Interference between tandem promoters with different arrangements relative to each other. (A) Interference by an RNAP occupying the downstream promoter on the activity of the elongating RNAP from upstream promoter. The TSSs need to be at least 36 bp apart (the length occupied by an RNAP when in 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

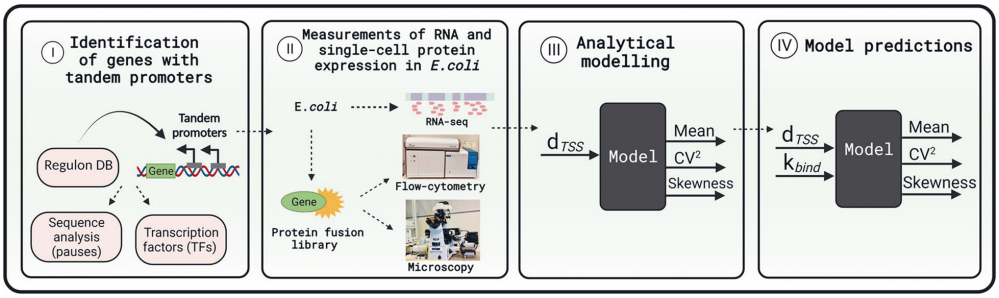

Fig 2. Workflow. (I) We identified genes controlled by tandem promoters in Regulon DB. (II) Next, we measured the single-cell protein levels of those genes with arrangements I and II that are tagged in the YFP strain library [28]. We also measured the mean RNA fold changes of these genes over time (S1 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 ω_d and (1- ω_d) in their rate constants, with ω_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

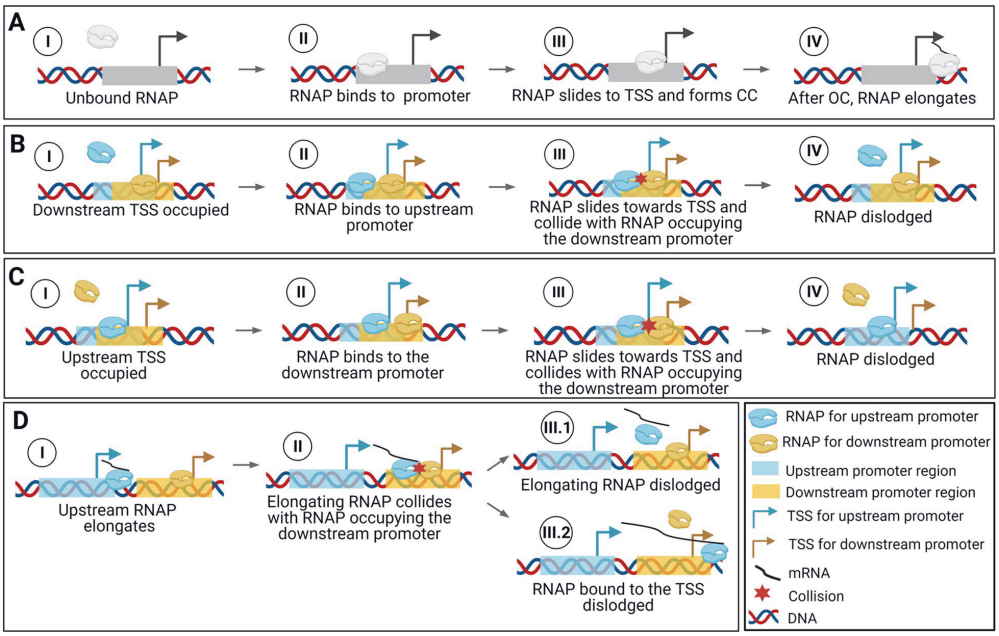

Fig 3. Events leading to transcriptional interference between tandem promoters. (**A**) Sequence of events in transcription in isolated promoters. A similar set of events occurs in tandem promoters, if only one RNAP interacts with them at any given time. (**B / C**) Interference due to the occlusion 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.

$$P^u_{free} \xrightarrow[\text{I}]{k^u_{bind}\cdot[R]} P^u_{bound} \xrightarrow[\text{II}]{k^u_{cc}} P^u_{cc} \xrightarrow[\text{III}]{k^u_{oc}} P^u_{oc} \xrightarrow[\text{IV}]{k^u_{escape}} P^u_{free} + R^u_{elong} \tag{1A1}$$

$$P^u_{cc} \xrightarrow{k_{unbind}} P^u_{free} \tag{1A2}$$

$$R^u_{elong} \xrightarrow{k^u_{elong}\cdot(1-\omega_d\cdot f)} RNA \tag{1A3}$$

$$R^u_{elong} \xrightarrow{k^u_{elong}\cdot\omega_d\cdot f} \emptyset \tag{1A4}$$

$$P^u_{bound} \xrightarrow{k^{u/d}_{occlusion}} P^u_{free} \tag{1A5}$$

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:

$$P^u_{free} \xrightarrow{k^u_{bind}\cdot[R]} P^u_{occupied} \xrightarrow{k^u_{after}} P^u_{free} + R^u_{elong} \tag{1B1}$$

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:

$$P^u_{free} \underset{k^{u/d}_{occlusion}+k^u_{unbind}}{\overset{k^u_{bind}\cdot[R]}{\rightleftharpoons}} P^u_{occupied} \xrightarrow{k^u_{after}} P^u_{free} + R^u_{elong} \tag{1C1}$$

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:

$$P^d_{free} \underset{k^{d/u}_{occlusion}+k^d_{unbind}}{\overset{k^d_{bind}\cdot[R]}{\rightleftharpoons}} P^d_{occupied} \xrightarrow{k^d_{after}} P^d_{free} + R^d_{elong} \tag{2A1}$$

$$R^d_{elong} \xrightarrow{k^d_{elong}} RNA \tag{2A2}$$

$$P^d_{free} \xrightarrow{k_{occupy}} P^d_{occupied} \tag{2A3}$$

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).

$$RNA \xrightarrow{k_p} Prot \tag{3}$$

$$RNA \xrightarrow{k_{rd}} \emptyset \tag{4A}$$

$$Prot \xrightarrow{k_{pd}} \emptyset \tag{4B}$$

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:

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

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

Here, $k_{ocl}^{\max}$ 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 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, ω 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")	$e^{-(b_1 \cdot d_{TSS})}$	$k_{ocl}^{\max} \cdot e^{-(b_1 \cdot d_{TSS})} \cdot \omega$
Exponential 2 ("Exp2")	$e^{-(b_1 \cdot d_{TSS} + b_2 \cdot d_{TSS}^2)}$	$k_{ocl}^{\max} \cdot e^{-(b_1 \cdot d_{TSS} + b_2 \cdot d_{TSS}^2)} \cdot \omega$
Step ("Step")	$1 - \frac{1}{1+e^{-m \cdot (d_{TSS} - L)}}$	$k_{ocl}^{\max} \cdot \left(1 - \frac{1}{1+e^{-(d_{TSS} - L)}}\right) \cdot \omega$, for m = 1 bp^{-1}
Zero order ("ZeroO")	k	$k_{ocl}^{\max} \cdot \omega$

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):

$$M_p = \frac{k_r \cdot k_p}{k_{rd} \cdot k_{pd}} \tag{8}$$

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:

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

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):

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

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:

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

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

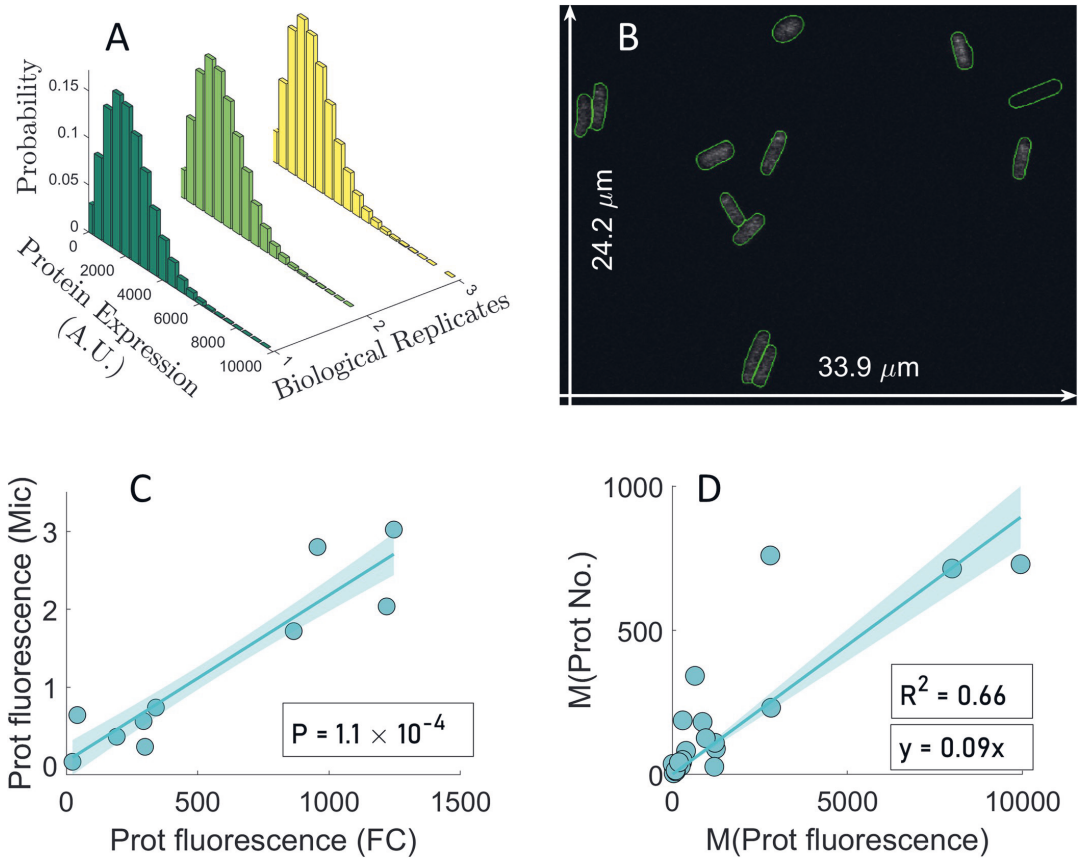

Fig 4. Single cell protein numbers by microscopy and flow-cytometry. (A) Example single-cell distributions (3 biological replicates) of fluorescence (in arbitrary units) of cells with a YFP tagged gene controlled by a pair of tandem promoters obtained by flow-cytometry, 'FC'. (B) Example confocal microscopy 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).

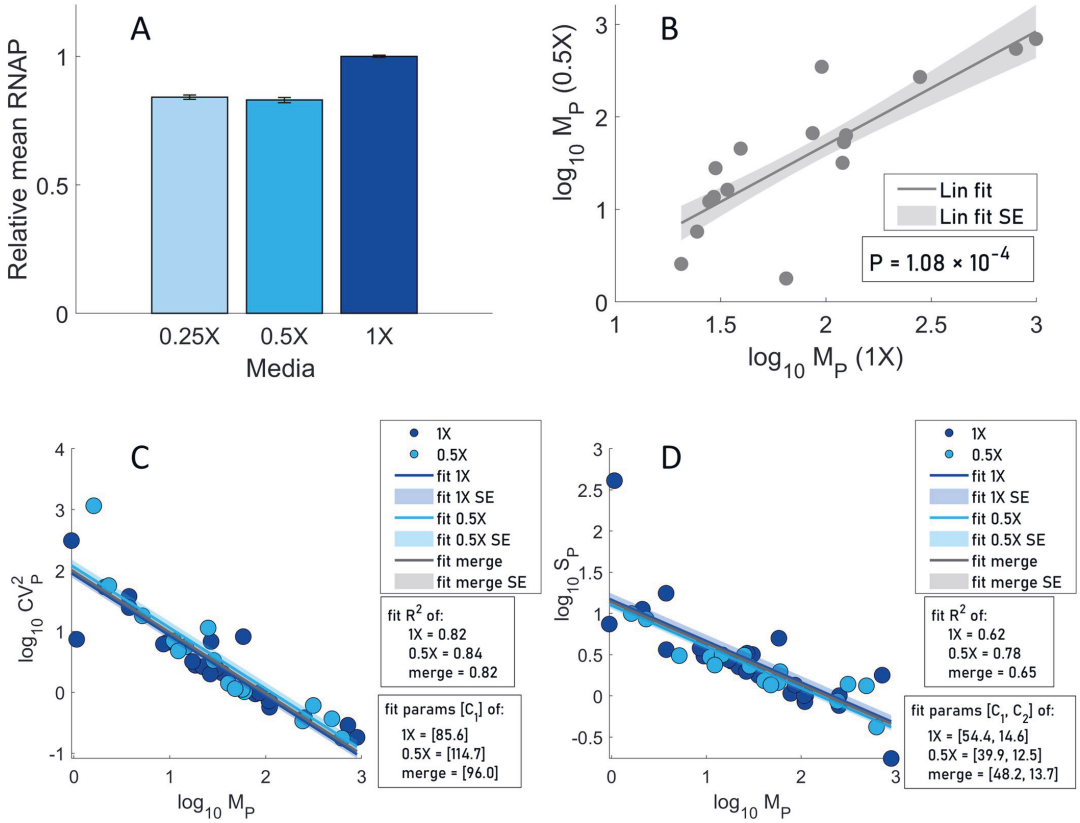

Fig 5. Relative RNAP concentrations along with the relationships between the moments of the single cell distributions of protein numbers. (A) Relative RNAP levels measured by flow-cytometry (Section 'flow-cytometry and data analysis' in the S1 Appendix) in three media. (B) Scatter plot between M_P in 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 s^{-1}	Differs between promoters. Since empirical data lacks, we used the data from *in vivo* single RNA measures for Lac-Ara-1 [20].
RNA and protein dilution due to division	$k_{dil} = \frac{\ln(2)}{D}$	1.005×10^{-4} s^{-1}	Legend of Fig H in the S2 Appendix.
RNA degradation	k_{rdeg}	2.3×10^{-3} s^{-1}	[28]
RNA decay due to dilution from cell division and due to degradation	$k_{rd} = k_{rdeg} + k_{dil}$	2.4×10^{-3} s^{-1}	From row 2.
Protein degradation	k_{pdeg}	2.93×10^{-5} s^{-1}	[46], estimates it to be from ~6×10^{-5} to ~2×10^{-5}. We used the value in [47], in that interval.
Protein decay due to dilution by cell division and degradation	$k_{pd} = k_{pdeg} + k_{dil}$	1.3×10^{-4} 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	$k_p = C_1\times(k_{pd}+k_{rd})$	0.18 s^{-1}	C_1 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 ~1000 free RNAPs per cell [41]. Meanwhile, for doubling times of 60 min in minimal medium, there are ~144 [40]. In both our media, we observed a doubling time of ~115 mins (Fig 5B). Thus, we expect the free RNAP in 1X to also be ~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 ~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 $k^{\max} = k^{\max}_{occupy} = k^{\max}_{ocl}$, 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^2_p 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^2_p 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^2_p, 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

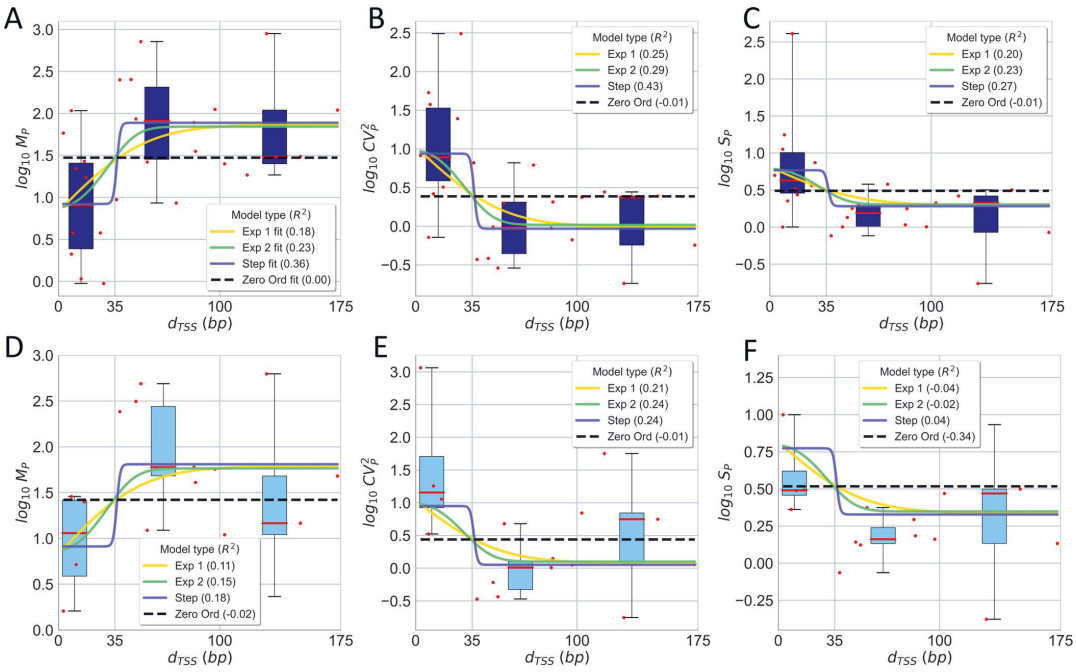

Fig 6. Empirical data and analytical model of how d_{TSS} influences the single-cell protein numbers of genes controlled by tandem promoters. **(A)** Mean, **(B)** CV^2, and **(C)** S of single protein numbers in the **1X** media as a function of d_{TSS}. **(D)**, **(E),** and **(F)** show the same for 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 3rd and 1st 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 $k_{bind}\cdot[R]$, 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 (M, CV^2, S) 1X medium	Average R^2 (M, CV^2, S) 0.5X medium
Exponential 1	k_{bind}·[R] = 1.09×10^{-2} s^{-1} × (cell vol)$^{-1}$ k_{bind} = 7.53×10^{-5} s^{-1} k_{unbind} = 0.84 s^{-1} k^{max} = 677.7 s^{-1} b_1 = 5.08×10^{-2} bp^{-1}	0.21 (Fig 6A–6C)	0.09 (Fig 6D–6F)
Exponential 2	k_{bind}·[R] = 9.71×10^{-3} s^{-1} × (cell vol)$^{-1}$ k_{bind} = 6.74×10^{-5} s^{-1} k_{unbind} = 0.80 s^{-1} k^{max} = 554.8 s^{-1} b_1 = 7.92×10^{-8} bp^{-1} b_2 = 1.47×10^{-3} bp^{-2}	0.25 (Fig 6A–6C)	0.12 (Fig 6D–6F)
Step	k_{bind}·[R] = 6.62×10^{-3} s^{-1} × (cell vol)$^{-1}$ k_{bind} = 4.60×10^{-5} s^{-1} k_{unbind} = 0.49 s^{-1} k^{max} = 313.4 s^{-1} L = 35.11 bp (by best fitting, which corresponds to 35 bp)	0.35 (Fig 6A–6C)	0.15 (Fig 6D–6F)
zero order	k_{bind}·[R] = 4.63×10^{-3} s^{-1} × (cell vol)$^{-1}$ k_{bind} = 3.22×10^{-5} s^{-1} k_{unbind} = 0.57 s^{-1} k^{max} = 6.48 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 (ω) 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 ≤ 35, and to 100 bp to represent promoters with $d_{TSS} > 35$. Then, for each cohort, we changed ω from 0.01 to 0.99 (i.e., nearly all possible values). In addition, we estimated these moments when $k_{occlusion}$, k_{occupy}, and ω 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 ω.

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

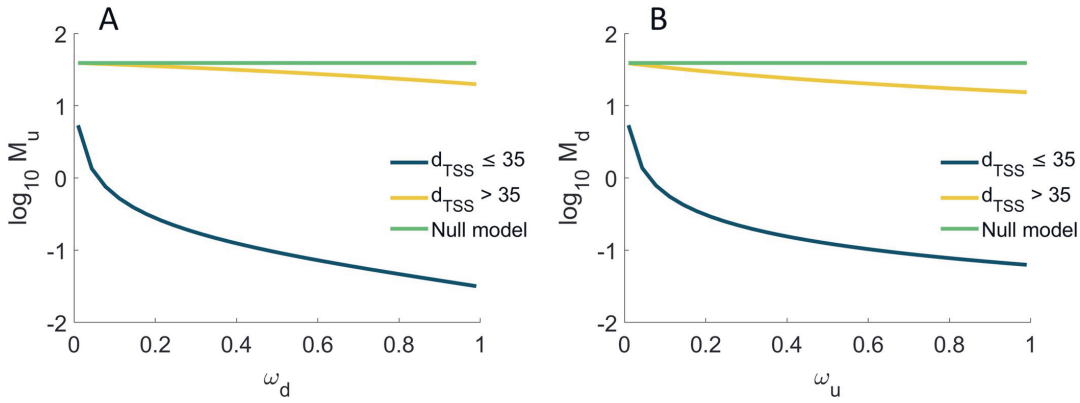

Fig 7. Mean protein numbers produced as a function of other promoter's occupancy. M_P of the single-cell distribution of the number of proteins produced (**A**) by the upstream promoter alone, and (**B**) by the downstream promoter alone. Results are shown as a function of the fraction of times that 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 ω 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, ω, 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 ω. Eqs 6A and 6B define the occupancies of the upstream and downstream promoters, ω_u and ω_d, respectively. For simplicity, here we refer to both of them as ω. Fig 8A shows that ω 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, ω 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, ω can change significantly within $10^{-2} < k_{bind}\times[R] < 10\ s^{-1}$ and $10^{-2} < k_{after}^{-1} < 10^{2}$ s. For these ranges, we expect RNA production rates (k_r, Eqs 5A, 5B, 6B, 7 and 9) to vary from ~10^{-5} (if $d_{TSS} \leq 35$) and ~10^{-4} (if $d_{TSS} > 35$) until 10 s^{-1}. In agreement, in *E. coli*, promoters have RNA production rates from ~10^{-3} to 10^{-1} s^{-1} when induced [20–21,39,50–51] and ~10^{-4} to 10^{-6} s^{-1} when non-fully active [28]. Thus, ω 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 ω 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 ω, when and only when $d_{TSS} \leq 35$ (Fig 8B).

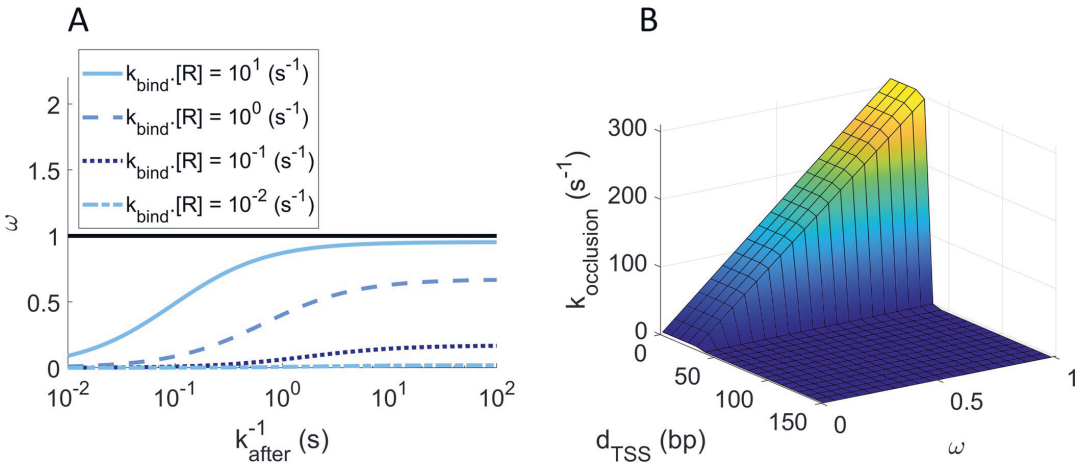

Fig 8. Promoter occupancy ω estimated for the step model. (A) ω as a function of the rate constant for a *free* RNAP to bind to the *unoccupied* promoter (k_{bind}·[R]) and of the time for that RNAP to start elongation after commitment to OC, k^{-1}_{after}. The horizontal black line at $\omega = 1$, 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 ω and d_{TSS}. Since $k_{occlusion}$ increases with ω 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^2_p, and S_P) of the upstream, 'u', and downstream, 'd', promoters changes with ω_u, ω_d, and d_{TSS}. Here, ω_u and ω_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 ω_d while also increasing ω_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, ω_d, becomes ineffective.

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

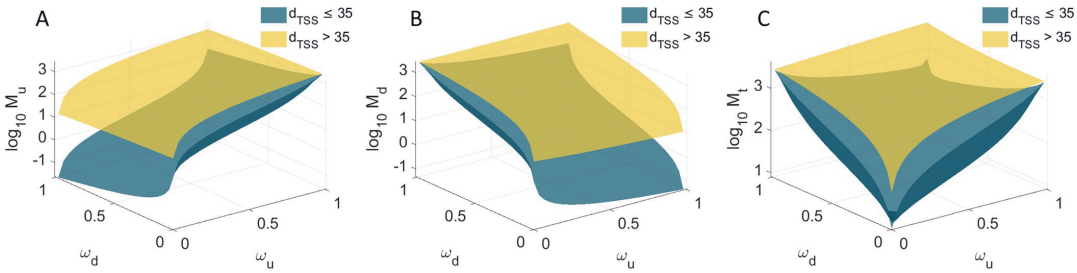

Fig 9. Mean protein expression as a function of both promoters' occupancy. Expected mean protein numbers due to the activity of: (**A**) the upstream promoter alone, (**B**) the downstream promoter alone, and (**C**) both promoters. M_P is shown as a function of the fraction of times that the upstream ($0 \leq \omega_u \leq 1$) 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, ω_u, becomes ineffective.

Finally, from Fig 9C, regardless of d_{TSS}, for small ω_d and ω_u, as the occupancies increase, M_t increases quickly and in a non-linear fashion. However, as both ω_d and ω_u reach high values, M_t decreases for further increases, if $d_{TSS} \leq 35$ bp. Instead, if $d_{TSS} > 35$ bp, Mt 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 (~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 ω, 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 ω 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 ω 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 30th 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.

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 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. Korbel 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. Biotechnol 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, Yli-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 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 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.bbagrm.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, Kortkhonjia 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 PMID: 11893332

34. Margeat E, Kapanidis AN, Tinnefeld P, Wang Y, Mukhopadhyay J, Ebright 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 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 PMID: 3542569

39. Kandavalli VK, Tran H, Ribeiro AS. Effects of σ factor competition are promoter initiation kinetics dependent. Biochim Biophys Acta. 2016; 1859: 1281–1288. https://doi.org/10.1016/j.bbagrm.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 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.bbagrm.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-Tejeida 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

S1 Appendix: Extended Materials and Methods

Selection of natural genes controlled by tandem promoters

We define a pair of tandem promoters as two promoters in a head-to-tail formation transcribing the same gene, as in [1]. In order to find them in the genome of *E. coli*, from RegulonDB, we obtained the lists of all known transcription units (TUs), promoters (defined as stretches of 60 upstream and 20 downstream nucleotide sequences from a TSS), gene sequences, TFs, and terminators [2].

From the list of TUs (3560), we extracted all genes (510) under the control of two and only two promoters in tandem formation with known TSS and DNA strand (information from the promoters' list). Then, we calculated the nucleotide distance between their pair of TSSs (d_{TSS}) and obtained the start and end positions of their sequence in the DNA. As a side note, we found additional 321 genes controlled by more than two promoters in tandem formation, which are not accounted for as they are not included in the model, for simplicity.

Next, we removed all genes with another gene or promoter sequence (associated to a TU) located in the opposing strand anywhere between the start of the upstream promoter and the end of the gene sequence (186 out of 510) since their dynamics may be subject to interference from convergent RNAPs [1,3,4]

Out of the remaining 324 genes, only 152 are in the first position of a TU or in a TU with only one gene. Since evidence suggests that the existence of multiple genes in a TU influences their transcription significantly, due to premature terminations, distance to the promoter etc. [5,6], we opted for keeping only those 152 genes. Subsequently, from the list of terminators, we obtained their start and end positions and DNA strand and filtered out (9 out of 152) genes with a terminator sequence in between the beginning of the upstream promoter and the end of the gene sequence, due to potential enhanced premature terminations. Finally, from these, we only considered promoter pairs (102 out of the 143 genes) such that no gene is coded in the regions containing them or the space in between them (Fig 1), so that elongation of other genes do not perturb their transcription.

Finally, of these 102 genes, we measured the expression levels at the single-cell level of 30 of them (Table A in S3 Appendix) using a YFP strain library [7]. These genes are of the categories 'I' (9 genes) and 'II' (21 genes) in Fig 1. Their d_{TSS} range from 84 to 173, and from 3 to 73 nucleotides, respectively.

Selection of natural genes controlled by single promoters

To select natural genes controlled by single promoters in the genome of *E. coli,* from RegulonDB, we obtained the lists of all known transcription units (TUs), promoters, gene sequences and terminators [2]. From the list of TUs (3560), we extracted all genes (1760) under the control of one and only one promoter with known TSS and DNA strand (information from the promoters' list). Next, we filtered out all genes with another gene or promoter sequence (associated to a TU) located in the opposing strand

anywhere between the start of the promoter and the end of the gene sequence (446 out of 1760) since their dynamics may be subject to interference from convergent RNAPs [1,3,4] Out of the remaining 1314 genes, only 649 are in the first position of a TU or in a TU with only one gene and no other promoter sequence (associated to another TU) between the promoter and the end of the gene of interest. Since evidence suggests that the existence of multiple genes in a TU influences their transcription significantly, due to premature terminations, distance to the promoter etc. [5,6], we opted for keeping only those 649 genes. Subsequently, from the list of terminators, we obtained their start and end positions and DNA strand and filtered out (36 out of 649) genes with a terminator sequence in between the promoter and the end of the gene sequence, due to potential enhanced premature terminations. Finally, of these 613 genes, we obtained data on the expression levels of 126 genes from [7], which we used to compare expression levels of genes controlled by tandem promoters and genes controlled by single promoters.

Meanwhile, for purposes of validating the scaling factor between protein fluorescence and numbers, of these 613 genes, we measured the expression levels at the single-cell level of 10 of them, randomly selected (Table B in S3 Appendix) [7].

Gene Conservation

From a list of 5443 reference bacterial genomes [8], we used the Rentrez package [9] to obtain which genes are present in each genome. Next, we removed those genomes without gene entries (1310). Using the remaining genomes, we estimated the evolutionary conservation of each gene in the genome of MG1655 (GCF_000005845.2_ASM584v2), including those controlled by tandem promoters, by the ratio between the number of genomes where the gene is present, and the total number of genomes considered. Fig Q in S2 Appendix shows the conservation levels as a function of d_{TSS} of the tandem promoters controlling the genes' expression.

Gene Ontology (GO)

For gene ontology representations, we performed overrepresentation tests using the PANTHER Classification System [10], which finds statically significant overrepresentations using Fisher's exact tests. For p-values < α (here set to 0.05), the null hypothesis that there are no associations between the gene cohort and the corresponding GO of the biological process is rejected, which we interpret as the gene cohort being associated with corresponding GO of the biological process.

Network topological properties

By 'network topological property' we refer to some feature of a gene that is related to how that gene is integrated with the network formed by TFs linking genes. We used *Cytoscape* [11] to extract these features for the genes controlled by tandem promoters from the *known* transcription factor (TF) network

of *E. coli*, using information from RegulonDB v10.5 on all known transcription factors (TFs) and their binding sites [2].

Next, for the two cohorts of genes with d_{TSS} larger or not than 35 bps, based on definitions in [12], we calculated (Table C in S3 Appendix) the mean and standard error of each cohort's average shortest path length (minimum number of edges between pairs of genes), clustering coefficient (fraction of input nodes to a node that are also linked), eccentricity (maximum non-infinite shortest path length between the node and another node in the network), edge count (number of edges/nodes that are connected to the node), indegree (number of incoming edges), neighbourhood connectivity (average connectivity of all nearest neighbours), and outdegree (number of outgoing edges).

For each feature, we also obtained a p-value, which is the probability that the genes of the cohort have a smaller mean than the mean from all genes of *E. coli*. This probability is estimated from 10^5 cohorts assembled from random samples from all genes with replacement, using a non-parametric bootstrap method. The sample size is equal to the size of the cohort being compared with.

Media and chemicals

Measurements were performed in Luria-Bertani (LB) and M9 media (standard and diluted). The chemicals, such as tryptone, sodium chloride, agarose, MEM amino acids (50X), MEM Vitamin solution (100X), Glucose and antibiotic chloramphenicol, etc. were purchased from Sigma Aldrich. Yeast extract was purchased from Lab M (Topley House, Bury, Lancashire, UK). The components of LB medium were 10 g tryptone, 10 g NaCl, and 5 g yeast extract in 1000 mL distilled water. For M9 medium, the components were 1x M9 Salts, 2 mM MgSO4, 0.1 mM CaCl2; 5x M9 Salts with 34 g/L Na2HPO4, 15 g/L KH2PO4, 2.5 g/L NaCl, 5 g/L NH4Cl supplemented with 100X vitamins, 0.2% Casamino acids and 0.4% glucose. We also used '0.5X' and '0.25X' media by diluting the M9 medium to 1:1 and to 1:3 respectively, using autoclaved distilled water [13-16].

Strains and growth conditions

To measure RNA polymerase (RNAP) levels at different medium, we used the RL1314 strain with RpoC endogenously tagged with GFP (generously provided by Robert Landick), which was engineered from the W3110 strain (used here to measure background fluorescence).

To measure single-cell protein levels of genes controlled by tandem promoters, we used genes endogenously tagged with the YFP coding sequence from the YFP fusion library [7]. These were purchased from the *E. coli* genetic stock center (CGSC) of Yale University, U.S.A. (Table B in S3 Appendix), which has wild type MG1655 cells as the reference genome (and thus was used to measure cellular background fluorescence). Measurements of protein levels using this library are expected to be precise for a wide range of expression levels, given evidence for strong correlation in single gene expression levels when measured by RNA-fish, RNA-seq, mass spectrometry and flow cytometry (taken using the YFP library) [7]. The lesser accurate estimations occur for the weakest expressing genes

[7][17], due to their values being near the level of cellular autofluorescence. For this reason as well, we do not consider all of the 30 genes in our analysis as described in the Results section.

From a glycerol stock (-80°C), cells were streaked on LB agar plates with the appropriate antibiotics and incubated at 37°C overnight. From the plates, a single colony was picked, inoculated in LB medium and supplemented with appropriate antibiotics and incubated at 30°C overnight with shaking at 250 rpm. Next, overnight cultures were diluted into freshly prepared tailored media (see 'Media and Chemicals'), with appropriate antibiotics with an $O.D_{600}$ of 0.03 (Optical Density, 600 nm; Ultrospec 10, Amersham biosciences, UK) and allowed to grow at 30°C with shaking at 250 rpm until reaching the mid-exponential phase ($O.D_{600}$ ~0.4-0.5). At this stage, measurements of protein levels were conducted using flow-cytometry and/or microscopy.

Growth curves

Growth curves were measured by $O.D_{600}$ using a spectrophotometer (Ultrospec 10; GE Healthcare). From the overnight culture, cells were diluted (1:10000) into the respective fresh media and allowed to grow while shaking (250 rpm). O.D.'s were recorded for 450 min. every 30 min. We performed 3 biological replicates for each condition. We found negligible variability between replicates. The results shown are the averages and standard error of the mean.

Microscopy and image analysis

When reaching the mid-exponential growth phase, cells were pelleted by centrifugation (10000 rpm for 1 min), and the supernatant was discarded. The pellet was re-suspended in 100 µL of the remaining medium Next, 3 µL of cells were placed in between 2% agarose gel pad and a coverslip and imaged using a confocal microscopy with a 100X objective. The fluorescence was measured with a 488 nm laser and a 514/30 nm emission filter. Phase-contrast images were simultaneously acquired for purposes of segmentation and to assess health, morphology, and physiology.

Using the software *CellAging* [18], from phase contrast images, we segmented cells semi-automatically, correcting errors manually. Next, phase-contrast and corresponding fluorescence images were aligned to extract single-cell fluorescence intensities (example image in Fig 4B). We then performed background subtraction, i.e., from each cell's total fluorescence we subtracted the mean fluorescence of control cells, not expressing YFP.

RNA-seq measurements and data analysis

We searched for correlations between the LFCs over time of genes controlled by tandem promoters ('Tg') and the LFCs over time of their output genes ('Og') as well as their input genes ('Ig').

Given known rates of RNA and protein production and degradation in *E. coli* [7, 19-22], we expect changes in RNA numbers to take at least 60 min. on average, to propagate to protein numbers. Thus,

we performed RNA-seq of cells in exponential growth phase at moments '0 min', and then 20 and 180 mins. later. We then calculated LFCs between 0 and 20 min, and between 0 and 180 min.

Specifically, to assess if LFCs in Ig propagate to Tg, we compared changes in Ig between moments 0 and 20, with changes in Tg between moments 0 and 180 min. Similarly, to assess LFCs in Tg propagate to Og, we compared changes in Tg between moments 0 and 20, with changes in Og between moments 0 and 180 min. Results are shown in Panels A and B of Fig D in S2 Appendix.

Sample preparation

For RNA-seq experiments, single colonies of K12 MG1655 cells were picked from LB Agar plates and inoculated into 5 ml of LB medium. Cultures were grown overnight with shaking at 250 rpm. Next, these cultures were diluted to $O.D_{600}$ of 0.05 in fresh LB medium and incubated, with a 250 rpm agitation. RNA-seq was performed over time (0, 20 and 180 min). Total RNA from 3 independent biological replicates in each medium was extracted using RNeasy kit (Qiagen). RNA was treated twice with DNase (Turbo DNA-free kit, Ambion) and quantified using Qubit 2.0 Fluorometer RNA assay (Invitrogen, Carlsbad, CA, USA). Total RNA amounts were determined by gel electrophoresis, using a 1% agarose gel stained with SYBR safe (Invitrogen). RNA was detected using UV with a Chemidoc XRS imager (Biorad).

Sequencing was performed by GENEWIZ, Inc. (Leipzig, Germany). The RNA integrity number (RIN) was obtained with the Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA). Ribosomal RNA depletion was performed using Ribo-Zero Gold Kit (Bacterial probe) (Illumina, San Diego, CA, USA). RNA-seq libraries were constructed using NEBNext Ultra RNA Library Prep Kit (NEB, Ipswich, MA, USA). Sequencing libraries were multiplexed and clustered on 1 lane of a flow-cell. Samples were sequenced using a single-index, 2x150 bp paired-end (PE) configuration on an Illumina HiSeq instrument. Image analysis and base calling were conducted with HiSeq Control Software (HCS). Raw sequence data (.bcl files) were converted into fastq files and de-multiplexed using Illumina bcl2fastq v.2.20. One mismatch was allowed for index sequence identification.

Data analysis

RNA-seq data analysis pipeline was: i) RNA sequencing reads were trimmed with Trimmomatic [23] v.0.39 to remove possible adapter sequences and nucleotides with poor quality. ii) Trimmed reads were mapped to the reference genome, *E. coli* MG1655 (NC_000913.3), using the using the STAR aligner v.2.5.2b, which outputs BAM files [24]. iii) Then, '*featureCounts'* from the Rsubread R package v.1.34.7 was used to calculate unique gene hit counts [25]. iv) These counts were used for the differential expression analysis. Genes with less than 5 counts in more than 3 samples, and genes whose mean counts are less than 10 were removed from further analysis. We used the DESeq2 R package v.1.24.0 [26] to compare gene expression between groups of samples and calculate p-values and $\log_2$ of fold changes using Wald tests (function *nbinomWaldTest*). P-values were adjusted for multiple hypotheses testing (Benjamini–Hochberg, BH procedure, [27]).

Flow-cytometry and data analysis

We measured single-cell fluorescence using a ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA). Upon reaching the mid-exponential phase (OD~0.4-0.5), cells were diluted (1:10000) into 1 mL of phosphate buffer saline (PBS) solution and vortexed for 5 s. For a single run, 50000 events were collected at a flow rate of 14 μL/minute and a core diameter of 7.7 mm using the Novo Express software using a blue laser (488 nm) for excitation. We obtained the height of the fluorescein isothiocyanate channel (FITC-H) (530/30 nm filter). A PMT voltage of 600 volts was set for FITC. To avoid background signal from particles smaller than bacteria, the detection threshold was set to 5000 for FSC-H analyses. Three biological replicates were performed per condition.

We applied unsupervised gating [28] to the flow-cytometry data, setting the fraction of single-cell events used in the analysis, α, to 0.99. We proved to be enough to remove non-cell events due to debris, doublets, fragments, cell clumps, and other undesired events. Reducing α did not change the results qualitatively.

To remove outliers from the flow-cytometry distributions, we applied secondary gating. In detail, we sorted the data based on FITC-H values and calculated the difference between consecutive samples. Then, we obtained the indices of those differing by more than 10000 (approximately 10 times the mean fluorescent level observed). Next, we obtained the minimum of those indices to define the upper bound. Finally, values above this index were considered an outlier and discarded. In all measurements, never more than 10000 events were discarded, thus, more than 40000 were used for the analysis.

Subtraction of background fluorescence from total protein fluorescence in flow-cytometry

First, we collected mean background fluorescence from distributions of cells not carrying YFP. Then we measured the distributions of fluorescence of cells carrying the protein tagged with YFP. Having this, the protein fluorescence ‘g’ of a gene is obtained by subtracting mean background fluorescence ‘bg’ from the (total ‘T’) measured fluorescence. For the mean (M) protein fluorescence from a cell population, we write:

$$M(g) = M(T) - M(bg) \tag{1}$$

Similarly, the variance ‘Var’ is obtained by:

$$Var(g) = Var(T) - Var(bg) \tag{2}$$

The CV^2 of the distribution protein fluorescence of a gene after background subtraction is:

$$CV^2(g) = \frac{Var(g)}{M(g)^2} \quad (3)$$

Finally, the third moment of protein fluorescence and the skewness after background subtraction are given by:

$$\mu_3(g) = \mu_3(T) - \mu_3(bg) \quad (4)$$

$$S(g) = \frac{\mu_3(g)}{Var(g)^{\frac{3}{2}}} \quad (5)$$

After background subtraction, any genes with negative means, variance or third moment, will not be included in the data (except in Fig F in S2 Appendix for illustrative purposes).

Conversion of protein fluorescence into protein numbers

To convert protein fluorescence into protein numbers, we made a correlation plot between the mean protein fluorescence measured in our lab (after background subtraction) and the mean protein numbers reported in [7] for the same genes. We fitted a line to the data points by forcing the intercept with the Y axis to be at zero. The slope of the fitted line is used as a scaling factor (~0.09) with an R^2 value of 0.68 (Fig 4D). For protein fluorescence to protein numbers correction only the mean gets changed whereas the normalised moments CV^2 and S remain unchanged.

Analytical model of mean RNA levels controlled by a single promoter in the absence of a closely spaced promoter

From Reactions 1c1 and 1a4 in the main manuscript, for an isolated promoter, one would have:

$$P_{free} \underset{k_{unbind}}{\overset{k_{bind}\cdot[R]}{\rightleftarrows}} P_{occupied} \xrightarrow{k_{after}} P_{free} + R_{elong} \quad (6)$$

$$R_{elong} \xrightarrow{k_{elong}} RNA \quad (7)$$

At steady state $P_{occupied}$ is:

$$\frac{dP_{occupied}}{dt} = P_{free} \times k_{bind} \cdot [R] - P_{occupied} \times \left(k_{unbind} + k_{after}\right) = 0 \quad (8)$$

$$P_{free} = P_{occupied} \cdot \frac{\left(k_{unbind} + k_{after}\right)}{k_{bind} \cdot [R]} \quad (9)$$

Since necessarily:

$$P_{free} + P_{occupied} = 1 \quad (10)$$

From equations 9 and 10:

$$P_{occupied} \cdot \left(1 + \frac{k_{unbind} + k_{after}}{k_{bind} \cdot [R]}\right) = 1 \quad (11)$$

$$P_{occupied} = \frac{k_{bind} \cdot [R]}{k_{bind} \cdot [R] + k_{unbind} + k_{after}} \quad (12)$$

Note that, by definition (main manuscript, equations 6a and 6b), the fraction of time that an RNAP is bound to the promoter, ω, should equal $P_{occupied}$ in (12). Meanwhile, at steady state, R_{elong}becomes:

$$\frac{dR_{elong}}{dt} = P_{occupied} \times k_{after} - R_{elong} \times k_{elong} = 0 \quad (13)$$

$$R_{elong} = \frac{P_{occupied} \times k_{after}}{k_{elong}} \quad (14)$$

From equations 12 and 14:

$$R_{elong} = \frac{k_{bind} \cdot [R]}{k_{bind} \cdot [R] + k_{unbind} + k_{after}} \times \frac{k_{after}}{k_{elong}} \quad (15)$$

At steady state, the mean *RNA* numbers, M_{RNA}, is:

$$\frac{dM_{RNA}}{dt} = R_{elong} \times k_{elong} - M_{RNA} \times k_{rd} = 0 \quad (16)$$

From equations 15 and 16s:

$$M_{RNA} = \frac{k_{bind} \cdot [R]}{k_{bind} \cdot [R] + k_{unbind} + k_{after}} \times \frac{k_{after}}{k_{elong}} \times \frac{k_{elong}}{k_{rd}} \quad \text{(S7)}$$

$$M_{RNA} = \frac{k_{bind} \cdot [R]}{k_{bind} \cdot [R] + k_{unbind} + k_{after}} \times \frac{k_{after}}{k_{rd}} \quad (18)$$

From S18, the RNA numbers at steady state do not depend on k_{elong}.

Derivation of mean protein numbers at steady state produced by a pair of tandem promoters

For the upstream promoter, from (1c1), (1a3), and (1a4) in the main manuscript, at steady state:

$$\frac{d(RNA)}{dt} = R^u_{elong} \times k^u_{elong} \cdot (1 - \omega_d \cdot f) - RNA \times k_{rd} = 0 \quad (19)$$

From this and equation 6b in the main manuscript:

$$RNA = \frac{k^u_{bind} \cdot [R]}{k^{u/d}_{occlusion} + k^u_{bind} \cdot [R] + k^u_{unbind} + k^u_{after}} \times \frac{k^u_{after} \cdot (1 - \omega_d \cdot f)}{k_{rd}} \quad (20)$$

Meanwhile, for the downstream promoter, from reactions (2a1), (2a2), and (2a3) in the main manuscript, at steady state:

$$\frac{d(RNA)}{dt} = R^d_{elong} \times k^d_{elong} - RNA \times k_{rd} = 0 \quad (21)$$

$$RNA = \frac{k^d_{bind} \cdot [R]}{k^{d/u}_{occlusion} + k^d_{occupy} + k^d_{bind} \cdot [R] + k^d_{unbind} + k^d_{after}} \times \frac{k^d_{after}}{k_{rd}} \quad (22)$$

Having this, since at steady state the RNA numbers produced by a pair of tandem promoters should equal the sum of RNA numbers from the upstream (S20) and downstream (S22) promoters, we have:

$$M_{RNA} = \left(\frac{k^{u}_{bind} \cdot [R] \times k^{u}_{after} \cdot (1 - \omega_d \cdot f)}{k^{u/d}_{occlusion} + k^{u}_{bind} \cdot [R] + k^{u}_{unbind} + k^{u}_{after}} + \frac{k^{d}_{bind} \cdot [R] \times k^{d}_{after}}{k^{d/u}_{occlusion} + k_{occupy} + k^{d}_{bind} \cdot [R] + k^{d}_{unbind} + k^{d}_{after}} \right) \cdot \frac{1}{k_{rd}} \quad (23)$$

Thus, the mean protein numbers is:

$$M_P = M_{RNA} \cdot \frac{k_p}{k_{pd}} \quad (24)$$

If the upstream and downstream promoters have similar strengths, i.e., if $k^{d}_{bind} \approx k^{u}_{bind}$, $k^{d}_{unbind} \approx k^{u}_{unbind}$, and $k^{d}_{after} \approx k^{u}_{after}$, we can expect that: $\omega_d = \omega_u$, $k^{d/u}_{occlusion} = k^{u/d}_{occlusion}$. If so, the equation above becomes:

$$M_P = \left(\frac{k_{bind} \cdot [R] \times k_{after} \cdot (1 - \omega_d \cdot f)}{k_{occlusion} + k_{bind} \cdot [R] + k_{unbind} + k_{after}} + \frac{k_{bind} \cdot [R] \times k_{after}}{k_{occlusion} + k_{occupy} + k_{bind} \cdot [R] + k_{unbind} + k_{after}} \right) \cdot \frac{k_p}{k_{rd} \cdot k_{pd}} \quad (25)$$

Here, the symbols “u” and “d” are removed, as they no longer imply potentially different amounts. Having this, let k_r be the effective transcription rate constant of a pair of tandem proteins. It should equal:

$$k_r = \left(\frac{k_{bind} \cdot [R] \times k_{after} \cdot (1 - \omega_d \cdot f)}{k_{occlusion} + k_{bind} \cdot [R] + k_{unbind} + k_{after}} + \frac{k_{bind} \cdot [R] \times k_{after}}{k_{occlusion} + k_{occupy} + k_{bind} \cdot [R] + k_{unbind} + k_{after}} \right) \quad (26)$$

Thus, from equation 25 and 26:

$$M_P = \frac{k_r \cdot k_p}{k_{rd} \cdot k_{pd}} \quad (27)$$

CV^2 and skewness of the distribution of single-cell protein numbers of model tandem promoters

The distributions of protein numbers in *E. coli* cells, can, in general, be well approximated by a Gamma or by a negative binomial distribution [7]. We assume here a negative binomial distribution. For a given number of events, if r is the number of failures, p is the probability of success per event, and an 'event' is an attempt to produce a protein, then the mean, variance, and skewness of the single-cell distribution of protein numbers should equal:

$$M_P = \frac{pr}{1-p} \tag{28}$$

$$Var_P = \frac{pr}{(1-p)^2} \tag{29}$$

$$S_P = \frac{1+p}{\sqrt{pr}} \tag{30}$$

The relationship between the mean, CV^2 could be written as:

$$CV_P^2 = \frac{1}{M_P} \cdot \left(\frac{Var_P}{M_P} \right) \tag{31}$$

Substituting (S28) and (S29) in (S31)

$$CV_P^2 = \frac{\left(\frac{1}{1-p} \right)}{M_P} \tag{32}$$

Rewriting the above equation by assuming a scaling factor C_1 as:

$$C_1 = \frac{1}{1-p} \tag{33}$$

$$CV_P^2 = \frac{C_1}{M_P} \tag{34}$$

Taking $\log_{10}$ on both sides

$$\log_{10}\left(CV_P^2\right)=\log_{10}(C_1)-\log_{10}\left(M_P\right) \tag{35}$$

From [17], C_1 is approximated as

$$C_1=\frac{M_P}{M_{RNA}}\cdot\frac{\frac{1}{\tau_p}}{\frac{1}{\tau_p}+\frac{1}{\tau_{RNA}}} \tag{36}$$

$\tau_p=\frac{1}{k_{pd}}$ and $\tau_{RNA}=\frac{1}{k_{rd}}$ are the lifetimes of proteins and RNAs, respectively. The above equation is rewritten as:

$$C_1=\frac{k_p}{k_{pd}}\cdot\frac{k_{pd}}{k_{pd}+k_{rd}} \tag{37}$$

$$C_1=\frac{k_p}{k_{pd}+k_{rd}} \tag{38}$$

From (S28) and (S30), the relationship between the mean, skewness could be written as:

$$S_P=\frac{\frac{1+p}{\sqrt{1-p}}}{\sqrt{M_P}} \tag{39}$$

The equation can be rewritten assuming constant C_2 as:

$$C_2=\frac{1+p}{\sqrt{1-p}} \tag{40}$$

$$S_P=\frac{C_2}{\sqrt{M_P}} \tag{41}$$

Taking $\log_{10}$ on both sides

$$\log_{10}\left(S_P\right)=\log_{10}\left(C_2\right)-\frac{1}{2}\cdot\log_{10}\left(M_P\right) \tag{42}$$

The constants C_1 and C_2 are related as follows. From equation 33:

$$p = 1 - \frac{1}{C_1} \tag{43}$$

Inserting S43 in S40:

$$C_2 = \frac{2 - \frac{1}{C_1}}{\sqrt{\frac{1}{C_1}}} \tag{44}$$

The equation can be rewritten as

$$C_2 = 2\sqrt{C_1} - \frac{1}{\sqrt{C_1}} \tag{45}$$

Stochastic simulations for the step inference model

Stochastic simulations of the models were done using the stochastic gene network simulator SGNS2 [29]. These stochastic models were compared to the analytical solutions to assess how much variability can there be in $k_{bind} \cdot [R]$ without the analytical solution deviating too much.

First, to compare analytical and stochastic solutions, we set d_{TSS} between 0 and 180 with an increment of 30. For each d_{TSS}, we calculated the occlusion rate constant ($k_{occlusion}$) for upstream and downstream promoters (Equations 5a and 5b in the main manuscript). The other parameters are listed in Tables 2 and 3 in the main manuscript. To obtain protein numbers at steady state, we have set the simulation time to 10^5 seconds and performed 1000 runs per condition. From these runs, for each condition, we calculated the mean, CV^2 and skewness, along with their standard errors using bootstrapping (10^4 resampling with replacement). Additional runs would slightly decrease the deviation between the two solutions.

References

1. Shearwin KE, Callen BP, Egan JB. Transcriptional interference--a crash course. Trends Genet. 2005;21: 339–345. doi: 10.1016/j.tig.2005.04.009
2. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida 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. doi:10.1093/nar/gky1077

3. Crampton N, Bonass WA, Kirkham J, Rivetti C, Thomson NH. Collision events between RNA polymerases in convergent transcription studied by atomic force microscopy. Nucleic Acids Res. 2006;34: 5416–5425. doi:10.1093/nar/gkl668
4. Ward DF, Murray NE. Convergent transcription in bacteriophage λ: Interference with gene expression. J Mol Biol. 1979;133: 249–266. doi:10.1016/0022-2836(79)90533-3
5. Lewin B. Genes IX. 9th ed. Sudbury, Mass: Jones and Bartlett Publishers; 2008.
6. Turnbough CL Jr. Regulation of bacterial gene expression by transcription attenuation. Microbiol Mol Biol Rev. 2019;83. doi:10.1128/MMBR.00019-19
7. 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. doi:10.1126/science.1188308
8. Xavier JC, Gerhards RE, Wimmer JLE, Brueckner J, Tria FDK, Martin WF. The metabolic network of the last bacterial common ancestor. Commun Biol. 2021;4: 413. doi:10.1038/s42003-021-01918-4
9. Winter D. rentrez: An R package for the NCBI eUtils API. R J. 2017;9: 520. doi:10.32614/rj-2017-058
10. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47: D419–D426. doi:10.1093/nar/gky1038
11. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 2003;13: 2498–2504. doi:10.1101/gr.1239303
12. Doncheva NT, Assenov Y, Domingues FS, Albrecht M. Topological analysis and interactive visualization of biological networks and protein structures. Nat Protoc. 2012;7: 670–685. doi:10.1038/nprot.2012.004
13. 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. doi:10.1093/dnares/dsw009
14. Kandavalli VK, Tran H, Ribeiro AS. Effects of σ factor competition are promoter initiation kinetics dependent. Biochim Biophys Acta. 2016;1859: 1281–1288. doi:10.1016/j.bbagrm.2016.07.011
15. 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. doi:10.1016/j.bbagrm.2018.12.005
16. Oliveira SMD, Goncalves NSM, Kandavalli VK, Martins L, Neeli-Venkata R, Reyelt J, et al. Chromosome and plasmid-borne P LacO3O1 promoters differ in sensitivity to critically low temperatures. Sci Rep. 2019;9: 1–15. doi:10.1038/s41598-019-39618-z
17. 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. doi:10.1038/ng1807

18. Häkkinen A, Muthukrishnan A-B, Mora A, Fonseca JM, Ribeiro AS. CellAging: a tool to study segregation and partitioning in division in cell lineages of Escherichia coli. Bioinformatics. 2013;29: 1708–1709. doi:10.1093/bioinformatics/btt194
19. Bernstein JA, Khodursky AB, Lin P-H, Lin-Chao S, Cohen SN. Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci U S A. 2002;99: 9697–9702. doi:10.1073/pnas.112318199
20. Balleza E, Kim JM, Cluzel P. Systematic characterization of maturation time of fluorescent proteins in living cells. Nat Methods. 2018;15: 47–51. doi:10.1038/nmeth.4509
21. Hebisch E, Knebel J, Landsberg J, Frey E, Leisner M. High variation of fluorescence protein maturation times in closely related Escherichia coli strains. PLoS One. 2013;8: e75991. doi:10.1371/journal.pone.0075991
22. Maurizi MR. Proteases and protein degradation in Escherichia coli. Experientia. 1992;48: 178–201. doi:10.1007/BF01923511
23. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114–2120. doi:10.1093/bioinformatics/btu170
24. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, et al. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2012;29: 15–21. doi:10.1093/bioinformatics/bts635
25. Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019;47: e47. doi:10.1093/nar/gkz114
26. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15: 550. doi:10.1186/s13059-014-0550-8
27. Benjamini Y, Hochberg Y. Controlling the false discovery rate: A practical and powerful approach to multiple testing. J R Stat Soc. 1995;57: 289–300. doi:10.1111/j.2517-6161.1995.tb02031.x
28. Razo-Mejia M, Barnes SL, Belliveau NM, Chure G, Einav T, Lewis M, et al. Tuning Transcriptional Regulation through Signaling: A Predictive Theory of Allosteric Induction. Cell Syst. 2018;6: 456-469.e10. doi:10.1016/j.cels.2018.02.004
29. Lloyd-Price J, Gupta A, Ribeiro AS. SGNS2: a compartmentalized stochastic chemical kinetics simulator for dynamic cell populations. Bioinformatics. 2012;28: 3004–3005. doi:10.1093/bioinformatics/bts556

S2 Appendix: Supporting Figures

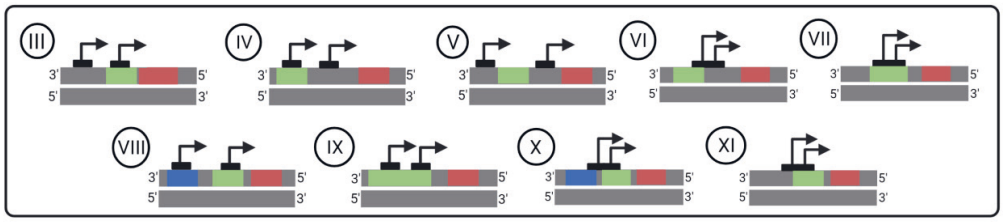

Fig A. Other arrangements of tandem promoters in *E. coli.* Unlike the arrangements I and II in Fig 1 in the main manuscript, the arrangements here (III-XI) allow for overlaps with or in between other gene(s). The red, green, and blue rectangles are DNA regions coding for RNA. These arrangements are not considered in this study. Figure created with BioRender.com.

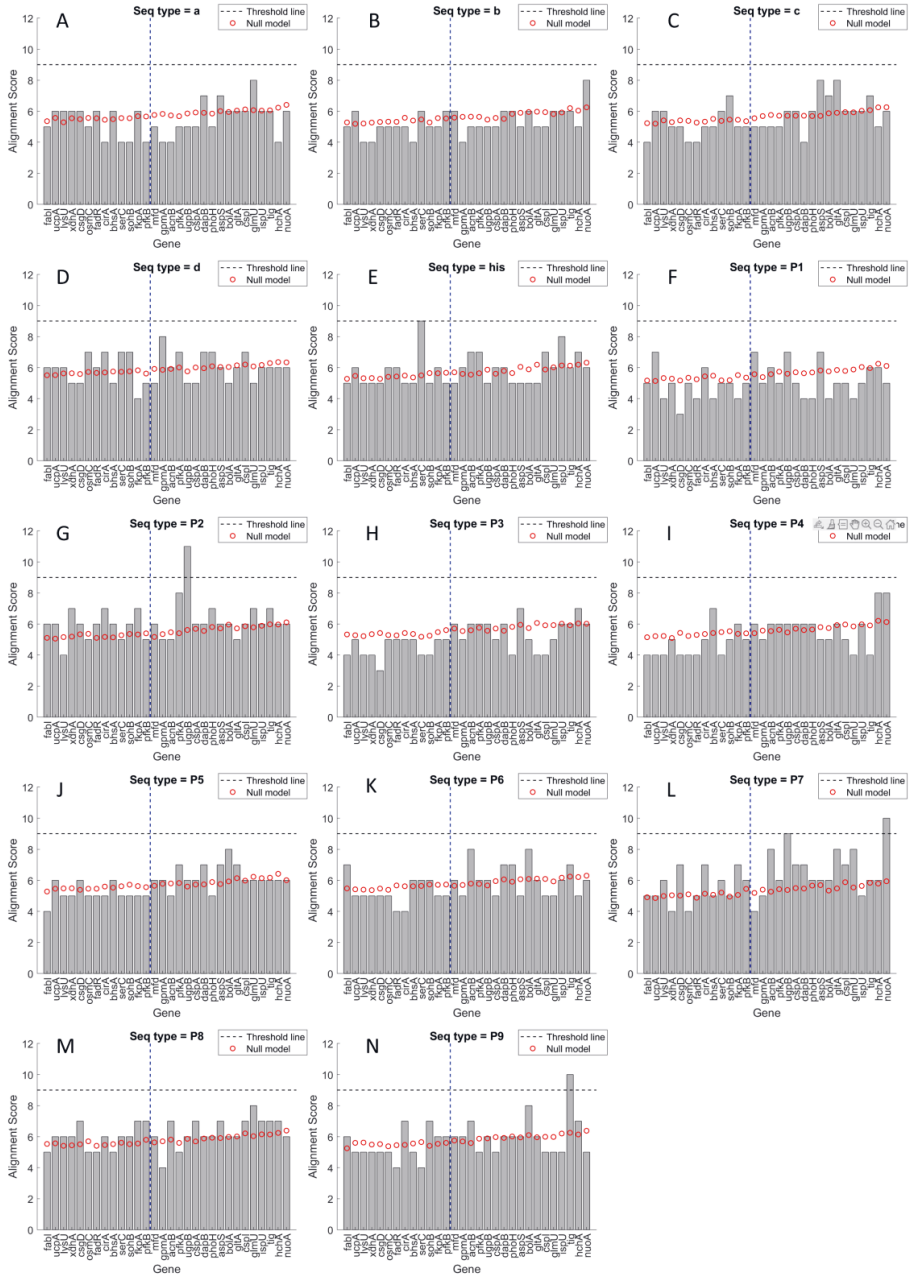

Fig B. Local alignment scores. Local alignment scores between known pause sequences and the sequences in between the tandem promoter regions (grey bars). Also shown by red circles are the alignment scores between each pause sequence and randomly generated sequences with the same d_{TSS} as the natural genes. The minimum alignment score to be considered significant is shown by a dashed black line. Finally, the blue vertical dashed line at d_{TSS} = 35 bp shows the separation between genes subject to occlusion or not.

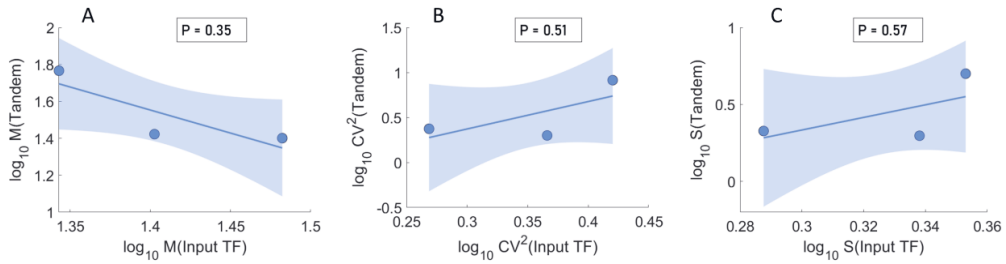

Fig C. Correlation of the moments of the single-cell protein numbers between genes and their input TFs. Scatter plots between the moments of the single-cell protein numbers (in $\log_{10}$ scale) of genes regulated by tandem promoters ('Tandem') and their input TFs. (A) Mean, (B) CV^2, and (C) Skewness. The blue line is the best linear fit, and its shadow is the standard error of the fit. The p-value, *P* is the probability that the slope of the line equals 0. If $P < 0.05$, there is a statistically significant correlation. The genes used in these results are listed in Table E in S3 Appendix. The axes differ widely in scales between the figures to facilitate visualization of the relationships.

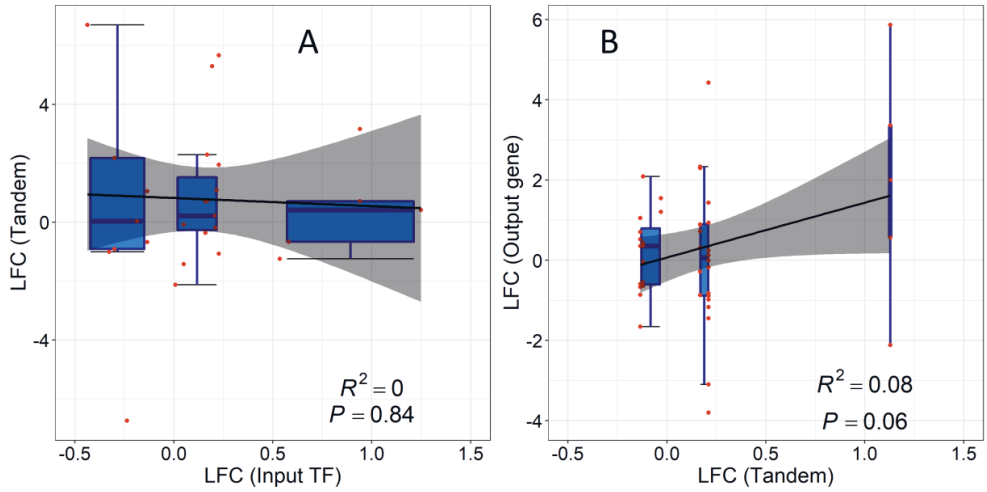

Fig D. Correlation of RNA fold changes of genes and their input TFs. Correlation plots between the LFCs of the RNA numbers of genes controlled by tandem promoters with their input and output genes. (A) LFCs (from 0 to 20 min) of 29 genes expressing input TFs plotted against the corresponding LFCs (from 0 to 180 min) of the genes controlled by tandem promoters. (B) LFCs (from 0 to 20 min) of genes controlled by tandem promoters plotted against the corresponding LFCs of their output genes (from 0 to 180 min). A total of 43 TF-gene interactions were analysed. RNA-seq measurements described in section "RNA-seq Measurements and Analysis in S1 Appendix". The black line is the best linear fit and the grey shadow area is the standard error of the fit. The blue horizontal lines inside the boxes are the median, the top of the boxes are the 3rd quartile (Q3) and the bottom of the boxes are the first quartile (Q1). The error bars at the top and bottom range from (Q3+1.5*IQR) to (Q1-1.5*IQR),

with an interquartile range: IQR = Q3 – Q1. The three box plots correspond to the data points with LFCs < 0, LFC between 0 and 0.5, and LFC > 0.5. Related to Table E and F in S3 Appendix.

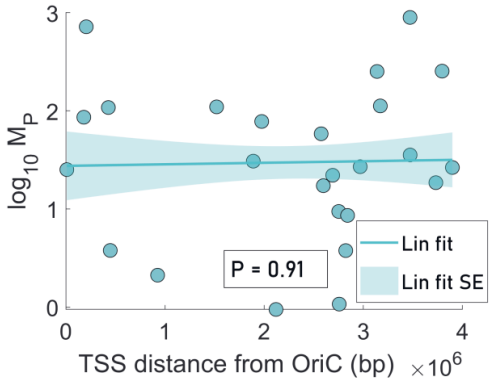

Fig E. Relationship between expression levels of the genes controlled by tandem promoters and the distance in nucleotides (bp) from the upstream promoter and the Oric region in the DNA. Data from 25 genes for the 1X condition. Also shown in a linear fit and the corresponding 1 standard error of the fit (shadow area). The p-value, *P*, is the probability that the slope of the line equals 0.

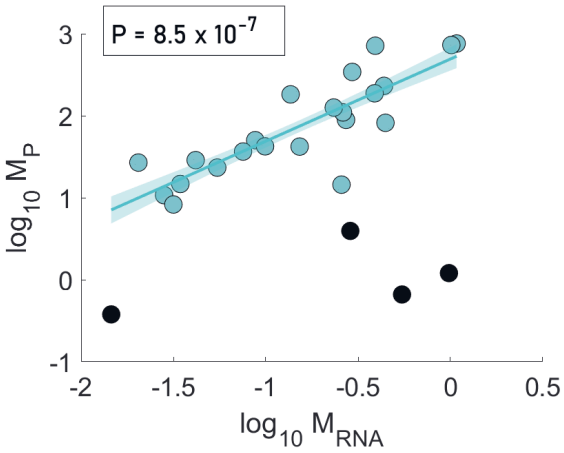

Fig F. Correlation plot between the mean single-cell RNA levels (M_{RNA}) and the mean single-cell protein numbers (M_P). Both data are obtained from Ref. [28] in main manuscript and are processed to include only genes controlled by tandem promoters (classes I and II, Table H in S3 Appendix). The line is the best linear fit to the data, and its shadow area is the standard error of the fit. The p-value, *P* is the probability that the slope of the line equals 0. Since P < 0.05, we conclude that M_{RNA} and M_P are significantly correlated. The black balls correspond to 4 genes that were not considered when fitting the line, due to being outliners. In our own data, cells carrying these same 4 genes exhibited a fluorescence that was equal or lower than the cellular background fluorescence in either 1X or 0.5X media.

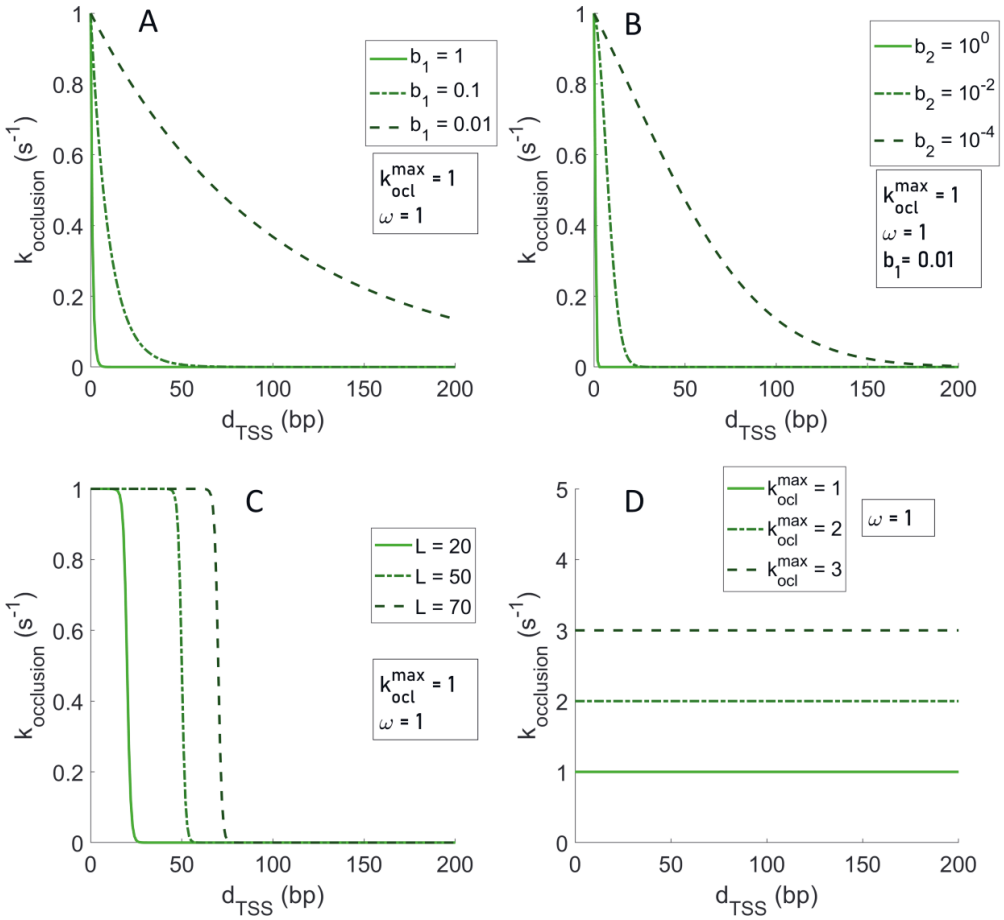

Fig G. Models of transcription interference. Models of transcription interference between RNAPs in tandem promoters as a function of the d_{TSS} between them. (A) 'Exponential 1' as a function for different values of 'b_1'. (B) 'Exponential 2' as a function at different values of 'b_2. (C) Continuous 'step-like' function for different values of 'L' (which is the d_{TSS} at which the step occurs). (D) Zero order polynomial for different values of $k_{ocl}^{\max}$. See Table 1 in the main manuscript for the definitions of these models and variables within.

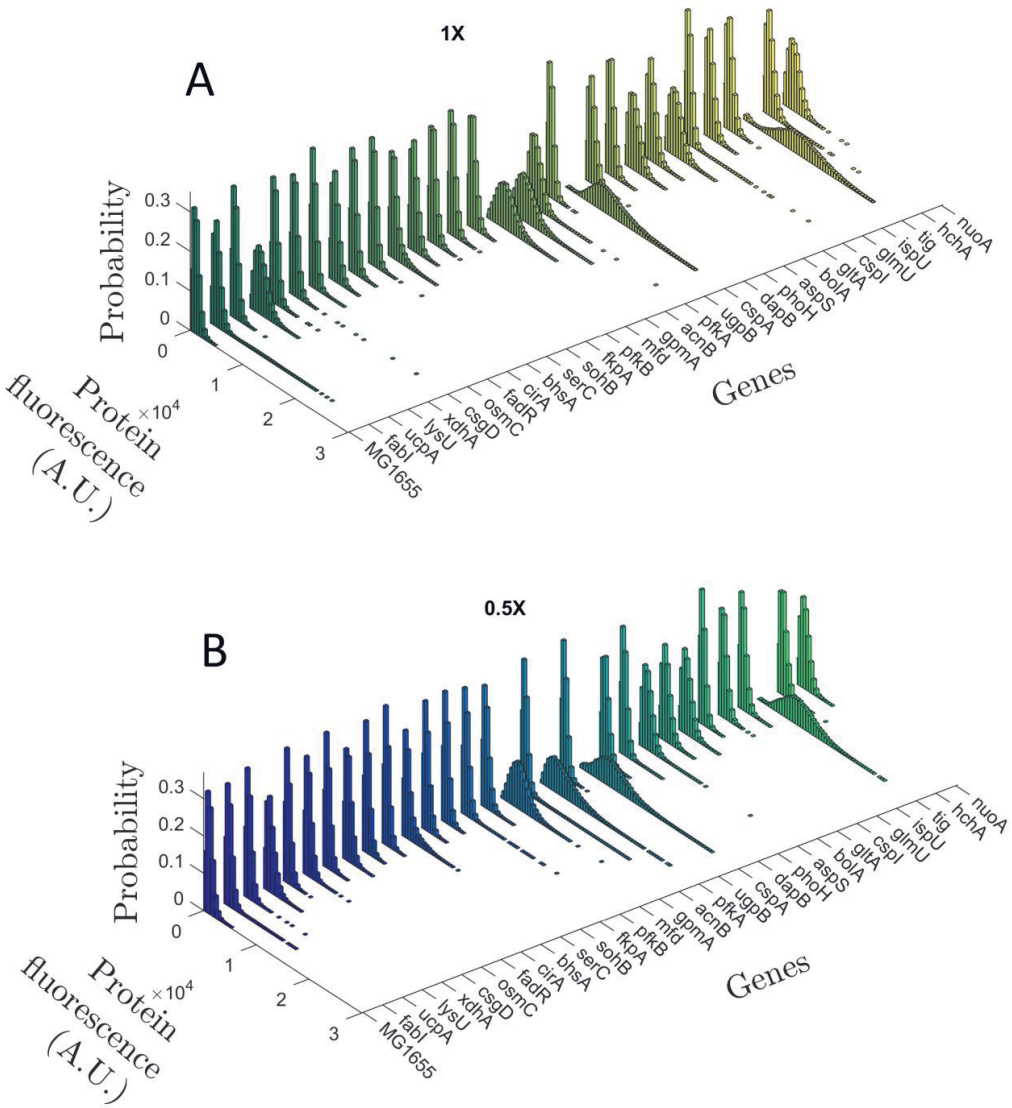

Fig H. Protein florescence distributions. Protein florescence distributions of genes controlled by tandem promoters measured by flow-cytometry**.** Each protein is tagged with a YFP (YFP strain library). Only 1 of 3 biological replicates is shown per gene. (A) M9 medium (1X). (B) Diluted M9 medium (0.5X). 'MG1655' are control cells, not carrying YFP. Protein fluorescence is shown in arbitrary units.

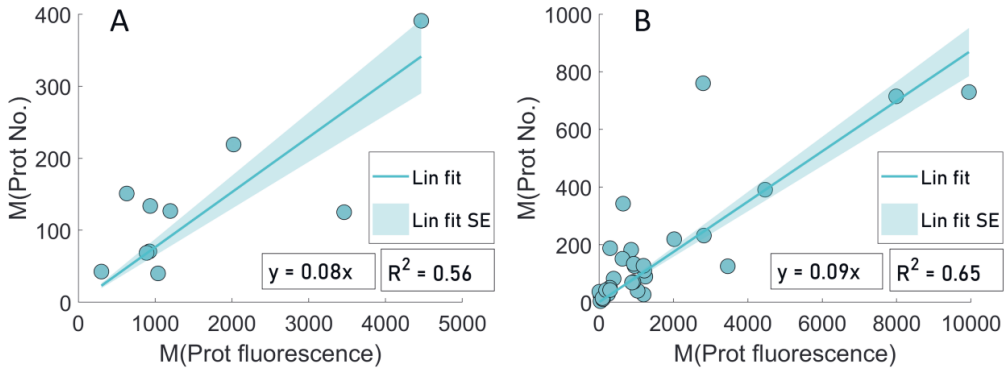

Fig I. Estimation of scaling factors using data from genes controlled by single promoters. A) Mean single-cell protein fluorescence (own measurements of genes controlled by single promoters) plotted against the corresponding mean single-cell protein numbers reported in [28]. From the equation of the best fitting line without y-intercept (y-intercept = 0), we obtained a scaling factor, s_f, equal to 0.08. B) Same as (A) but the own measurements are of both single promoters and tandem promoters, merged. From the equation of the best fitting line without y-intercept (y-intercept = 0), we obtained a scaling factor, s_f, equal to 0.09.

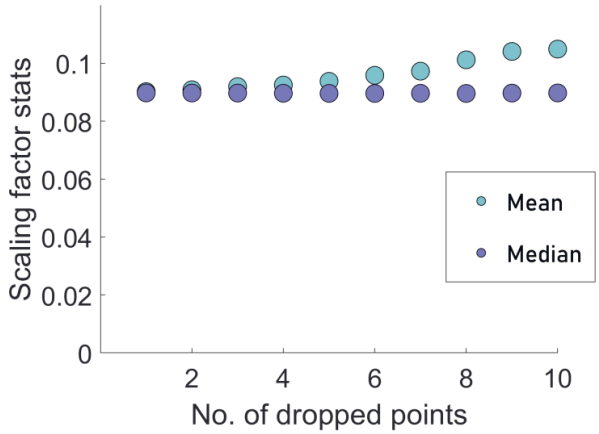

Fig J. Sensitivity test. Mean and median of scaling factor varies as a function of number of data points randomly dropped.

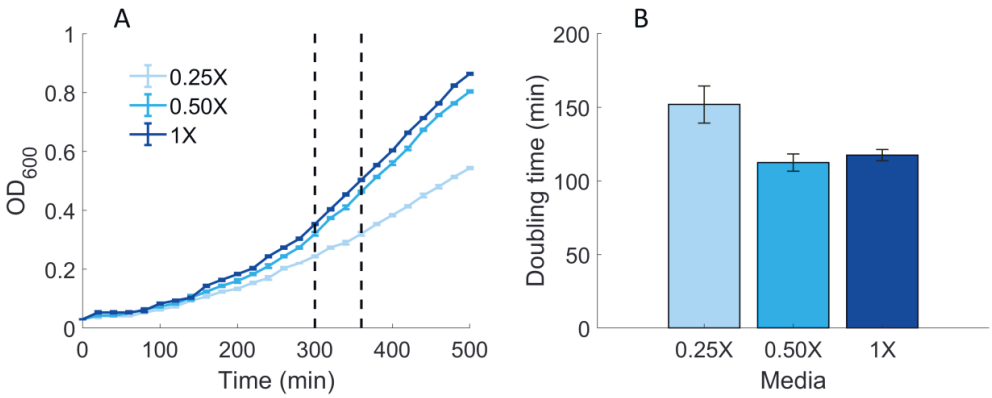

Fig K. Growth curves and doubling times. A. Optical density (OD_{600}) curves of *E. coli* MG1655 cells grown in 0.25X,0.5X and 1X media (section 'Media and Chemicals' in S1 Appendix). B. From these curves, the doubling time was estimated to be ~112 min in 0.5X and ~118 min in 1X. We used 115 min doubling time in the models. The estimation is made using the formula $D = \frac{\ln(2)}{\ln\left(\frac{OD(t_2)}{OD(t_1)}\right)} \times (t_2 - t_1)$, with t_2 and t_1 being the end and start times (in minutes), respectively. They are marked by two vertical dashed black lines. The error bars denote the standard error of the mean. Ref. [28] in main manuscript reported ~150 min using 96 well-plates in the same conditions. The fact that we used culture tubes may explain the difference.

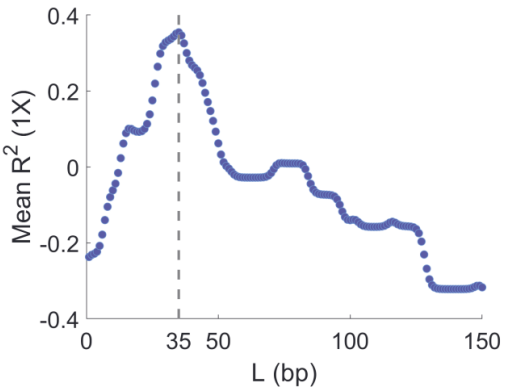

Fig L. Mean R^2 of the step interference model. Mean R^2 of the step interference model to the 1X data in Fig 6A, 6B, and 6C, as a function of L (d_{TSS} at which the step of the step function occurs). The Mean

R^2 is visibly maximized at L = 35, which marked by a grey dashed line. Relates to Fig 6 in the main manuscript.

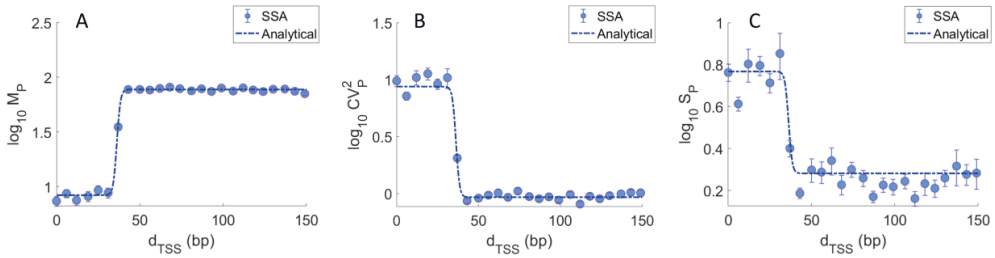

Fig M. Confronting the solutions of the analytical and stochastic model. (A) $\log_{10}$ of mean protein numbers, (B) $\log_{10}$ of CV^2 of protein numbers and (C) $\log_{10}$ of Skewness of protein numbers as a function of d_{TSS}. The blue line is the analytical solution of the step model. The blue dots are the mean results of stochastic simulations of the step model. The parameters used are shown in Tables 2 and 3 in the main manuscript. See Section 'Stochastic simulations for the step interference model' in S1 Appendix.

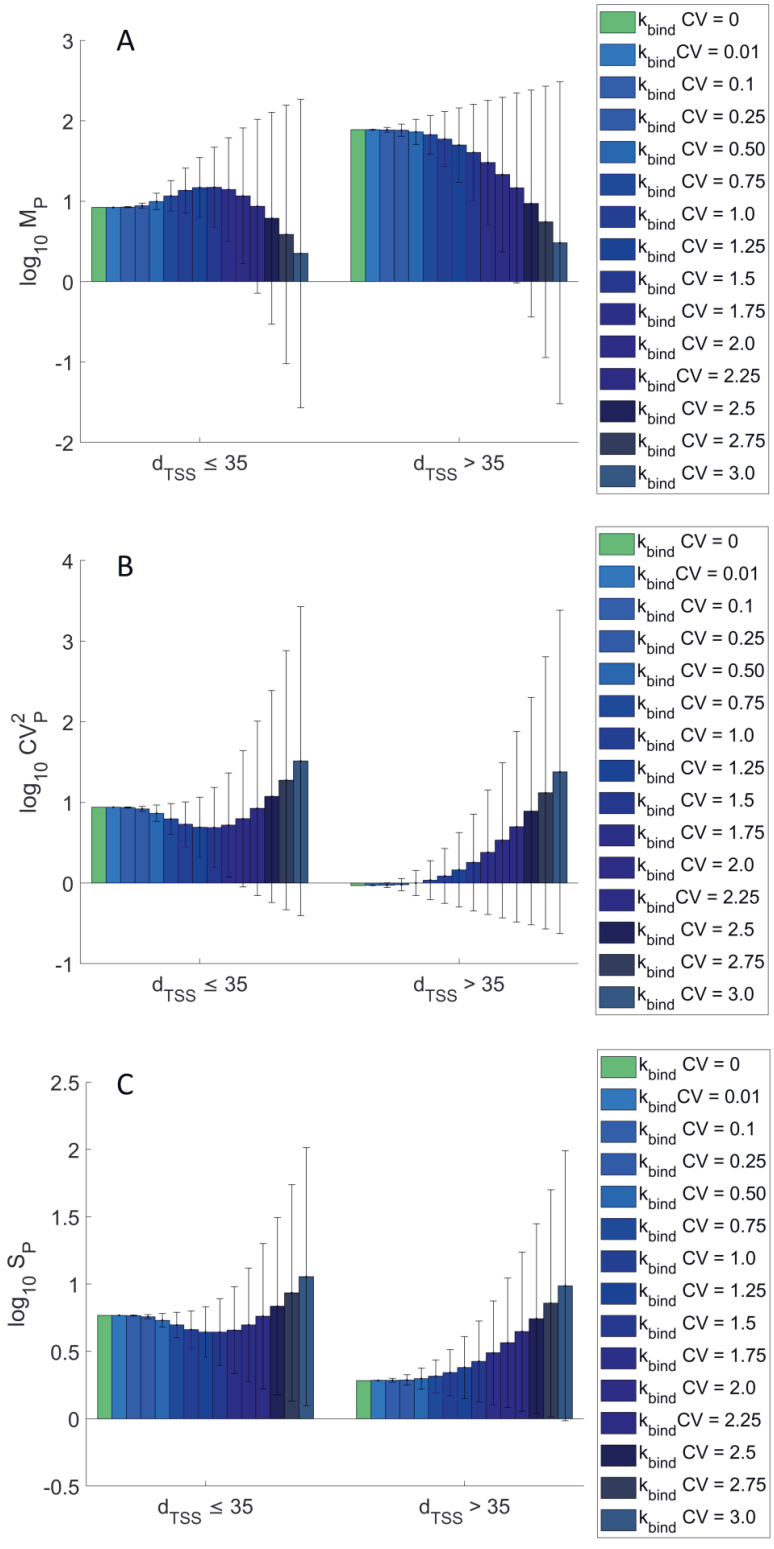
A
B
C
log10 MP
log10 CV2P
log10 SP
dTSS ≤ 35
dTSS > 35
kbind CV = 0
kbind CV = 0.01
kbind CV = 0.1
kbind CV = 0.25
kbind CV = 0.50
kbind CV = 0.75
kbind CV = 1.0
kbind CV = 1.25
kbind CV = 1.5
kbind CV = 1.75
kbind CV = 2.0
kbind CV = 2.25
kbind CV = 2.5
kbind CV = 2.75
kbind CV = 3.0

Fig N. Solutions of the analytical model for different levels of variability of $k_{bind} \cdot [R]$. (Top) Mean, (Middle) CV^2 and (Bottom) S of single-cell protein numbers produced by tandem promoters when $d_{TSS} \leq 35$ (left) and $d_{TSS} > 35$ (right). The green bar is the analytical solution with $CV\left(k_{bind} \cdot [R]\right) = 0$. The other bars are from analytical solutions for various degrees of variability of $k_{bind} \cdot [R]$ of each promoter.

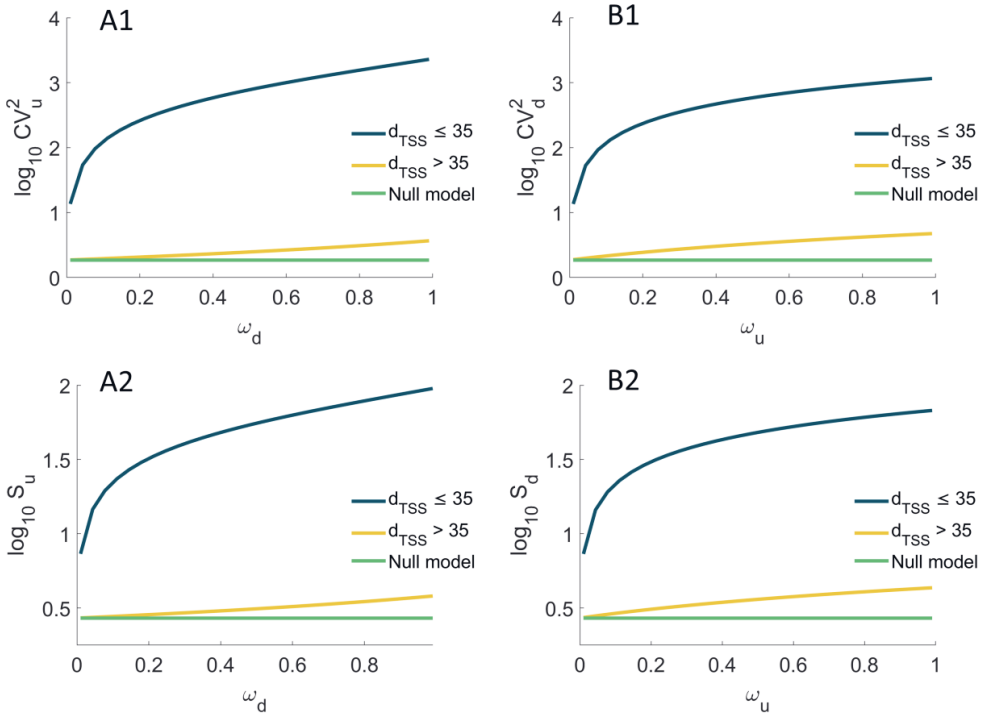

Fig O. Variability and skewness in single-cell protein numbers produced from an upstream and from a downstream promoter as a function of promoter occupancy of the other promoter. CV_P^2 and S_P of the single-cell distribution of the number of proteins produced (**A1 and A2**) by the upstream promoter alone, and (**B1 and B2**) 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 $k_{occlusion}$, $k_{sitting}$, and ω to zero.

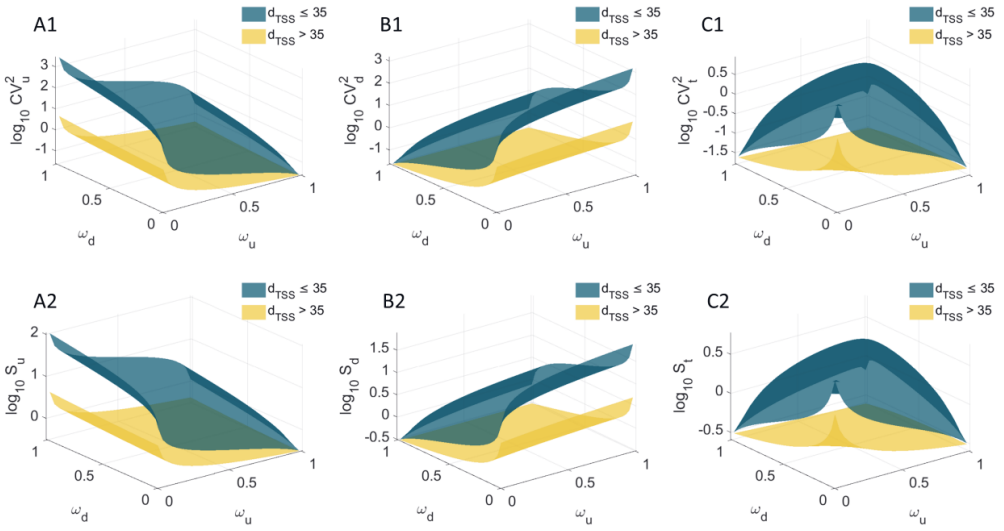

Fig P. Variability and skewness in single-cell protein numbers as a function of promoter occupancy. Expected variability (CV^2) and skewness (S) of the single cell distribution of protein numbers due to the activity of, respectively: (**A1** and **A2**) the upstream promoter alone, (**B1** and **B2**) the downstream promoter alone, and (**C1** and **C2**) both promoters. Shown is CV^2, S 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.

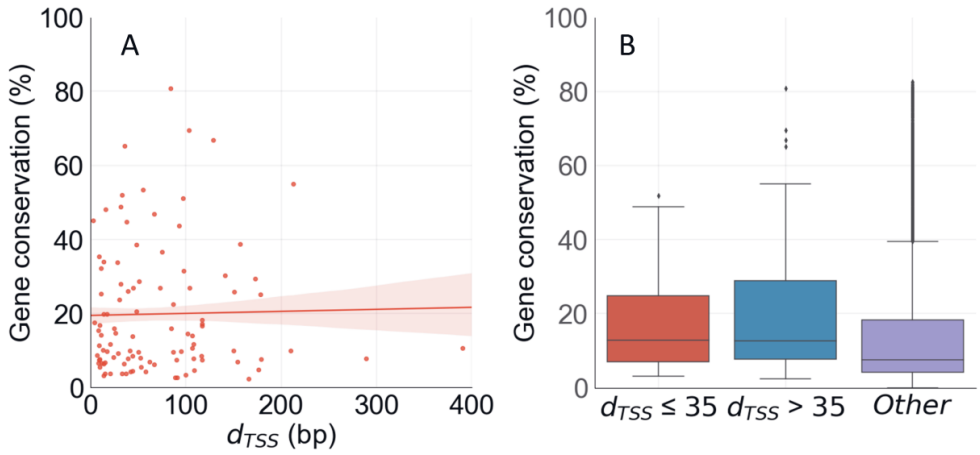

Fig Q. Gene conservation levels. (A) Correlation between d_{TSS} (bp) of the pairs of tandem promoters and the evolutionary conservation level of the gene that they express. The line shown is the best linear fit to the data, and its shadow is the standard error of the fit. (B) Box plot of the gene conservation levels of the cohorts of genes with $d_{TSS} > 35$ and with $d_{TSS} \leq 35$, along with genes other than those in tandem formation. The horizontal black line inside each box marks the median, the top of the box shows the 3rd quartile (Q3), and the bottom of the box shows the first quartile (Q1) of each gene cohort. The error bar above the box marks the range of values within (Q3+1.5*IQR), while the error bar below the bottom shows the range of values within (Q1-1.5*IQR). Here, IQR = Q3 – Q1.

S3 Appendix: Supporting Tables

Table A. List of genes controlled by tandem promoters.

S. No	Configuration (see Fig 1 main manuscript)	Gene	Promoters (upstream/ downstream)	Distance between TSS's (bp)
1	I	aspS	aspSp1/aspSp	84
2	I	bolA	bolAp2/bolAp1	85
3	I	cspI	cspIp/cspIp2	100
4	I	glmU	glmUp2/glmUp1	103
5	I	gltA	gltAp1/gltAp2	97
6	I	hchA	hchAp2/hchAp	150
7	I	ispU	ispUp1/ispUp2	117
8	I	tig	tigp1/tigp3	129
9	I	nuoA	nuoAp1/nuoAp2	173
10	II	acnB	acnBp/acnBp2	45
11	II	bhsA	bhsAp9/bhsAp	14
12	II	cirA	cirAp2/cirAp1	13
13	II	csgD	csgDp1/csgDp2	9
14	II	cspA	cspAp1/cspAp2	51
15	II	dapB	dapBp2/dapBp1	55
16	II	fabI	fabIp/fabIp1	3
17	II	fadR	fadRp/fadRp2	11
18	II	fkpA	fkpAp1/fkpAp2	26
19	II	gpmA	gpmAp2/gpmAp	38
20	II	lysU	lysUp1/lysUp2	8
21	II	mfd	mfdp1/mfdp2	36
22	II	osmC	osmCp1/osmCp2	10
23	II	pfkA	pfkAp2/pfkAp1	48
24	II	pfkB	pfkBp2/pfkBp1	28
25	II	phoH	phoHp1/phoHp2	73
26	II	serC	serCp2/serCp	16
27	II	sohB	sohBp1/sohBp2	17
28	II	ucpA	ucpAp2/ucpAp1	7
29	II	ugpB	ugpBp2/ugpBp1	48
30	II	xdhA	xdhAp/xdhAp2	8

List of genes controlled by tandem promoters whose single-cell protein numbers were measured by flow-cytometry using cells of the YFP strain library. Also shown are their promoters in tandem formation, their configuration, and the distance in base pairs (bp) between their TSSs.

Table B. List of strains of the YFP strain library observed by flow-cytometry.

S. No.	Strain name	Genotype	Source
1	acnB [SX1900]	F-, acnB791-YFP(::cat), Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13455)
2	argP [SX1436]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], argP794-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12991)
3	aspS [SX1044]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], aspS793-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12599)
4	bhsA [SX1979]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], bhsA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13534)
5	bolA [SX1087]	F-, Δ(argF-lac)169, bolA791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12642)
6	cirA [SX1509]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], cirA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13064)
7	csgD [SX1465]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], csgD791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13020)
8	cspA [SX1097]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, cspA791-YFP(::cat), rph-1	Yale CGSC (CGSC # 12652)
9	cspI [SX1106]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], cspI797-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12661)
10	dapB [SX1910]	F-, dapB792-YFP(::cat), Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13465)
11	fabD [SX2002]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], fabD793-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13557)
12	fabH [SX1474]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], fabH795-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13029)
13	fabI [SX1038]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], fabI796-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12593)
14	fadR [SX1521]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], fadR795-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13076)
15	fkpA [SX2015]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, fkpA791-YFP(::cat), rph-1	Yale CGSC (CGSC # 13570)
16	fur [SX1916]	F-, Δ(argF-lac)169, fur-791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13471)

17	glmU [SX1004]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, glmU792-YFP(::cat)	Yale CGSC (CGSC # 12559)
18	gltA [SX1925]	F-, Δ(argF-lac)169, gltA791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13480)
19	gpmA [SX1553]	F-, Δ(argF-lac)169, gpmA791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13108)
20	hchA [SX1988]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], hchA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13243)
21	ispU [SX1052]	F-, ispU796-YFP(::cat), Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12607)
22	lysU [SX1127]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, lysU793-YFP(::cat)	Yale CGSC (CGSC # 12682)
23	mfd [SX1072]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], mfd-791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12627)
24	mreB [SX1466]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], mreB791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13021)
25	nagC [SX1561]	F-, Δ(argF-lac)169, nagC791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13116)
26	nlpA [SX1615]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, nlpA791-YFP(::cat)	Yale CGSC (CGSC # 13170)
27	nuoA [SX1772]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], nuoA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13327)
28	osmC [SX1758]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], osmC791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13313)
29	pepD [SX1530]	F-, pepD792-YFP(::cat), Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC #13085)
30	pfkA [SX1349]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, pfkA791-YFP(::cat)	Yale CGSC (CGSC # 12904)
31	pfkB [SX1761]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], pfkB792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13316)
32	phoH [SX1752]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], phoH791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13307)
33	serC [SX1390]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], serC791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12945)
34	sohB [SX1707]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], sohB791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13262)
35	tig [SX1140]	F-, Δ(argF-lac)169, tig-791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12695)

36	ucpA [SX1211]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], ucpA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12766)
37	ugpB [SX1574]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, ugpB791-YFP(::cat), rph-1	Yale CGSC (CGSC # 13129)
38	wrbA [SX1718]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], wrbA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13273)
39	xdhA [SX1671]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], xdhA792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13226)
40	yccJ [SX1975]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yccJ791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13530)
41	yccT [SX1368]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yccT792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12923)
42	aldA [SX1901]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], aldA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13456)
43	elaB [SX1695]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], elaB792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13250)
44	feoA [SX1781]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, feoA791-YFP(::cat), rph-1	Yale CGSC (CGSC # 13336)
45	gcvT [SX1674]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gcvT792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13229)
46	glpD [SX1550]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, glpD792-YFP(::cat), rph-1	Yale CGSC (CGSC # 13105)
47	pepN [SX1519]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], pepN794-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13074)
48	wrbA [SX1718]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], wrbA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13273)
49	ybeL [SX1822]	F-, Δ(argF-lac)169, ybeL794-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13377)
50	ydfG [SX1986]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], ydfG791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13541)
51	yjbQ [SX1859]	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, yjbQ792-YFP(::cat)	Yale CGSC (CGSC # 13414)

Table C. Average 'network' properties of genes with 1 or more TFs.

Network properties	**Genes controlled by tandem promoters with $d_{TSS} \leq 35$**	**Genes controlled by tandem promoters with $d_{TSS} > 35$**	**All promoters of genes with 1 or more TF interactions**

	Mean ± SEM	Random set from all genes Mean ±SEM (p-value)	Mean ± SEM	Random set from all genes Mean ±SEM (p-value)	Mean ± SEM
Average Shortest PathLength	0.31 ± 0.16	0.17 ± 0.11 (0.23)	0.13 ± 0.05	0.17±0.08 (0.60)	0.17 ±0.01
Clustering Coefficient	0.09 ± 0.03	0.11 ± 0.03 (0.68)	0.10 ± 0.03	0.11 ± 0.03 (0.62)	$0.11 \pm 4.34\times10^{-3}$
Eccentricity	0.56±0.31	0.25 ± 0.20 (0.22)	0.15 ± 0.06	0.26 ± 0.16 (0.73)	0.26 ± 0.03
Edge Count	5±1.64	4.64 ± 3.4 (0. 33)	3.3 ± 0.83	4.64 ± 2.73 (0.67)	4.63 ± 0.43
Indegree	2.33±0.48	2.32 ± 0.34 (0.52)	2.02 ± 0.17	2.31 ± 0.27(0.83)	2.32 ± 0.04
Neighborhood Connectivity	161.76 ± 29.09	131.95 ± 21.74 (0.20)	134.63 ± 15.1	131.87 ± 17.36 (0. 44)	131.91 ± 2.74
Outdegree	2.66 ± 1.34	2.33 ± 3.4 (0.30)	1.28 ± 0.83	2.31 ± 2.71 (0. 59)	2.32 ± 0.43

Shown are the network properties for genes controlled by tandem promoters at a distance $d_{TSS} \leq 35$ bp and at a distance $d_{TSS} > 35$ bp. For comparison, we show the same properties, when averaged from all genes of *E. coli*'s TF network. Genes without TF's are not considered. Note that all p-values are larger than 0.05.

Table D: Genes controlled by tandem promoters without input TFs.

S. No.	Gene	Availability in the YFP strain library
1	ampH	
2	ansP	
3	aroK	
4	aspS	✓
5	bepA	
6	cfa	
7	cobU	
8	crfC	
9	degQ	
10	fkpA	✓
11	ispU	✓
12	lpp	
13	mepS	
14	mfd	✓
15	narU	

16	opgG	
17	panD	
18	pfkB	✓
19	serW	
20	tig	✓
21	ucpA	✓
22	xapR	
23	ybgI	
24	ygiM	
25	yheO	
26	yobF	

Genes controlled by tandem promoters without input TFs. Those genes whose proteins are tagged with YFP in the YFP strain library are marked with the symbol '✓'.

Table E. Genes controlled by tandem promoters regulated by one and only one input TF.

	Tandem promoter's genes	**Availability in YFP strain library**	**Input TF**	**Availability in YFP strain library**
1	argR		argR	
2	cvpA		purR	
3	cysK		cysB	✓
4	dapB	✓	argP	✓
5	fabI	✓	fadR	✓
6	fadR	✓	fadR	✓
7	fliL		flhdC	
8	ftnB		cpxR	✓
9	glgS		crp	
10	glk		cra	
11	glmU	✓	nagC	✓
12	gpmA	✓	fur	✓
13	hchA	✓	h-ns	
14	ibaG		mlrA	✓
15	iraP		csgD	✓
16	leuL		leuO	
17	livK		lrp	
18	lysU	✓	lrp	
19	mqsR		mqsA	
20	ompA		crp	
21	ompX		fnr	
22	osmB		rcsB	✓
23	pfkA	✓	cra	

24	phoH	✓	phoB	
25	potF		ntrC	
26	slyB		phoP	
27	sohB	✓	crp	
28	wza		rcsaB	
29	xdhA	✓	fnr	
30	ydbK	✓	soxS	
31	yeaG		ntrc	
32	yhbT		csgD	✓
33	yqjA		cpxR	✓

When the proteins of these genes and of their input TFs can be measured using strains of the YFP strain library, they are flagged with the symbol '✓'.

Table F. Genes controlled by, and only by, a TF expressed by tandem promoters.

	Genes controlled by tandem promoters	**Availability in YFP strain library**	**Genes regulated by the protein expressed by the gene controlled by tandem promoters**	**Availability in YFP strain library**
1	argR		argA	✓
2	argR		argB	
3	argR		argC	
4	argR		argE	✓
5	argR		argF	
6	argR		argH	
7	argR		argI	
8	argR		argR	
9	argR		artI	
10	argR		artJ	
11	argR		artM	
12	argR		artP	✓
13	argR		artQ	
14	argR		lysO	
15	bolA	✓	ampC	
16	bolA	✓	dacC	
17	bolA	✓	mreB	✓
18	bolA	✓	mreC	
19	bolA	✓	mreD	
20	csgD	✓	dgcC	
21	csgD	✓	iraP	
22	csgD	✓	nlpA	✓
23	csgD	✓	pepD	✓
24	csgD	✓	wrbA	✓
25	csgD	✓	yccJ	✓
26	csgD	✓	yccT	✓

27	csgD	✓	yhbS	
28	csgD	✓	yhbT	
29	evgA		frc	
30	evgA		oxc	✓
31	evgA		yegR	✓
32	evgA		yegZ	
33	evgA		yfdE	
34	evgA		yfdV	
35	evgA		yfdX	
36	fadR	✓	accA	
37	fadR	✓	accD	
38	fadR	✓	fabD	✓
39	fadR	✓	fabG	
40	fadR	✓	fabH	✓
41	fadR	✓	fabI	✓
42	fadR	✓	fadM	
43	fadR	✓	fadR	✓
44	xapR		xapA	
45	xapR		xapB	

Table G. Protein levels and d_{TSS} of 10 genes as measured by Microscopy and Image Analysis.

Gene	TSS distance (d_{TSS})	Mean single-cell protein level (Microscopy)
xdhA	8	0.04
csgD	9	0.64
serC	16	0.24
sohB	17	0.37
pfkA	48	2.8
dapB	55	0.57
aspS	84	1.72
gltA	97	3.02
hchA	150	0.74
nuoA	173	2.04

Related to Fig 4C in the main manuscript.

Table H. Number of genes controlled by a pair of tandem promoters in each configuration.

Configuration	Number (in RegulonDB)	Present in the YFP strain library (measured here by flow-cytometry)
I	40	9(9)
II	62	21(21)
III	7	3

IV	4	1
V	6	2
VI	0	0
VII	3	1
VIII	2	2
IX	4	1
X	0	0
XI	9	2
Other	6	0

Related to Fig 1 in the main manuscript and Fig A in S2 Appendix.

Table I. Coefficient of variation, CV, of the gamma distribution.

CV ($k_{bind} \cdot [R]$)	$Mean\left(abs\left(\begin{matrix} k^u_{bind} \cdot [R] - \\ k^d_{bind} \cdot [R] \end{matrix} \right) \right)$	$Mean\left(\frac{abs\left(\begin{matrix} k^u_{bind} \cdot [R] - \\ k^d_{bind} \cdot [R] \end{matrix} \right)}{k^u_{bind} \cdot [R]} \right) \times 100\%$
0.01	7.52×10^{1}	1.14 %
0.1	7.64×10^{-4}	1.16×10^{1} %
0.25	1.86×10^{-3}	2.98×10^{1} %
0.5	3.63×10^{-3}	7.33×10^{1} %
0.75	5.27×10^{-3}	1.99×10^{2} %
1	6.62×10^{-3}	2.05×10^{3} %
1.25	7.81×10^{-3}	5.15×10^{4} %
1.5	8.66×10^{-3}	1.95×10^{7} %
1.75	9.41×10^{-3}	6.19×10^{12} %
2.0	9.89×10^{-3}	1.48×10^{15} %
2.25	1.04×10^{-2}	1.77×10^{17} %
2.5	1.10×10^{-2}	6.60×10^{18} %
2.75	1.12×10^{-2}	4.00×10^{24} %
3.0	1.20×10^{-2}	6.03×10^{30} %

Coefficient of variation, CV, of the gamma distribution from which $k_{bind} \cdot [R]$ of each promoter in tandem configuration is sampled from. Also shown is the resulting expected mean absolute difference in $k_{bind} \cdot [R]$ between the upstream and downstream promoters. Furthermore, the last column shows how much larger (in percentage) is one of the $k_{bind} \cdot [R]$ values compared to the other.

Table J. Location of the tandem promoters relative to the oriC.

Genes controlled by tandem promoters	Distance between the upstream TSS and the oriC
aspS	1975043
bolA	3471395
cspI	2286932
glmU	10418
gltA	3170977
hchA	1890114
ispU	3730960
nuoA	1520409
tig	3470751
acnB	3794225
bhsA	2756725
cirA	1678802
csgD	2822400
cspA	205855
dapB	3897456
fabI	2574623
fadR	2690839
fkpA	448219
gpmA	3138074
lysU	428830
mfd	2751716
osmC	2369148
pfkA	181499
pfkB	2119421
phoH	2840879
serC	2968165
sohB	2596460
ucpA	1381073
ugpB	333318
xdhA	925487

S4 Appendix: Supplementary Results

Pause sequences

We investigated if the nucleotide sequence of and in between the natural tandem promoters is coding for specific sequences known to perturb RNAP elongation. There are several events that compete with stepwise elongation. However, arrest, misincorporation and editing, pyrophosphorolysis, and premature termination are too rare in optimal growth conditions (rate constants listed in [1]) to be influential in several genes, and/or are not sequence dependent. Only sequences known to enhance transcriptional pausing [2] could fit both of these requirements. In *E. coli*, the mean rate of non-sequence specific pauses is 1 per 100 base pairs. These last 3 s on average [3-4]. However, a few sequences can enhance pausing frequency and/or duration (up to 15 or more seconds) [5] via various mechanically processes, which explains their variability in half-life and frequency of occurrence. For example, '*his*' pauses occur when the assembling RNA forms a hairpin-like loop, while '*ops*' pauses do not require it. Likely because of it, *his* pauses have longer half-life [6]. We searched in (and in between) the sequences of the 102 pairs of tandem promoters for the 14 sequences (each 12 nucleotides long) known to enhance pausing [7] (section 'Sequences prone to causes transcriptional pauses' in S1 Appendix) but found none. Thus, sequence-dependent transcriptional pausing should not be a common phenomenon in the tandem promoters of arrangements I and II. Even when allowing for 3 or less mistakes (sequence gaps, misalignments, duplicates, etc.), we only found 5 matches in the 30 of the 102 tandem promoter pairs studied with protein measurements below (Fig B in the S2 Appendix, note the 5 bars crossing the threshold).

Over-representation test

We performed an over-representation test to search for biological functions (as defined in [8,9] that are overrepresented by genes controlled by tandem promoters (using PANTHER 14 [10]). While based on a Fisher test, some biological processes appear to be overrepresented in our genes of interest (e.g., regulation of catabolic processes), none of them were significant to 'FDR correction' (FDR < 0.05, [10]. As such, we failed to identify a biological process significantly associated to genes controlled by tandem promoters (S1 Table).

Input-output transcription factor relationships

From time-lapse RNA-seq data, we assessed if the 102 genes controlled by tandem promoters (arrangements I and II, Fig 1) are affected by their input TFs. To facilitate this, we considered only those that have one and only input TF. I.e., we did not consider the 26 genes that do not have known input TFs (Table D in S3 Appendix), neither the 43 genes that have more than one input TF, making the detection of input-output relationships problematic. As such, of the 102, we considered only 33 genes (Table E in S3 Appendix). In these, we did not observe influences from input TFs (Fig C, Panel A in Fig

D in the S2 Appendix). Finally, and similarly, we observed genes whose only input TF is expressed by tandem promoters (Table F in S3 Appendix). Again, we found no correlation (Panel B in Fig D in the S2 Appendix). Note that, while we did not find influences from TF interactions in the conditions of our measurements, we expect these interactions to become active in other conditions (e.g., stress conditions).

Proteins with membrane-related positionings

From RegulonDB [11], of the 30 genes measured by flow-cytometry (Table A in S3 Appendix), only 3 are known to be related to membrane transportation and binding: bhsA, which is an outer membrane protein that is involved in copper permeability, stress resistance and biofilm formation, cirA, which is also an outer membrane transporter, and ugpB which is a periplasmic binding protein. Such membrane localizations could affect their quantification by YFP fusion, potentially by enhancing effects from avidity due to weakened diffusion.

However, none of these proteins significantly affect our results since, first, cirA and ugpB were removed from our analysis of the 1X condition, after preprocessing (gating, background subtraction and protein number conversion) (marked in red in S2 Table). Meanwhile, all three genes were removed from our analysis of the 0.5X condition after preprocessing (marked in red in S2 Table). Specifically, their removal was due to lack of expression above background autofluorescence.

Relationship with the OriC region

From EcoCyc [12], the OriC region has a length of 232 base pairs and is located in positions 3 925 744 and 3 925 975 in the DNA of *E. coli*. We calculated the shortest distance between the TSS of the upstream promoter and the Oric region. These positions in the DNA are shown in Table J in the S3 Appendix. Meanwhile, the corresponding protein expression levels of these genes in the 1X condition are shown in the S2 Table. Finally, we show a Fig E in the S2 Appendix of these distances from OriC plotted again $\log_{10} M_p$ which shows that the two quantities do not correlate statistically.

Regulation by H-NS

From RegulonDB [11], we investigated how many of the 102 genes controlled by tandem promoters (arrangements I and II) and how many of 30 of them observed by flow-cytometry are expected to be regulated by H-NS.

Of the 102 genes, 14 are regulated by H-NS (14%). Meanwhile, of the 30 genes, 5 are regulated by H-NS (17%). From this, we conclude that H-NS is not consistently a master regulator of these genes.

Nevertheless, of 4698 genes in E. coli, only 4 % are regulated by H-NS. This is significantly lower than in the case of the genes controlled by tandem promoters (p-value < 0.05 based on a Fisher test). As

such, one could argue that H-NS regulation does occur higher than expected by chance. Future studies of the dynamics of those genes during environmental changes may thus be of interest.

References

1. Rajala T, Häkkinen A, Healy S, Yli-Harja O, Ribeiro AS. Effects of transcriptional pausing on gene expression dynamics. PLoS Comput Biol. 2010;6: e1000704. doi: 10.1371/journal.pcbi.1000704
2. Herbert KM, La Porta A, Wong BJ, Mooney RA, Neuman KC, Landick R, et al. Sequence-resolved detection of pausing by single RNA polymerase molecules. Cell. 2006;125: 1083–1094. doi: 10.1016/j.cell.2006.04.032
3. Greive SJ, von Hippel PH. Thinking quantitatively about transcriptional regulation. Nat Rev Mol Cell Biol. 2005;6: 221–232. doi:10.1038/nrm1588
4. Neuman KC, Abbondanzieri EA, Landick R, Gelles J, Block SM. Ubiquitous Transcriptional Pausing Is Independent of RNA Polymerase Backtracking. Cell. 2003;115: 437–447. doi:10.1016/S0092-8674(03)00845-6
5. Herbert KM, Greenleaf WJ, Block SM. Single-molecule studies of RNA polymerase: motoring along. Annu Rev Biochem. 2008;77: 149–176. doi: 10.1146/annurev.biochem.77.073106.100741
6. Artsimovitch I, Landick R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. Proc Natl Acad Sci U S A. 2000;97: 7090–7095. doi:10.1073/pnas.97.13.7090
7. Gabizon R, Lee A, Vahedian-Movahed H, Ebright RH, Bustamante CJ. Pause sequences facilitate entry into long-lived paused states by reducing RNA polymerase transcription rates. Nat Commun. 2018;9: 2930. doi:10.1038/s41467-018-05344-9
8. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Michael Cherry J, et al. Gene Ontology: tool for the unification of biology. Nat Genet. 2000;25: 25–29. doi:10.1038/75556
9. The Gene Ontology Consortium, Carbon S, Douglass E, Good BM, Unni DR, Harris NL, et al. The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 2020;49: D325–D334. doi:10.1093/nar/gkaa1113
10. Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 2019;47: D419–D426. doi:10.1093/nar/gky1038
11. Santos-Zavaleta A, Salgado H, Gama-Castro S, Sánchez-Pérez M, Gómez-Romero L, Ledezma-Tejeida 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. doi:10.1093/nar/gky1077
12. Karp PD, Weaver D, Paley S, Fulcher C, Kubo A, Kothari A, et al. The EcoCyc Database. EcoSal Plus. 2014;6. doi:10.1128/ecosalplus.ESP-0009-2013

PUBLICATION IV

Positive supercoiling buildup is a trigger of *E. coli's* short-term response to cold shock

S. Dash*, C.S.D. Palma*, I.S.C. Baptista, M.N.M. Bahrudeen, B.L.B. Almeida, V. Chauhan, R. Jagadeesan, and A.S. Ribeiro. *Equal contributions.

bioRxiv, 2021
DOI: 10.1101/2021.12.22.473827

Positive supercoiling buildup is a trigger of *E. coli*'s short-term response to cold shock

Suchintak Dash[1,†], Cristina S.D. Palma[1,†], Ines S.C. Baptista[1], Mohamed N.M. Bahrudeen[1], Bilena L.B. Almeida[1], Vatsala Chauhan[1], Rahul Jagadeesan[1] and Andre S. Ribeiro[1,2,*]

[1] Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland.

[2] School of Sciences and Technology and Uninova CTS, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal.

[†] Equal contributions.

* To whom correspondence should be addressed. E-mail: andre.sanchesribeiro@tuni.fi.

Present Address: Andre S. Ribeiro, Arvo Ylpön katu 34, 33520, Tampere University, Finland.

Abstract

Adaptation to cold shock (CS) is a key survival skill of gut bacteria of warm-blooded animals. In *E. coli*, this skill emerges from a complex transcriptional program of multiple, timely-ordered shifts in gene expression. We identified short-term, cold shock repressed (CSR) genes by RNA-seq and provide evidence that their variability in evolutionary fitness is low and that their responsiveness to cold emanates from intrinsic features. Given that their single-cell variability in protein numbers increases after CS, we hypothesized that the responsiveness of a large portion of CSR genes is triggered by the high propensity for transcription locking due to positive supercoiling buildup (PSB). We then proposed a model of this phenomenon and, in support, show that nearly half of CSR genes are highly responsive to Gyrase inhibition. Also, their response strengths to CS and Gyrase inhibition correlate and most CSR genes increase their single-cell variability in protein numbers. Further, during CS, the cells' nucleoid density increases (in agreement with increased numbers of positive supercoils), their energy levels become depleted (while the resolving of positive supercoils is ATP dependent), and the colocalization of Gyrases and the nucleoid increases (in agreement with increased time length for resolving supercoils). We conclude that high sensitivity to PSB is at the core of the short-term, cold shock responsive transcriptional program of *E. coli* and propose that this gene feature may be useful for providing temperature sensitivity to chromosome-integrated synthetic circuits.

Keywords

Cold shock response; Positive supercoiling buildup; RNA-seq; Single-cell Flow-cytometry; Model fitting; Bacterial gene expression.

1. Introduction

E. coli is widely found in the gut of warm-blooded animals in all natural habitats. It usually propagates to new hosts when the original host excretes (or perishes) [Phadtare et al., 1999]. For this, it becomes airborne until encountering new hosts. Thus, it will face (sometimes extreme) temperature downshifts. To cope with these, it has evolved a complex transcriptional program involving many genes [Jones et al., 1987; Phadtare et al., 2004]. Their responses are likely subject to regulatory mechanisms yet to be decoded, which are responsible for the implementation of physiological changes that enhance the chances of survival.

As other prokaryotes, *E. coli* halts cell division and undergoes an "acclimation phase", during which changes occur at a multi-scale level, from heterogeneous changes in the kinetics of transcription [Oliveira et al., 2016; Charlebois et al., 2018] and translation [Giuliodori et al., 2004; Farewell and Neidhardt, 1998; Phadtare et al., 1999; Keto-Timonen 2016; Madrid et al., 2002], up to decreasing in membrane fluidity [Mansilla et al., 2004; Yamanaka 1999] and increasing cytoplasmic viscosity [Oliveira et al 2016; Parry et al, 2014].

Measurements of transcriptomes at non-optimal temperatures revealed broad responses by specific gene cohorts [Phadtare et al., 2004; Arsène et al., 2000]. During cold shock (CS), a small gene cohort has a fast, transient response, another has a long-term response, while most other genes (including essential genes) remain stable [Phadtare et al., 2004]. This diversity of single-gene responses may be explained by the likely existence of multiple causes for their alterations in expression rates during CS. For example, studies using synthetic gene constructs suggest that temperature can affect the kinetics of rate-limiting steps in transcription initiation, such as the closed and open complex formations [Oliveira et al., 2016], and such effects can differ between promoters [Oliveira et al., 2019]. Other studies showed that temperature affects chromosomal DNA compaction [Goldstein et al., 1984, López-García et al., 2000], which is associated with supercoiling buildup [Stuger et al., 2002; Holmes et al., 2000]. Changing supercoiling buildup levels can cause genome-wide disturbances in gene expression [Travers et al., 2005; Dorman, 2006; Dorman et al., 2016; Peter et al., 2004]. Other influences may be indirect, e.g., temperature affects energy-dependent events, such as interactions between nucleoid-associated proteins (NAPs) and chromosomal DNA [Amit et al., 2003] which affect DNA topology, and thus transcription kinetics [Pruss et al., 1989; Liu et al., 1987; Ma et al., 2013].

Changes in DNA supercoiling may be a quick, efficient means to tune gene expression during stresses, including osmotic shifts [Cheung et al., 2003], oxidative stress [Weinstein-Fischer et al., 2000] and starvation [Drlica et al., 1992]. Many promoters of stress-inducible genes (such as virulence genes in pathogenic bacteria) are sensitive to changes in DNA supercoiling [Dorman, 1995; Dorman 1996]. Thus, it is possible that temperature-dependent changes in DNA superhelical density may be responsible for the responsiveness of some cold shock repressed (CSR) genes.

In agreement, a recent study [Oliveira et al, 2019] tracked RNA production at the molecular level by synthetic variants of the Lac promoter. It was shown that, at low temperatures, RNA production kinetics is weaker and noisier when the gene is chromosome integrated than when it is plasmid borne (in plasmids, supercoiling buildup should be much slower due to the annihilation of positive and negative supercoils [Liu et al., 1987]). They also showed the same phenomenon under Gyrase or Topoisomerase I repression, as well as in energy-depleted cells. Finally, by integrating data from [Phadtare et al., 2004] and [Peter et al., 2004] they hypothesized that CSR genes may exhibit atypical supercoiling-sensitivity.

Here, we subjected *E. coli* cells to CS and identified CSR genes by RNA-seq. We then investigated their common features and their transcription factor (TF) regulation during CS (Figure 1, step 1). Next, we used a YFP fusion library [Taniguchi et al., 2010] to measure CS effects on the single-cell protein levels of 30 CSR genes (Figure 1, step 2). Afterwards, we performed RNA-seq, under Gyrase inhibition, to detect positive supercoiling sensitive (PSS) genes and then identify which genes are both CSR and PSS (Figure 1, step 3). We further collected biophysical data on the chromosome structure and cell energy levels, to support the hypothesis that high supercoiling sensitivity (SS) may provide genes with enhanced short-term CS responsiveness (Figure 1, step 4). Finally, we proposed an analytical model of the dynamics of short-term CSR genes with high PSS (Figure 1, step 5).

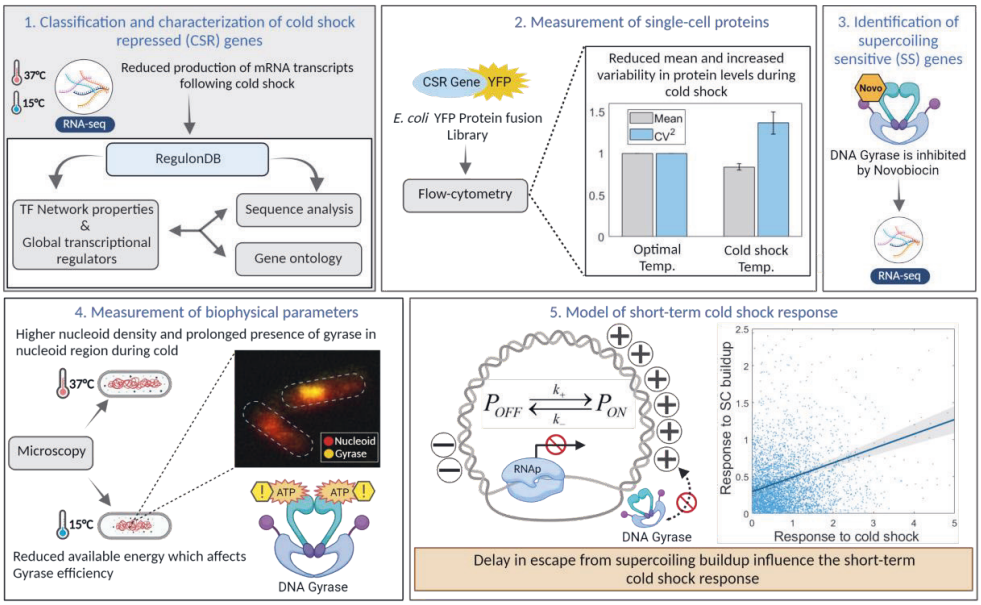

Figure 1: Workflow illustration. (1) Identification of short-term CSR genes from RNA-seq data and analysis of their functionality, sequences and, regulation by their direct input TFs and by global transcription regulators (e.g., RNAP). **(2)** Single-cell protein measurements of 30 CSR genes using a YFP fusion library in optimal and CS conditions. **(3)** Identification of PSS genes by RNA-seq following Gyrase inhibition. **(4)** Spectrophotometry and microscopy measurements of biophysical parameters such as ATP levels, cell size, and overlap between the nucleoid and Gyrase. **(5)** Analytical model of short-term CSR due to CS-enhanced locking of promoters due to PSB.

2. Material and Methods

2.1 Bacterial strains, growth conditions, and gene expression measurements

We used *E. coli* K-12 MG1655 for RNA and protein measurements, since it is the control strain of the YFP fusion library (Supplementary Table S1) [Taniguchi, et al. 2010]. From a glycerol stock (at -80°C), cells were streaked on LB agar plates and incubated at 37°C overnight. The next day, a single colony was picked from the plate, inoculated in fresh LB medium supplemented with antibiotics (34 μg/mL chloramphenicol for YFP tagged strains) and incubated at 30°C overnight with shaking at 250 RPM. Overnight culture cells were then diluted into fresh M9 media, supplemented with 0.4 % glucose, amino acids, and vitamin solutions, until reaching 0.03 OD (Optical Density at 600 nm measured by Ultrospec 10, Amersham biosciences, UK) and allowed to grow at 30°C with aeration until reaching the mid-exponential phase of growth (OD of 0.3). At this moment, the temperature was

downshifted (Innova® 40 incubator, New Brunswick Scientific, USA) and cells were incubated for another 180 mins. Cold-shock conditions are imposed by placing cells at 10-15°C [Phadtare et al., 2004]. Culture temperatures were monitored using a thermometer.

For measurements under Gyrase inhibition (Figure 6D), the antibiotic Novobiocin was added (50 μg/mL) upon reaching the OD ~0.3. This did not disturb the growth rate (Supplementary Figure S6). To measure RpoS, we used a MG*mCherry* (*rpoS*:mCherry) strain (kind gift from James Locke [Patange et al., 2018]), where the *rpoS* gene codes for σ^{38}, which is endogenously tagged with mCherry. For intracellular ATP measurements, we used the QUEEN 2m, a kind gift from Hiromi Imamura [Yaginuma et al., 2014] (Supplementary Table S1 for details).

Since all strains used contain the gene acrA, 50 μg/mL of Novobiocin is not expected to affect cell division rate [Ma et al., 1995]. We further verified this by measuring growth rates by OD_{600}. In agreement, growth rates only increase for 200 μg/ml or higher (Supplementary Figure S6).

We measured RNA and protein expression levels by RNA-seq (Supplementary Section I) and by Flow-cytometry (Supplementary Section II), respectively. We used pulse width data from flow-cytometry as a proxy for cell volume [Bahrudeen MNM et al., 2019; Cunningham et al., 1990; Traganos et al., 1984], which assisted the estimation of protein concentrations. We verified these results using microscopy data and image analysis (Supplementary Section III).

Finally, we measured the regions occupied by Gyrase and RNAP, using strains from the YFP fusion library, as described in Supplementary Section III.

2.2. Nucleoid visualization by DAPI

To study the effect of cold shock on nucleoid size (Figure 7A), cells were fixed with 3.7% formaldehyde in phosphate-buffered saline (PBS, pH 7.4) for 30 min at room temperature, followed by washing with PBS to remove excess formaldehyde. The pellets were suspended in PBS, and DAPI (4′,6-diamidino-2-phenylindole) (2 μg/mL) was added to the suspension to stain the nucleoid. After incubating for 20 min in the dark, cells were centrifuged and washed twice with PBS to remove excess DAPI. Cells were then re-suspended in PBS and 3 μL of these cells were placed on a 1% agarose gel pad for microscopy [Chazotte et al., 2011].

2.3. Cellular ATP levels

QUEEN-2m cells (Supplementary Table S1) were grown as described in Methods Section 2.1. We tracked ATP levels (Supplementary Figure S11) using a Biotek Synergy HTX Multi-Mode Reader (spectrophotometer). The solution was excited at 400 nm and emission was recorded at 513 nm. Similarly, the solution was re-excited at 494 nm and emission was recorded at 513 nm. The ratio of

513 nm emission intensity at these 2 excitation wavelengths, denoted as "400ex/494ex", is used to quantify cellular ATP concentration as proposed in [Yaginuma H et al., 2014].

2.4. Stochastic Model of cold shock response

We used stochastic simulations to estimate the expected noise in gene expression (as measured by the squared coefficient of variation, CV^2, of gene expression levels in individual cells), assuming the models described in Figure 6A. Simulations were performed using SGNSim [Ribeiro et al., 2007], whose dynamics follows the Stochastic Simulation Algorithm [Gillespie, 1976; Gillespie, 1977]. The time length of each simulation was 10^6s, which sufficed to avoid fluctuations due to sources other than noise in gene expression [Häkkinen et al., 2016]. The results [Figure 6B] were collected from 100 runs for each model, which sufficed to obtain consistent results. Finally, at the start of each run, in addition to the parameter values in Supplementary Table S8, it was set that there is 1 promoter in the system. In Model 1.3 (Figure 6A), the promoter was initially in the "ON" state.

2.5. Information from RegulonDB

Our data was extracted from RegulonDB v10.9. The data includes information on TF interactions, operon organization, and nucleotide sequence.

3. Results

3.1 Cell morphology, physiology, and master transcription regulators during cold shock

Having subject cells to CS (Methods Section 2.1), we first studied physiological and morphological effects. Once at 15ºC or lower temperatures, cells no longer divided (Figure 2A). Meanwhile, their size was not affected, according to microscopy (Supplementary Figure S7B) and flow-cytometry (Figure 2C, Supplementary Section II) data. Nevertheless, these cells are not likely to be shifting to stationary growth, since RpoS concentrations remain low (Supplementary Figure S15) [Lange, 1991; Jishage et al 1996] (Methods Section 2.1), when compared to cells in optimal conditions and to cells in the stationary growth phase (Figure 2B).

Next, we examined potential short-term effects of CS on the concentrations of the master regulators of transcription, since if they change, it could influence the dynamics of CSR genes. In detail, we observed RNA polymerase (RNAp) by tracking a YFP tagged β subunit, which is the product of the rpoB gene (Supplementary Table S1). We also observed the two subunits of Gyrase (GyrA and GyrB) and of Topoisomerase I (TopA and TopB) using a YFP fusion library [Taniguchi et al., 2010], since they are the master regulators of DNA supercoiling levels [Gellert et al., 1976; Wang, et al., 1971]. As

such, they heterogeneously influence transcription at a genome-wide level. Further, evidence suggests that the efficiency of Gyrase and Topoisomerase is temperature sensitive [Drlica 1992; Wang et al., 1985; Oliveira et al., 2019].

Neither of these global regulators showed concentration changes during 80 mins after CS (Figure 2D), while the RNA-seq measurements reported below to identify CSR genes were performed 20 min after CS. As such, short-term CS responsiveness, is not expected to be activated by changes in the concentrations of these master transcription regulators.

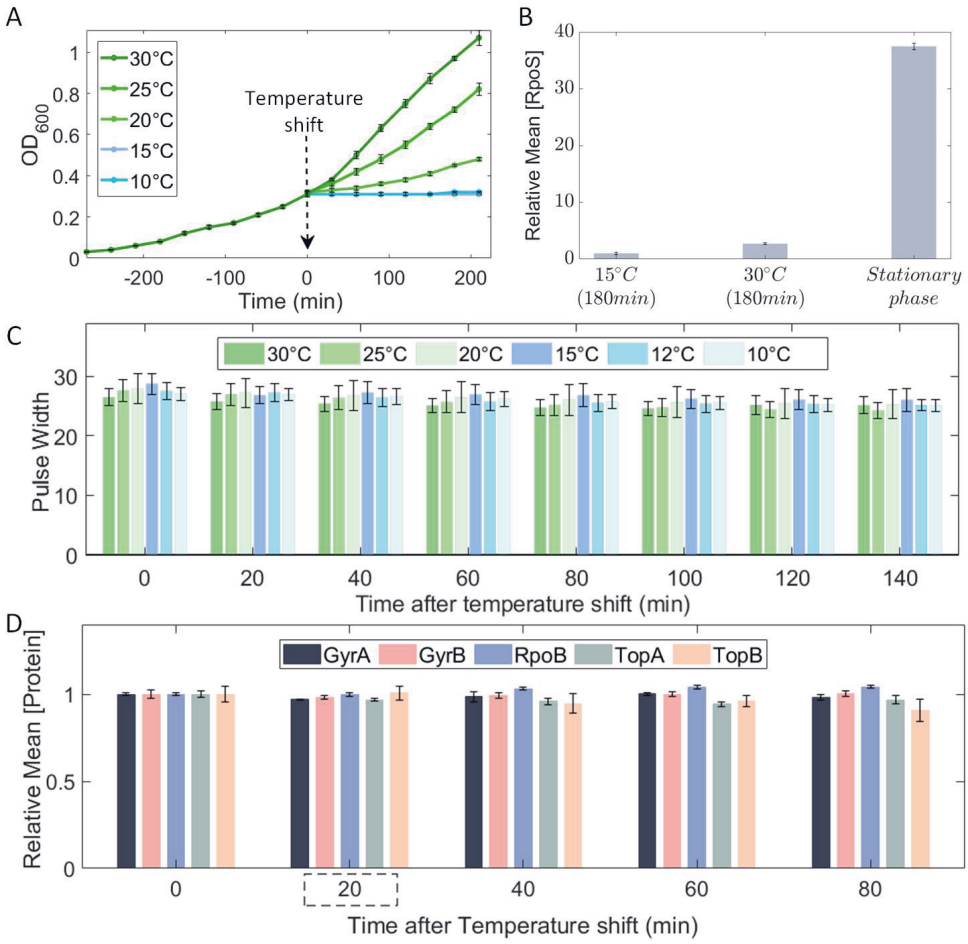

Figure 2. Effects of temperature shifts on cellular morphology, physiology, and global transcriptional regulators. (A) Growth curves at 10°C, 15°C, 20°C, 25°C, and 30°C following a temperature shift, set to be minute 0. **(B)** Mean RpoS concentration during CS and optimal conditions after 180 min, and during stationary growth (i.e., after 700 min). **(C)** Pulse width over time following temperature shifts (Methods Section 2.1). **(D)** Mean concentration of GyrA, GyrB, TopA, TopB and RpoB proteins over time after shifting temperature to 15°C. The vertical error bars are the standard error of the mean (SEM) from 3 biological repeats.

3.2 Identification of short-term CSR genes

We performed RNA-seq measurements (Supplementary Section I) at 0, 20, 80, and 180 min after shifting temperature to 15°C and under optimal (control) temperature (Methods Section 2.1).

We classified single-gene responses to CS as 'short-term' when they occur *prior to* influence from direct input TFs or global regulators, or from cell division. As such, based on cell doubling times (Figure 2A) and on known rates of transcription and translation in *E. coli* (see e.g. [Bernstein et al., 2002; Taniguchi et al., 2010]) we expect the changes in RNA numbers at 20 min. after the CS to be short-term, while subsequent changes at 80 and 180 min are here classified as being mid-term and long-term changes, respectively.

To identify short-term CSR genes, we obtained the RNA $\log_2$ fold changes (LFC_{CS}) at 20 mins after shifting to cold shock. We also obtained control LFCs (LFC_{CTRL}) after the same time interval when not shifting temperature.

We classified a gene as 'CSR' when its $LFC_{CS} < 0$ (with p-value < 0.05), provided that its corresponding $LFC_{CTRL} \geq 0$ (with p-value < 0.05) as this enhances the chance that the repression at CS was due to the CS. We found that 381 genes (Supplementary File X2) respected these conditions and, using the YFP fusion library [Taniguchi et al., 2010], one can measure the proteins levels of 124 of them.

From these 124, we selected genes that: i) have high expression under optimal conditions (resulting in higher fluorescence than cell backgrounds) and; ii) $LFC_{CS} < -0.23$, i.e., their RNA levels were reduced by 15% or more, relative to the same RNA in the control condition, to ensure significant downregulation during CS at the protein level. We found 30 of the 124 genes respected these conditions and, thus, we selected them for single-cell fluorescence measurements in the control and CS conditions.

Finally, we selected 6 of these 30 genes and additionally collected single-cell, time-lapse flow-cytometry data on their dynamics. Taken together, their expression levels cover the state space of protein expression levels of the 30 CSR genes.

3.3 Ontology and evolutionary fitness of short-term CSR genes

We investigated the ontology [Ashburner et al., 2000; Gene Ontology Consortium, 2021] of CSR genes to identify the most affected biological processes. From an over-representation test

(Supplementary Section XVI), we compared the number of CSR genes related to specific biological processes (quantified by the fold enrichment) with the *expected* number, given genome-wide numbers.

The significantly over-represented biological processes are listed in Table S11. Visibly, of 30 major biological processes in *E. coli* considered in GO studies [Ashburner et al. 2000; Gene Ontology Consortium, 2021], CSR genes are mainly associated with metabolism and response to external stimulus (Supplementary Figure S14). This agrees with reports that genes involved in metabolism are commonly affected during CS, which reduces growth rate and the rate of glycolysis [Gadgil et al., 2005; Andersen et al., 1980; Phadtare et al 2004].

Next, we studied the evolutionary fitness of the responsive genes (Supplementary section XVII). Interestingly, while their average fitness is the same as expected by chance, their fitness variability is smaller than in same-sized cohorts of randomly selected genes (Figure 3A). This is not because they are over-represented in two functional groups, since the fitness variability of random cohorts with the same distribution of gene functions (164 genes related to metabolism, 41 genes responsive to stimulus, 36 genes in both groups, and 140 with other functions) also have statistically distinguishable fitness variability from CSR genes. Given that the fitness is positively correlated to the evolutionary conservation (Supplementary Section XVII), we hypothesize that their evolutionary ages are likely to be more similar than expected by chance as is the fitness.

In addition to ontology and fitness, we also investigated the potential influences on CS responsiveness from TFs (Supplementary Section XVIII) and promoters' AT richness (Supplementary Section XIX). However, we failed to find any relationships.

3.4 Short-term responses of CSR genes can be partially explained by operon organization

Genes in the same operon commonly exhibit co-expression [Jacob, F. and Monod, J, 1961; Sabatti et al, 2002]. To verify if this influences the composition of the identified population of CSR genes, we confronted the positionings of CSR genes in the same operons with correlations in their dynamics.

Of the 381 CSR genes, 169 are not in operon structures (according to RegulonDB [Santos-Zavaleta A et al., 2019]), while the remaining 212 are organized in a total of 111 operons (Figure 3B, Supplementary File X3). As expected, the LFC_{CSR} of pairs of the CSR genes in the same operon are (similarly) correlated in both optimal and CS conditions (Figures 3C and 3D).

We confronted these data with a null model that assumes the same distribution of genes per operon as in Figure 3B, but with the genes composing those operons being randomly selected from the set of genes in operons. The random pairs showed no dynamic correlation (Figures 3C and 3D). We

conclude that the operons' organization might partially explain the numbers of CSR genes of *E. coli*, i.e., some genes might be CSR because they are located downstream a CSR gene in the same operon.

Nevertheless, there are 60 operons with only 1 CSR gene (Figure 3B). Thus, for a gene to be CSR, it does not suffice to be in the same operon as a CSR gene.

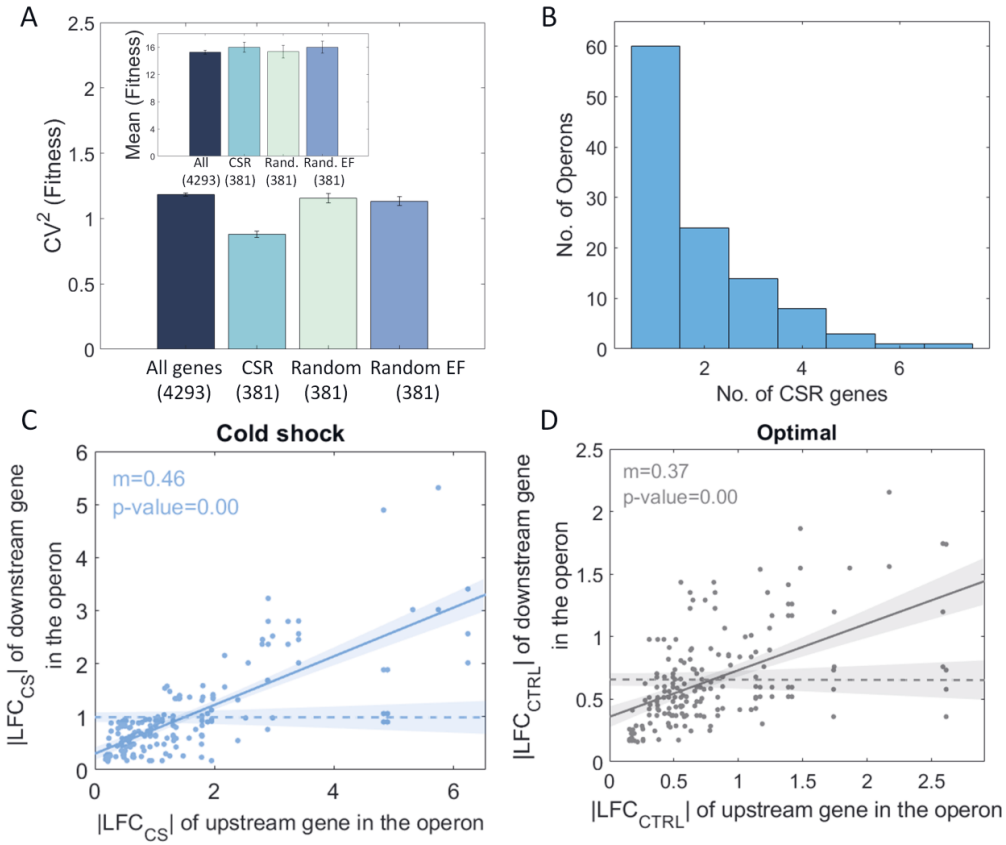

Figure 3. Characterization of CSR genes. (A) Bar plot of the variability, CV^2, of the fitness of all genes of *E. coli*'s genome (dark blue), CSR cohort (light blue), randomly selected cohort (light green) and a randomly selected cohort with the same size and same biological function (purple bar). The inset shows the mean fitness (in %) for each cohort. **(B)** Distribution of CSR genes in operons **(C)** Scatter plot between the $|LFC_{CS}|$ of pairs of CSR genes in the same operon during CS. **(D)** $|LFC_{CTRL}|$ of CSR genes upstream in the operon plotted against the $|LFC_{CTRL}|$ of CSR genes downstream in the same operon at optimal temperatures.

3.5 The scaling between noise and mean of single-cell CSR protein numbers is temperature dependent

Given that the short-term response of CSR genes was uncorrelated with their input TFs dynamics (Supplementary Section XVIII), individual gene features are more likely to be responsible for their repression during cold shock. We expect that, by repressing gene expression at the transcription level, these mechanisms will affect how noise and mean expression relate [Peccoud et al., 1995; Golding et al., 2005; Taniguchi et al., 2010]. To investigate this, we studied the single-cell distributions in protein numbers of 30 CSR genes (Methods Section 2.1, Supplementary Table S1, Supplementary File X1).

To quantify single-cell protein numbers, we first corrected the statistical moments of the distributions to account for cell auto-fluorescence (Supplementary Section IV). Then, we plotted the mean expression levels in optimal conditions against the corresponding protein numbers reported in [Taniguchi et al., 2010] (Supplementary Figure S3). Given the best fitting line, from here onwards we convert protein expression levels into protein numbers using a scaling factor of 0.1. Meanwhile, we did not find correlations between protein levels and cell size (Supplementary Figure S11A), in agreement with the lack of change in cell size with CS (Figure 2C and Supplementary Figure S7) implying that cell size is not affecting single-cell expression levels. Finally, as expected from the mechanical coupling between transcription and translation in *E. coli* (Miller et al., 1970), the changes with CS in these 30 protein numbers correlated to the changes in the corresponding RNA numbers (Supplementary Figure S11B, Supplementary section XI), indicating that protein levels can be used to study the effects of regulatory mechanisms of transcription.

We thus plotted the mean single-cell protein numbers of CSR genes, M, against the corresponding noise, as measured by CV^2, for each gene and best fitted the data by Ordinary Least Squares (OLS) with the function [Bar-Even et al., 2006; Taniguchi et al., 2010]:

$$CV^2=\Omega/M \qquad \text{(eq. 5.1)}$$

where Ω is a constant and M are mean protein numbers (estimated in Supplementary Figure S3). From Figure 4A, the inverse proportionality between CV^2 and M, previously observed in optimal conditions [Bar-Even et al., 2006; Taniguchi et al., 2010; Newman et al., 2006], is valid during CS, but CV^2 becomes higher for the same M (Ω ~26% higher than in optimal conditions). Meanwhile, since Ω does not change from 120 min to 180 min after the cold shock, changes likely occurred prior to 120 min (Supplementary Figure S2).

To further investigate how Ω changed following CS, we measured single-cell distributions of protein levels of 6 genes each 20 min for 140 min following the temperature shifts. These genes (aldA, feoA, manY, ndk, pepN, tktB) have mean protein levels that cover the state space of M of the 30 CSR genes. For each time moment, we extracted the corresponding Ω that best fits the data (Figure 4B, Supplementary Figure S4). Visibly, Ω increases with time during CS, but not at optimal temperatures (Supplementary Figure S16).

Namely, at T ≤ 15 ºC, 40 min after CS, there is a sharp increase in **Ω**, while at T > 15 ºC, **Ω** remains constant. In detail, for CS temperatures (10 °C, 12 °C and 15 °C), the data is best fit by a sigmoid curve (R^2= 0.96) of the type $\frac{L}{1+e^{-0.1\cdot(x-x_0)}}$, where L is the curve's maximum value and x_0 is the value of the sigmoid midpoint (we also attempted to fit polynomials up to the several order, but none fitted better). Meanwhile, for the set of control temperatures, the data is best fit by a first order polynomial (R^2= 0.91).

Overall, we suggest that, as CS is applied, a step emerges in transcription that is responsible for the strong repression of CSR genes, which not only reduces expression levels of CSR genes, but it also increases the scaling factor between noise and mean of protein numbers.

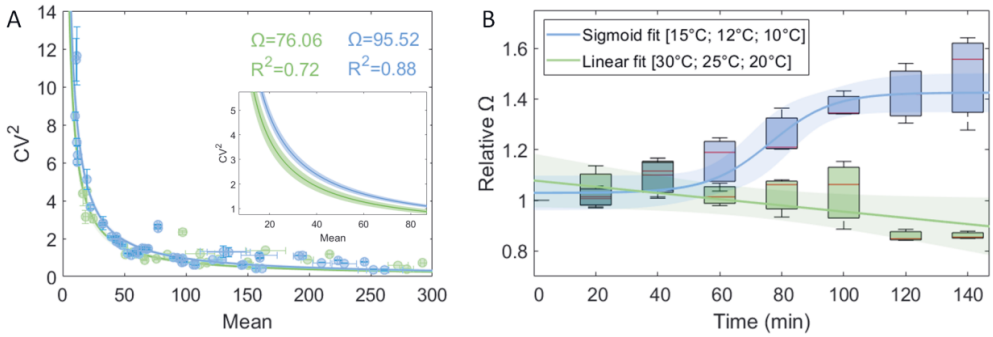

Figure 4. Relationship between CV^2 and mean protein numbers over time, at different temperatures. Blue corresponds to cold shock conditions, while green corresponds to optimal conditions. **(A)** Squared coefficient of variation (CV^2) versus mean protein numbers of 30 CSR genes (Supplementary file X1). Data at 120 and 180 min was merged as they did not differ (Supplementary Figure S2). We performed a 2-sample t-test to test the null hypothesis that there is no difference between **Ω** at 30°C and 15°C. The test rejected the null hypothesis with a p-value of 0.02. **(B)** Box plot of **Ω** over time at 'control' and 'cold shock' temperatures. The red line in the box is the median and the top and bottom of the box are one standard deviation (STD) above and below the median, respectively. For control and cold-shock temperatures, we fit the best fitting function. We performed an F-test on the regression model, which tests for the hypothesis that the 0 order polynomial fits significantly better than a 1st order polynomial. The test did not reject the null hypothesis that the best fit line is a horizontal line (p-value of 0.06). The lines correspond to the best-fit functions that maximize R^2.

Finally, from [Taniguchi et al., 2010], most protein number distributions in optimal conditions are well described by a Γ distribution. In these, both (eq. 5.1) is valid, as well the skewness can also be written as a function of M (Derivation in Supplementary Section Va) as follows:

$$S = \frac{2}{\sqrt{M}} \cdot \sqrt{\Omega} \qquad \text{(eq. 5.2)}$$

Given the Ω values above, we estimated S using (eq. 5.2) and compared to the empirical values of S in cold shock and control conditions (Figure 5A-B). We find that the two correlate linearly (see also Supplementary Figure S5), above the noise floor, which was estimated using the data in Figure 4A (Supplementary Figure S10, Supplementary Section XIV). This suggests that the effects of CS propagate up to the third moment of the single-cell distribution of protein numbers.

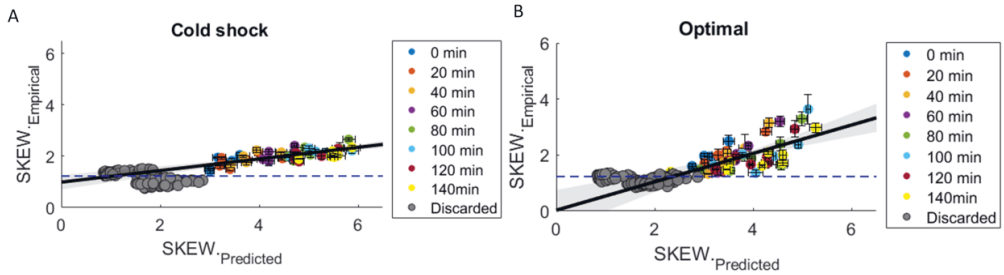

Figure 5. Correlation between empirical and predicted skewness. (A) CS temperatures (15°C, 12°C and 10°C). Skewness is predicted using equation 5.2 and the empirical values of Ω (Section 3.5). **(B)** Control temperatures (30°C, 25°C and 20°C). Meanwhile, empirical data on skewness is extracted from single-cell distributions obtained by flow-cytometry (Supplementary File X1) after being corrected for background noise. Blue dashed line is the estimated lower bound (Supplementary Section XIV). Grey circles are data points excluded from the fitting due to being below or crossing the noise floor.

3.6 An ON-OFF model can explain the short-term dynamics of CSR genes

From past studies [Taniguchi et al., 2010], *E. coli* transcription in optimal conditions can be well modeled as a one-step process (reaction 1.1 in Figure 6A). Using this reaction, along with reactions for translation (reaction 2 in Figure 6A) and RNA and protein decay due to degradation and dilution in cell division (reactions 3 and 4, respectively), one can model the dynamics of RNA and protein numbers in single-cells. Using this model, we derived an analytical solution for Ω (Supplementary Section Vb):

$$\Omega = 1 + \frac{k_2}{\lambda_1 + \lambda_2} \qquad \text{(eq. 6.1)}$$

where k_2 is the translation rate, and λ_1 and λ_2 are the RNA and protein decay, respectively. Given this, and since $\lambda_1 >> \lambda_2$ [Koch and Levy, 1955; Taniguchi et al., 2010], Ω would then necessarily be controlled by (k_2 / λ_1). Because of it, this model cannot explain the selection of the CSR gene cohort during CS.

First, we observed that the selection of this cohort occurs quickly, at the transcription level (Results Section 3.2), which excludes changes in k_2 as the main cause for the selection. In support, we observed that changes in RNA and protein numbers in CS are correlated (Supplementary Figure S11B) and thus, no particularly relevant regulation is expected to be occurring during translation. Finally, we failed to find any statistically significant differences in the RBS sequence of the RNAs coded by the CSR genes and randomly selected genes (Supplementary Figure S8, see also Supplementary Section VIII), in what regards their Shine-Dalgarno (Table S9) as well as their start codon sequences (Table S10).

Second, RNA degradation in *E. coli* does not correlate with RNA sequence, abundance, or metabolic function [Bernstein et al., 2002; Chen et al., 2015; Deutcher et al., 2006], nor with cell responses to acute events [Bernstein et al., 2002]. Thus, we do not expect that changes in λ_1 of CSR genes contributes to their selective responsiveness to CS.

We thus hypothesized that another mechanism, not present in the one-step model, ought to be responsible for the selective responsiveness of CSR genes, which includes the non-linear shift in the relationship between mean and noise (Figure 4B). We thus considered the emergence of an additional rate-limiting step in transcription initiation. A similar phenomenon has been reported [Buc and McClure, 1985] to occur on a synthetic promoter when shifting to temperatures lower than 20 ºC. Also, it has been reported that tuning the supercoiled state of the DNA template can oppose this effect [Buc and McClure, 1985].

If the origin of the reduction in transcription rates is supercoiling buildup, then it can be accounted for by an ON-OFF process (e.g., by replacing reaction 1.1 by reactions 1.3, Figure 6A) [Chong et al., 2014; Palma et al., 2020; Oliveira et al., 2016]. Else, if the reduction results from the slowdown of the forward kinetics of transcription initiation, e.g., due to an isomerization process preceding open complex formation (Buc and McClure, 1985), then it can be modeled by two forward, rate-limiting steps in RNA production (e.g., by replacing reaction 1.1 by reactions 1.2, Figure 6A).

To determine which model is more realistic, we estimated their noise (CV^2) for the same mean expression level. In detail, we tuned the three models so that they match in mean expression, by enforcing the relationships shown in the inset of Figure 6A between the rate constants (specific parameter values in Supplementary Table S8). Next, we performed stochastic simulations (Methods Section 2.4) and found that the ON-OFF model is the only one with higher noise than the one-step model and, thus, fits better the increase in Ω, at low temperatures (Figure 6B) meaning that, in CS, CV^2 is higher than at optimal temperatures (for the same mean expression levels).

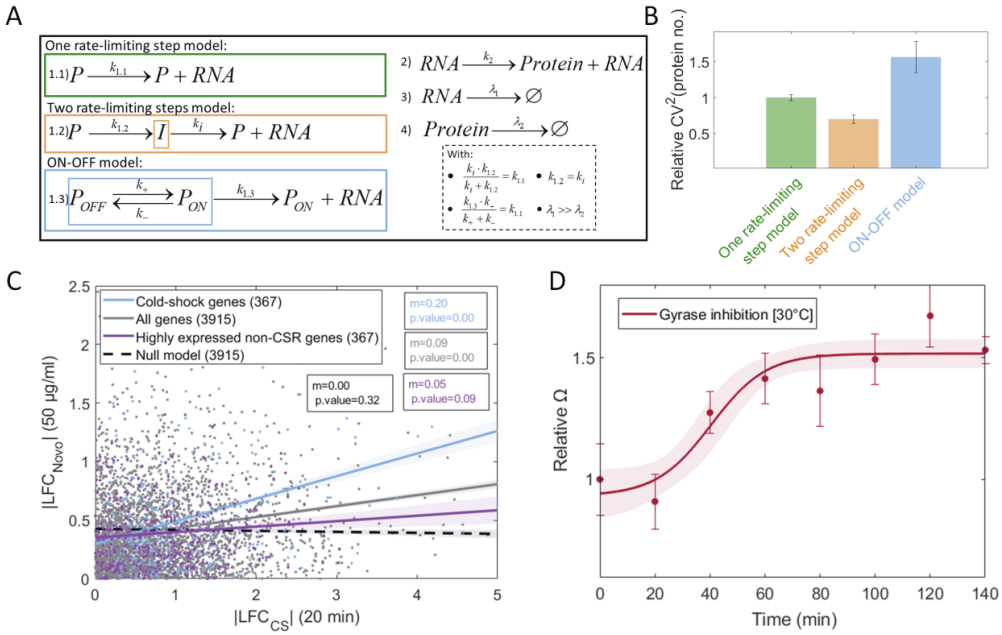

Figure 6. Nature of the short-term cold shock responses. (A) Three models were considered, differing in the reaction of transcription (reaction 1.1 for a one rate-limiting step model, reactions 1.2 for a two rate-limiting steps model, and reactions 1.3 for an ON-OFF model). All models include the same reactions for translation and RNA and protein decay (reactions 2, 3, and 4, respectively). The inset shows the conditions that the rate constants of the models must respect to impose identical mean protein numbers between them. **(B)** CV^2 of protein numbers (relative to the one step model), as predicted by each model, assuming the parameter values in Table S8. Vertical error bars are the SEM. **(C)** Scatter plots of $|LFC_{NOVO}|$ after subjecting cells to 50 µg/mL Novobiocin (relative to a control condition, absent of Novobiocin) versus the $|LFC_{CS}|$ of mRNAs 20 min after shifting to 15°C (Methods Section 2.1). The Temperature RNAseq data has 4328 genes, while the Novobiocin dataset has 3948 genes. Common to both datasets are 3915 genes (grey circles). From the 3915 genes found, the blue circles correspond to 367 CSR. To the data, we fitted by OLS the best fit line (grey line and blue line, respectively). As a null model, we randomized the |LFC| for grey circles and fit by OLS the best-fit line between the correlation of the random pairs (dashed grey line). For each fit, we performed a likelihood ratio test between the zero-order polynomial and higher-order polynomials. P-values < 0.05 reject the null hypothesis that the best fit line is a horizontal line. **(D)** Effects of Novobiocin over time on **Ω**, which is best fit by a sigmoid function ($\frac{L}{1+e^{-0.1\cdot(x-x_0)}}$), where L is the curve's maximum value and x_0 is the value of x of the midpoint of the sigmoid function. The red curve corresponds to the best-fit curve (that maximize R^2). Measurements are performed by flow cytometer every 20 min, for 140 min. For each time point, we fit the function $CV^2=\Omega/P$ [Bar-Even et al., 2006; Taniguchi et al., 2010].

3.7 Response strength to cold shock is correlated with Positive Supercoiling sensitivity

We next explored the hypothesis that short-term responses to CS emerge from positive supercoiling sensitivity. For this, we performed RNA-seq after subjecting cells to 50 μg/mL Novobiocin, which inhibits Gyrase [Gellert et al., 1976; Mizuuchi et al., 1978] (Methods Section 2.1) and, thus, would cause a similar effect as CS if the hypothesis holds true.

From Figure 6C, the response strengths of CSR genes to Novobiocin are positively correlated to their response strengths to CS (blue balls in Figure 6C, p-value < 0.05), which supports the hypothesis. Further, CSR genes are more sensitive to Novobiocin than the average gene, i.e., have stronger responses, which further supports that they are more PSS.

On the other hand, it could instead be because their original expression in the control condition was relatively high, when compared to the average gene. To test this, we compared the response strength to Novobiocin of genes that are *not* CSR but, have similar expression levels in optimal conditions. We selected random cohorts with the same number of genes as the CSR cohort and the same average expression level in the optimal condition. We found that the best fitting line between their responses to CS and Novobiocin (Figure 6C) has a smaller slope than the line for CSR genes (and the two slopes can be statistically distinguished). We conclude that it is not their high expression level that explains why CSR genes are also PSS.

Given the above, we hypothesized that PSS is a key underlying mechanism of the short-term transcriptional program of cold shock responsiveness. To find if CSR genes are also PSS, we considered the genes whose responses to CS were stronger. To select them, we set a threshold between weak and strong at $|LFC_{CS}| = 0.8$ (Supplementary Figure S19) since, below it, several p-values are close to the significance level (Supplementary Section I).

Next, to investigate if genes with strong CSR also have high PSS, we also needed to classify genes as having 'high' $|LFC_{NOVO}|$. For this, we considered that the inclination of the best fitting lines in Figure 6C likely differ with perturbation strengths (e.g., adding more than 50 μg/mL Novobiocin would cause stronger LFCs [Palma et al., 2020]). Since, on average, the response strength of CSR genes to 15 ºC was twice as strong as their response to 50 μg/mL Novobiocin, we classified responses of $|LFC_{NOVO}| > 0.4$ as 'strong' (if their p-value < 0.05). Given this, 1215 out of 3948 genes of *E. coli* (~31%) were classified as having a strong response to Novobiocin.

Given the classifications, of the 381 genes classified as CSR, 201 are strongly responsive to CS. Of these, we considered only 190, since the other 11 failed to obey the filtering criteria *iii*, in step I.b. in

Supplementary Section I. Of the 190, 92 are strongly responsive to Novobiocin. Thus, approximately 50% of the CSR genes are also PSS, which is higher than expected by random chance.

Given this, we hypothesized that high sensitivity to PSB is at the core of the short-term, cold shock responsive transcriptional program of *E. coli*. Nevertheless, we also conclude that PSB is not the only means by which genes can be part of the cohort of quickly repressed genes during cold shocks. Finally, it is worth noting that, 1215 out of 3948 genes of *E. coli* exhibited $|LFC_{NOVO}| > 0.4$ (p-value < 0.05), but only 92 of them were CSR. Thus, being PSS is not sufficient to be strongly, short-term CSR.

Given this, from *in vivo* single-cell, time lapse protein data (Methods section 2.1), we studied the dynamics of the 6 genes used to produce Figure 4B and investigated if their CS responsiveness is due to their PSS. In detail, if during CS, a rate-limiting step emerges in their dynamics (reaction 1.3 in the ON-OFF model in Figure 6A), we expect that the noise for a given mean expression level should be higher than during optimal conditions. For this, we further measured 4 additional CSR genes because they, in addition to pepN and ndk of the 6 genes, are the only ones out of the 381 CSR genes that: i) do not have any known input TFs and, thus, even in optimal conditions, should be less influenced by the TF network of *E. coli*; ii) their expression levels in control conditions are above background noise, and; iii) they are not integrated in a position of an operon structure other than the first one downstream the promoter.

Results in Supplementary Figure S18 show that, in accordance with the predictions, there is a decrease in mean protein levels during CS and Gyrase inhibition. Only two genes, pepN and feoA, exhibit increased levels, contrary to the model, after 60 and 100 mins following the addition of Novobiocin, respectively. This is, potentially, due to mid- and long-term phenomena (also respectively) occurring as part of the cellular response program to cold. For example, feoA has 4 input TFs, while pepN is closely spaced to another gene, ssuB, in a convergent configuration. Also, ssuB has no transcription termination site. As such, it can perturb pepN's expression, e.g., by first repressing and then stopping doing so, when under the effects of Novobiocin.

Meanwhile, the overall ratio between noise and mean following Gyrase inhibition (Figure 6D) fits well a sigmoid, as it did when subjecting cells to CS (Figure 4B). The main difference between Figures 4B and 6D is that it takes less 20 min for the shift to occur following Novobiocin addition. This might be due to the slowing down of metabolic events during cold.

Finally, we note that the similarity in the mean changes in Ω is not used as criteria to support that the underlying mechanism is the same, because we tuned the Novobiocin levels to make the average strong of the perturbations similar.

3.8 The overlap between Gyrase and Nucleoid during CS increases

If the ON-OFF Model (Figure 6A) is in accordance with the process of gene expression during CS and, if Gyrases are responsible for removing promoters from their OFF state due to PSB [Chong et al, 2014], one would expect Gyrases to be more present in the DNA region during CS. Instead, if it was the two-step model (Figure 6A) that best described CS effects, then the RNAP would take longer to complete transcription initiation events and thus, would spend longer time at the DNA region.

To verify this, we measured by microscopy (Supplementary section III), prior to and during CS, the cell areas occupied by GyrA and RpoB, respectively. We also assessed how these areas overlapped with the area occupied by the nucleoid, observed by DAPI staining (Methods section 2.2), since temperature is known to perturb chromosomal DNA compaction [Goldstein et al., 1984; López-García et al., 2000].

Soon after cells enter CS, the nucleoid area decreases for the entire period of our gene expression measurements (Figure 7A). Since past studies observed a similar phenomenon when inhibiting Gyrase [Palma et al., 2020] and since the nucleoid area is a good proxy for nucleoid density [Gray et al., 2019], whose increase is a common effect of PSB [Eriksson et al., 2002], we hypothesize that Gyrase duty cycles increase during CS. In agreement, the overlap between the Gyrase and nucleoid regions increase during CS (Figure 7D, see example cells), even though the 'Gyrase region' decreased in size relative to the nucleoid region (Figure 7B.1 and B.2). Contrary to this, the overlap between the RNAP and nucleoid region does not increase during CS (Figure 7E, see example cell), even though the 'RNAP region' increased in size relative to the nucleoid region (Figure 7C.1 and C.2) which is in line with reduced number of available promoters for transcription initiation.

Overall, these results support the hypothesis that the ON-OFF model explains the underlying mechanism of a large number of short-term CS responsive genes.

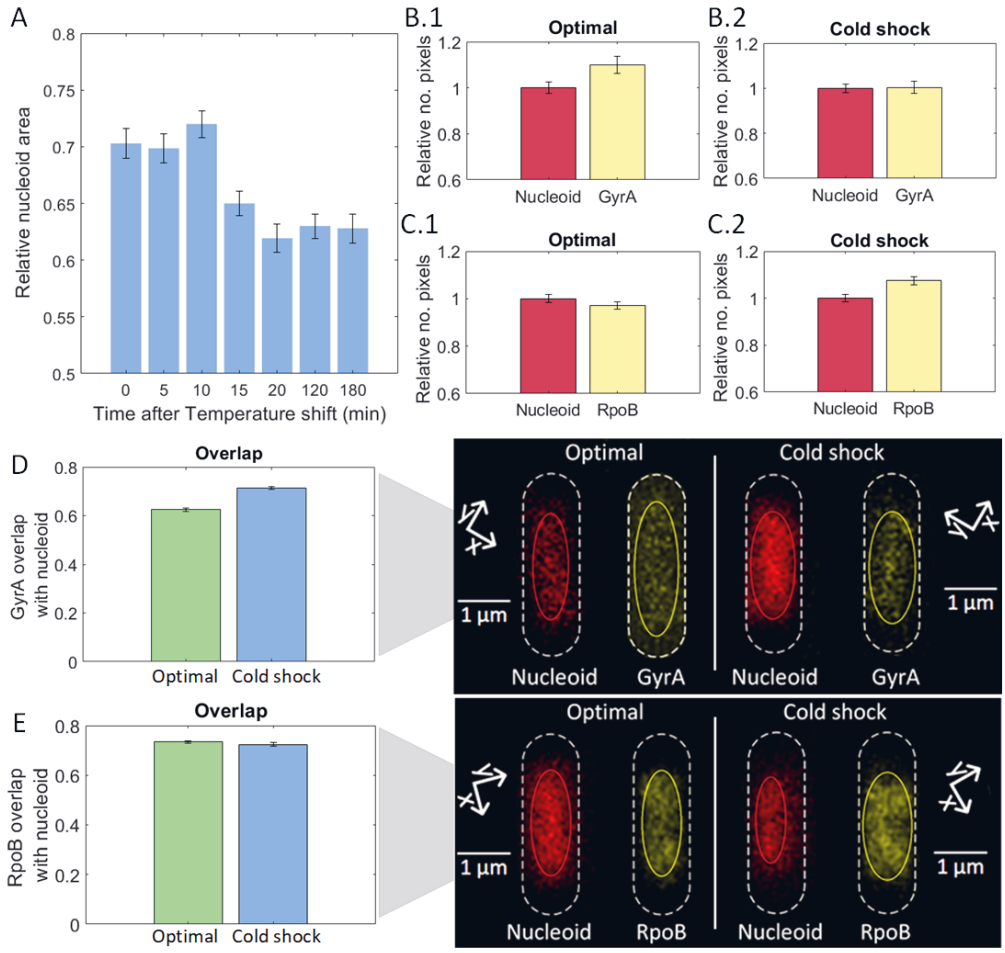

Figure 7. Model Study of biophysical parameters affected by cold shock; (A) Nucleoid area relative to cell area over time following CS, measured using 'SCIP' [Martins et al., 2018]. More than 500 cells analyzed per condition. **(B.1)** Size of YFP tagged GyrA relative to nucleoid size in optimal temperature. **(B.2)** Size of YFP tagged GyrA relative to nucleoid size after cold shock. **(C.1)** Size of YFP tagged RpoB (px) relative to nucleoid size (px) in optimal temperature. **(C.2)** Size of YFP tagged RpoB relative to nucleoid size after cold shock. **(D)** Relative overlapping between GyrA and the nucleoid in optimal and cold shock temperatures (Relative overlap is calculated as $\frac{R_{overlap}}{\left(R_{GyrA} + R_{nucleoid}\right)/2}$, where R is "size of the region occupied by"). **(E)** Relative overlapping between the RpoB and nucleoid in optimal and cold shock temperatures. Also shown, example microscopy images of the same cell, with the nucleoid stained with DAPI and the GyrA/RpoB proteins tagged with YFP in optimal and cold shock temperatures, respectively. More than 400 cells analyzed per condition. Vertical error bars correspond to the standard error of the mean. Size is defined by number of pixels.

3.9 Cellular energy levels decrease during cold shock

Gyrase ability to remove supercoils is ATP dependent [Rovinskiy et al., 2012; Gubaev et al., 2016]. Also, Gyrase numbers did not increase during CS (Figure 2D). As such, a decrease in ATP levels could contribute to make PSB an 'efficient' mechanism underlying CSR genes. To study this, we measured ATP levels in the control and CS conditions (Methods section 2.3). These levels quickly decrease during CS (Supplementary Figure S13). This furthers supports the ON-OFF Model, as it suggests that the average time to escape OFF states increases in CS.

3.10 Relative Ω as a function of OFF-ON rates

From above, we expect CS to alter how Ω is regulated due to the emergence of an ON-OFF step controlling transcription. We estimated the expected ratio between values of Ω at cold and optimal conditions assuming ON-OFF and one-step models (Figure 6A), respectively (Supplementary Section V.e.). From there:

$$\frac{\Omega_{15°C}}{\Omega_{30°C}} \propto \left(1 + \frac{k_1.k_-}{(k_+ + k_-)^2}\right) \qquad \text{(eq. 10.1)}$$

Equation 10.1 informs on how the ratio $\Omega_{15°C} / \Omega_{30°C}$ is expected to be affected by the rate constants controlling the ON-OFF steps, k_+ and k_- (Reactions 1.3 in Figure 6A), and the transcription activity from active promoters, k_1.

We do not expect k_1 to be a major regulator of this ratio, since this rate constant is present in the one-step model, which was unable to mimic the measurements. Meanwhile, of the two remaining events controlling promoter activity (Reactions 1.3, Figure 6A) only promoter escape from the OFF state (regulated by k_+) is energy consuming [Gubaev et al., 2016]. Thus, this event is expected to be most decelerated one during CS. We therefore hypothesized that k_+ is the most temperature sensitive parameter in equation 10.1.

We therefore investigated the relationship between k_+ and temperature. We explored four temperature sensitive models of k_+ which were fitted to the empirical data from Figure 4B. Models and best fitting parameter values are shown in Supplementary Table S12, while results of the fitting are shown in Supplementary Figure S12. From the R^2 values, the best fitting model assumes that k_+ changes over time following an exponential function.

4. Discussion

We identified a large number of short-term CSR genes and studied what causes their quick repression in CS. A few of them are likely responsive due to being in an operon with upstream CSR genes, but the majority appear to be independently responsive to CS. Interestingly, following CS, CSR genes rapidly decrease expression level, while their noise relative to the mean expression increases. This increase in noise is consistent with the emergence of transient locking events of transcription. Since a similar phenomenon was observed following Gyrase inhibition [Chong et al., 2014; Palma et al., 2020] and because we observed here that a large number of CSR genes is also highly sensitive to PSB, we hypothesized that the responsiveness of a large number of CSR genes emerges from their PSB responsiveness. Meanwhile, we observed that Gyrase converges to the nucleoid and that cell energy decreases during CS, suggesting that the number of promoters locked due to PSB increases during CS. We therefore proposed a model of the responsiveness of CSR genes based on their temperature sensitive PSB.

To our knowledge, temperature sensitive PSB is the first identified physical mechanism of how an *E. coli* gene can be CS responsive, and evidence suggests that it is present in nearly half of *E. coli*'s CSR genes. This finding, first hypothesized in [Oliveira et al., 2019], opens an avenue for the engineering of future synthetic, temperature sensitive and temperature resistant gene regulatory circuits, whose functioning could be tuned by the adaptive regulation of Gyrase activity. Further, we expect that it will contribute to learning how the short- and long-term transcriptional programs of *E. coli* responsive to CS have evolved.

Nevertheless, we found that not all short-term CSR genes (~ 50% of them) are responsive to PSB. As such, there must be other mechanisms by which genes become short-term CSR. Similarly, we also observed that not all genes responsive to PSB are also short-term CSR, implying that being responsive to PSB does not suffice to be CSR. Finally, it also remains unclear how PSB percolates genes in the same operon in a manner that, while some genes downstream a CSR gene are also CSR, many are not.

Given this, much study is still needed to identify the various possible set of features that can make a gene CSR. Potentially, genes could be CSR by CS-based locking of their rate-limiting steps during transcription initiation, as reported for a synthetic promoter [Buc and McClure, 1985]. This could explain how, in 4 out of the 30 genes measured at the protein level, noise did not increase although the mean levels decreased. However, we do not expect this to be common in most CSR genes, since the size of the cellular region occupied by RNAP during CS increased. Finally, it may be that, in some genes, their RNA or proteins have increased decay rates during CS, rather than altered production rates.

Bacterial transcriptional programs of cold shock responsiveness are critical survival skills that affect a wide range of vital Human activities [Phadtare et al., 2004]. It should be possible to perturb this ability by introducing it in the cells' synthetic circuits, with a temperature sensitivity based on the PSS

of its component genes. By interfering with CSR genes that rely on PSB, from our results, we expect that half of the transcriptional program of short-term response would be affected. As such, this is a viable strategy with potentially great rewards. For example, if we could activate natural bacterial CS transcriptional programs in the absence of cold, we would be able to slow down infections and food spoilage. We would also be able to enhance natural bacterial programs in low-temperature bioreactors (e.g., responsible for fermentation in the dairy industry), making them more cost-efficient. Meanwhile, if we could deactivate it when desired, we would be able to enhance bio fertilization and plant resistance to bacteria, among others.

Supplementary data

See supplementary file attached.

Author Contributions

A.S.R., C.S.D.P., and S.D. conceived the study and ASR supervised it. S.D. planned and executed the measurements, to which C.S.D.P. and V.C. contributed to. C.S.D.P. planned and executed the data analysis to which S.D., I.S.C.B., M.N.M.B., B.L.B.A. and R.J. contributed to. A.S.R., S.D., and C.S.D.P. drafted all documents, which were revised by all co-authors. S.D. and C.S.D.P. contributed equally and have the right to list their name first.

Funding

This work was supported by the Jane and Aatos Erkko Foundation [10-10524-38 to A.S.R.]; Suomalainen Tiedeakatemia to C.S.D.P.; Finnish Cultural Foundation [00200193 and 00212591 to I.S.C.B., and 50201300 to S.D.]; EDUFI Fellowship [TM-19-11105 to S.D.]; Tampere University Graduate Program to V.C., M.N.M.B., and B.L.B.A.. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of Interest

The authors have no competing interests.

References

Amit, R., Oppenheim, A.B., Stavans, J. (2003) Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. Biophys J. 84(4):2467-73. DOI: 10.1016/S0006-3495(03)75051-6.

Andersen KB, von Meyenburg K. (1908) Are growth rates of Escherichia coli in batch cultures limited by respiration? J Bacteriol. 144(1):114-23. DOI: 10.1128/JB.144.1.114-123.1980.

Arsène, F., Tomoyasu, T., Bukau, B. (2000) The heat shock response of Escherichia coli. Int J Food Microbiol. 55(1-3), 3-9. DOI: 10.1016/s0168-1605(00)00206-3.

Ashburner, M., et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25: 25–29. DOI:10.1038/75556

Bahrudeen, M.N.M., Chauhan, V., Palma, C.S.D., Oliveira, S.M.D., Kandavalli, V.K., Ribeiro, A.S. (2019) Estimating RNA numbers in single cells by RNA fluorescent tagging and flow cytometry. J Microbiol Methods 166:105745. DOI: 10.1016/j.mimet.2019.105745.

Bernstein, J. A., Khodursky, A.B., Lin, P.H., Lin-Chao, S., Cohen, S. N. (2002). Global analysis of mRNA decay and abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc Natl Acad Sci of the U.S.A., 99(15), 9697–9702. DOI:10.1073/pnas.112318199

Charlebois, D.A., Hauser, K., Marshall, S., Balázsi, G. (2018) Multiscale effects of heating and cooling on genes and gene networks. Proc Natl Acad Sci of the U.S.A.115(45):E10797-E10806. DOI: 10.1073/pnas.1810858115. Epub 2018 Oct 19.

Chazotte, B. (2011) Labeling nuclear DNA using DAPI. Cold Spring Harb Protoc 1: pdb prot5556

Chen, H., Shiroguchi. K., Ge H., Xie, X.S. (2015). Genome-wide study of mRNA degradation and transcript elongation in Escherichia coli. Mol. Syst. Biol., 11: 781. DOI: 10.15252/msb.20145794

Cheung, K.J., Badarinarayana, V., Selinger DW, Janse D, Church GM. A microarray-based antibiotic screen identifies a regulatory role for supercoiling in the osmotic stress response of Escherichia coli. Genome Res. 2003 Feb;13(2):206-15. DOI: 10.1101/gr.401003.

Cunningham, A., 1990. Fluorescence pulse shape as a morphological indicator in the analysis of colonial microalgae by flow cytometry. J. Microbiol. Methods 11, 27–36. DOI:10.1016/0167-7012(90)90044-7.

Dorman, C.J. (1995) Flemming Lecture. DNA topology and the global control of bacterial gene expression: implications for the regulation of virulence gene expression. Microbiology (Reading). 1995 141(6):1271-1280. DOI: 10.1099/13500872-141-6-1271.

Dorman, C.J. (1996) Flexible response: DNA supercoiling, transcription and bacterial adaptation to environmental stress. Trends Microbiol. 4(6):214-6. DOI: 10.1016/0966-842X(96)30015-2.

Dorman, C.J. (2006) DNA supercoiling and bacterial gene expression. Sci Prog. 89:151-66. DOI: 10.3184/003685006783238317.

Dorman, C.J., and Dorman, M.J. (2016) DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. Biophys Rev. 8(Suppl 1):89-100. DOI: 10.1007/s12551-016-0238-2.

Drlica, K., (1992) Control of bacterial DNA supercoiling. Mol Microbiol. 6(4):425-33. DOI: 10.1111/j.1365-2958.1992.tb01486.x.

Farewell, A., Neidhardt, F.C. (1998) Effect of temperature on in vivo protein synthetic capacity in Escherichia coli. J Bacteriol. 180(17):4704-10. DOI: 10.1128/JB.180.17.4704-4710.1998.

Gadgil, M., Kapur, V., Hu, W.S. (2005) Transcriptional response of Escherichia coli to temperature shift. Biotechnol Prog. 21(3):689-99. DOI: 10.1021/bp049630l.

Gellert, M., Mizuuchi, K., O'Dea, M.H., Nash, H.A. (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. Proc Natl Acad Sci U S A. 73(11):3872-6. DOI: 10.1073/pnas.73.11.3872.

Gellert, M., O'Dea, M.H., Itoh, T., and Tomizawa, J. (1976) Novobiocin and coumermycin inhibit DNA supercoiling catalyzed by DNA gyrase, Proc. Natl. Acad. Sci. U.S.A. 73 (12) 4474–4478, DOI: 10.1073/pnas.73.12.4474.

Gene Ontology Consortium. (2021) The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 8, 49(D1):D325-D334. DOI: 10.1093/nar/gkaa1113.

Gillespie, D.T. (1976) A general method for numerically simulating the stochastic time evolution of coupled chemical reactions, J. Comput. Phys. 22 (4), 403–434, DOI:10.1016/0021-9991(76)90041-3.

Gillespie, D.T. (1977) Exact stochastic simulation of coupled chemical reactions, J. Phys. Chem. 81 (1977) 2340–2361, DOI:10.1021/j100540a008.

Giuliodori, A. M., Brandi, A., Gualerzi, C. O., & Pon, C. L. (2004). Preferential translation of cold-shock mRNAs during cold adaptation. RNA (New York, N.Y.), 10(2), 265–276. DOI:10.1261/rna.5164904.

Goldstein, E., Drlica, K. (1984) Regulation of bacterial DNA supercoiling: plasmid linking numbers vary with growth temperature. Proc Natl Acad Sci U S A. 81(13):4046-50. DOI: 10.1073/pnas.81.13.4046.

Gray, W.T., Govers, S. K., Xiang, Y., Parry, B. R., Campos, M., Kim, S., & Jacobs-Wagner, C. (2019) Nucleoid Size Scaling and Intracellular Organization of Translation across Bacteria. Cell 177(6), 1632-1648.e20. DOI:10.1016/J.CELL.2019.05.017

Gubaev, A., Weidlich, D., & Klostermeier, D. (2016). DNA gyrase with a single catalytic tyrosine can catalyze DNA supercoiling by a nicking-closing mechanism. Nucleic Acids Research, 44(21), 10354. DOI:10.1093/NAR/GKW740

Häkkinen, A. and Ribeiro, A.S. (2016) Characterizing rate limiting steps in transcription from RNA production times in live cells, Bioinformatics 32 (9) 1346–1352, DOI: 10.1093/bioinformatics/btv744.

Holmes, V.F., Cozzarelli, N.R. (2000) Closing the ring: links between SMC proteins and chromosome partitioning, condensation, and supercoiling. Proc Natl Acad Sci U S A. 97(4):1322-4. DOI: 10.1073/pnas.040576797.

Jacob, F. and Monod, J. (1961) Genetic regulatory mechanisms in the synthesis of proteins. J. Mol. Biol. 3, 318–356.

Jones, P.G., VanBogelen, R.A., Neidhardt, F.C. (1987) Induction of proteins in response to low temperature in Escherichia coli. J Bacteriol. 169(5):2092-5. DOI: 10.1128/jb.169.5.2092-2095.1987.

Keto-Timonen, R., Hietala, N., Palonen, E., Hakakorpi, A., Lindström, M., Korkeala, H. (2016) Cold Shock Proteins: A Minireview with Special Emphasis on Csp-family of Enteropathogenic Yersinia. Front Microbiol. 22(7):1151. DOI: 10.3389/fmicb.2016.01151.

Koch L., Levy H.R. (1955) Protein turnover in growing cultures of Escherichia coli. J Biol Chem, 12;1955 217:947

Lange, R.P., and Hengge-Aronis, R. (1991) Identification of a central regulator of stationary-phasegene expression in Escherichia coli, Mol. Microbiol. 5, 49–59, DOI: 10.1111/j.1365-2958.1991.tb01825.x.

Liu, L.F., Wang, J.C. (1987) Supercoiling of the DNA template during transcription. Proc Natl Acad Sci of the U.S.A. 84(20):7024-7. DOI: 10.1073/pnas.84.20.7024.

López-García, P., Forterre, P. (2000) DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. Bioessays 22(8):738-46. DOI: 10.1002/1521-1878(200008)22:8<738:AID-BIES7>3.0.CO;2-5.

Ma, J., Bai, L., Wang, M.D. (2013) Transcription under torsion. Science 340(6140):1580-3. DOI: 10.1126/science.1235441.

Ma, D., Cook, D.N., Alberti, M., Pon, N.G., Nikaido, H., and Hearst, J.E. (1995) Genes acrA and acrB encode a stress-induced efflux system of Escherichia coli, Mol. Microbiol. 16 (1) 45–55, DOI: 10.1111/j.1365-2958.1995.tb02390.x.

Madrid, C., Nieto, J.M., Paytubi, S., Falconi, M., Gualerzi, C.O., Juárez, A. (2002) Temperature- and H-NS-dependent regulation of a plasmid-encoded virulence operon expressing Escherichia coli hemolysin. J Bacteriol. 184(18):5058-66. DOI: 10.1128/jb.184.18.5058-5066.2002.

Mansilla, M.C., Cybulski, L.E., Albanesi, D., de Mendoza, D. (2004) Control of membrane lipid fluidity by molecular thermosensors. J Bacteriol. 186(20):6681-8. DOI: 10.1128/JB.186.20.6681-6688.2004.

Miller, O. L., Hamkalo, B. A., and Thomas, C. A. (1970) Visualization of bacterial genes in action, Science 169, 392–395. DOI: 10.1126/science.169.3943.392

Mizuuchi, K., O'Dea, M.H., Gellert, M. (1978) DNA gyrase: subunit structure and ATPase activity of the purified enzyme. Proc Natl Acad Sci U S A. 75(12):5960-3. DOI: 10.1073/pnas.75.12.5960.

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., and Weissman, J. S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:7095, 441(7095), 840–846. DOI:10.1038/nature04785

Oliveira, S.M., Häkkinen, A., Lloyd-Price, J., Tran, H., Kandavalli, V., 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. DOI: 10.1371/journal.pcbi.1005174.

Oliveira, S.M.D., Goncalves, N.S.M., Kandavalli, V.K., Martins, L., Neeli-Venkata, R., Reyelt, J., Fonseca, J.M., Lloyd-Price, J., Kranz, H., Ribeiro, A.S. (2019) Chromosome and plasmid-borne $P_{LacO3O1}$ promoters differ in sensitivity to critically low temperatures. Sci Rep. 9(1):4486. DOI: 10.1038/s41598-019-39618-z.

Palma CSD, Kandavalli V, Bahrudeen MNM, Minoia M, Chauhan V, Dash S, et al. Dissecting the 681 in vivo dynamics of transcription locking due to positive supercoiling buildup. Biochimica et 682 Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2020;1863: 194515. doi: 683 10.1016/j.bbagrm.2020.194515

Parry, B.R., Surovtsev, I.V., Cabeen, M.T., O'Hern, C.S., Dufresne, E.R., and Jacobs-Wagner, C. (2014) The bacterial cytoplasm has glass-like properties and is fluidized by metabolic activity. Cell 156: 1–12. DOI: 10.1016/j.cell.2013.11.028

Patange, O., Schwall, C., Jones, M., Villava, C., Griffith, D.A., Phillips, A., Locke, J.C.W. (2018) Escherichia coli can survive stress by noisy growth modulation. Nat Commun. 9(1):5333. DOI: 10.1038/s41467-018-07702-z.

Phadtare, S. (2004) Recent developments in bacterial cold-shock response. Curr Issues Mol Biol. 6(2):125-36.

Phadtare, S., Alsina, J., Inouye, M. (1999) Cold-shock response and cold-shock proteins. Curr Opin Microbiol. 2(2):175-80. DOI: 10.1016/S1369-5274(99)80031-9.

Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O., Cozzarelli, N.R. (2004) Genomic transcriptional response to loss of chromosomal supercoiling in Escherichia coli. Genome Biol. 5(11):R87. DOI: 10.1186/gb-2004-5-11-r87.

Pruss, G.J., Drlica, K. (1989) DNA supercoiling and prokaryotic transcription. Cell. 56(4):521-3. DOI: 10.1016/0092-8674(89)90574-6.

Ribeiro, A.S., Lloyd-Price, J. (2007) SGN Sim, a stochastic genetic networks simulator, Bioinformatics 23 (6), 777–779, DOI:10.1093/bioinformatics/btm004.

Rovinskiy, N., Agbleke, A.A., Chesnokova, O., Pang, Z., Higgins, N.P. (2012) Rates of gyrase supercoiling and transcription elongation control supercoil density in a bacterial chromosome, PLoS Genet. 8 (8) e1002845, DOI: 10.1371/ journal.pgen.1002845.

Sabatti, C., Rohlin, L., Oh, M.K. and Liao, J.C. (2002) Co-expression pattern from DNA microarray experiments as a tool for operon prediction. Nucleic Acids Res. 30, 2886–2893. DOI: 10.1093/nar/gkf388

Santos-Zavaleta, A. et al. (2019). RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12., Nucleic Acids Res. 47(D1):D212-D220. DOI: 10.1093/nar/gky1077.

Stuger, R., Woldringh, C.L., van der Weijden, C.C., Vischer, N.O., Bakker, B.M., van Spanning, R.J., Snoep, J.L., and Westerhoff HV (2002) DNA supercoiling by gyrase is linked to nucleoid compaction. Mol Biol Rep. 29(1-2):79-82. DOI: 10.1023/a:1020318705894.

Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science 329(5991):533-8. DOI: 10.1126/science.1188308.

Traganos, F. (1984) Flow cytometry: principles and applications. I. Cancer investigations. Cancer Investig. 2, 149–163.

Travers, A., Muskhelishvili, G. (2005) DNA supercoiling - a global transcriptional regulator for enterobacterial growth? Nat Rev Microbiol. 3(2):157-69. DOI: 10.1038/nrmicro1088.

Wang, J.C. (1971) Interaction between DNA and an Escherichia coli protein omega. J Mol Biol. 55(3):523-33. DOI: 10.1016/0022-2836(71)90334-2.

Weinstein-Fischer, D., Elgrably-Weiss, M., Altuvia, S. (2000) Escherichia coli response to hydrogen peroxide: a role for DNA supercoiling, topoisomerase I and Fis. Mol Microbiol. 35(6):1413-20. DOI: 10.1046/j.1365-2958.2000.01805.x.

Yaginuma, H., Kawai, S., Tabata, K.V., Tomiyama, K., Kakizuka, A., Komatsuzaki, T., Noji, H., Imamura, H. (2014) Diversity in ATP concentrations in a single bacterial cell population revealed by quantitative single-cell imaging. Sci Rep. 4:6522. DOI: 10.1038/srep06522.

Yamanaka, K. (1999) Cold shock response in Escherichia coli. J Mol Microbiol Biotechnol. 1(2):193-202.

Supplementary Information for

Positive supercoiling buildup is a trigger of *E. coli*'s short-term response to cold shock

Suchintak Dash[1†], Cristina S.D. Palma[1†], Ines S.C. Baptista[1], Mohamed N.M. Bahrudeen[1], Bilena L.B. Almeida[1], Vatsala Chauhan[1], Rahul Jagadeesan[1] and Andre S. Ribeiro[1,2,*]

[1] Laboratory of Biosystem Dynamics, Faculty of Medicine and Health Technology, Tampere University, Tampere, 33520, Finland.

[2] School of Sciences and Technology and Uninova CTS, NOVA University of Lisbon, Campus de Caparica, 2829-516 Caparica, Portugal.

[†] Equal contributions.

* To whom correspondence should be addressed. E-mail: andre.sanchesribeiro@tuni.fi.

Present Address: Andre S. Ribeiro, Arvo Ylpön katu 34, 33520, Tampere University, Finland.

Supplementary Methods

I. RNA-Seq experiments and analysis

I.a) Sample preparation

We shifted the temperature for cells in the mid-exponential growth phase (OD=0.3) to 15°C (here named minute 0) and collected cells from three independent colonies, as well as after 20, 80 and, 180 minutes, respectively. The same was done with cell colonies not subjected to the temperature shift, for purposes of control. Meanwhile, when subjecting cells to the antibiotic Novobiocin (50 μg/mL), the cells were only collected 120 minutes after the treatment, to reduce cell-to-cell diversity due to e.g., different absorption times, as in [Chong et al., 2014].

After collecting the samples, 5 ml of the culture was immediately treated with a double volume (10 mL) of RNA protect bacteria reagent (Qiagen, Germany) for 5 minutes at room temperature, to prevent RNA degradation. Next, treated cells were pelleted and frozen at -80°C overnight. The next morning, total RNA was extracted using the RNeasy kit (Qiagen, Germany).

I.b) Sequencing

Extracted RNA was treated twice with DNase (Turbo DNA-free kit, Ambion, USA) and quantified using Qubit 2.0 Fluorometer RNA assay (Invitrogen, Carlsbad, CA, USA). The total RNA quality was determined using a 1% agarose gel stained with SYBR Safe (Invitrogen, Carlsbad, CA, USA), where RNA was detected using UV in a Chemidoc XRS imager (Biorad, USA). RNA integrity was measured by the Agilent 4200 TapeStation (Agilent Technologies, Palo Alto, CA, USA).

RNA library preparations, sequencing, and quality control analysis of sequenced data were conducted at GENEWIZ, Inc. (Leipzig, Germany). In detail, ribosomal RNA depletion was performed using Ribo-Zero Gold Kits (Bacteria probe) (Illumina, San Diego, CA, USA), while the RNA sequencing library was prepared using the NEBNext Ultra RNA Library Prep Kit.

The sequencing libraries were multiplexed and clustered on one lane of a Flowcell, which was loaded on an Illumina HiSeq 4000 instrument (after cold shock) or on an Illumina NovaSeq 6000 instrument (after Gyrase inhibition). In both instruments, the samples were sequenced using a single-index 2x150 Paired-End (PE) configuration. Image analysis and base calling were conducted by the HiSeq Control Software (Illumina HiSeq) and by the NovaSeq Control Software v1.7 (Illumina NovaSeq). The raw sequence data (.bcl files) was converted into “fastq” files and de-multiplexed using Illumina bsl2fastq v.2.20. One mismatch was allowed for index sequence identification.

I.c) RNA-seq data analysis pipeline

i) RNA sequencing reads were trimmed to remove possible adapter sequences and nucleotides with poor quality using Trimmomatic [Bolger et al., 2014] v.0.36. ii) Trimmed reads were mapped to the reference genome, *E. coli* MG1655 (NC_000913.3), using the STAR aligner v.2.5.2b (after cold shock) or the Bowtie2 aligner v.2.3.5.1 (after Gyrase inhibition), generating BAM files [Dobin et al., 2013; Langmead et al., 2012]. iii) Unique gene hit counts were calculated with 'featureCounts' from the Rsubread R package (v.1.34.7) [Liao et al., 2019]. Genes with less than 5 counts in more than 3 samples, and genes whose mean counts were smaller than 10 were removed from further analysis. iv) The read counts were used for downstream differential expression analysis. The DESeq2 R package (v.1.24.0) [Love, et al., 2014] was used to calculate log2 of fold changes (LFC) of RNA levels between group of samples and calculate p-values using Wald tests (function 'nbinomWaldTest'). We also calculated relative abundances of mRNA in a given condition using the Transcripts Per Million (TPM) normalization [Li et al., 2011].

Finally, the sequencing platforms for the cold shock experiment and for the Novobiocin experiment differ for logistical reasons (similarly, the aligners differ). Consequently, we avoided quantitative comparisons between genes responses (e.g., between specific numbers). Our conclusions are based on qualitative comparisons (i.e., we checked if the changes in the two perturbations are linearly correlated or not).

II. Flow cytometry and data analysis

We measured cells fluorescence using an ACEA NovoCyte Flow Cytometer (ACEA Biosciences Inc., San Diego, USA). Cells were diluted (1:10000) into 1 mL of Phosphate buffer saline (PBS) solution, vortexed for 10 seconds. In each condition, 3 biological replicates were obtained. In each replicate, we collected data from 50 000 cells. The flow rate was set to 14 µL/minute. The data was collected by the Novo Express software (ACEA Biosciences Inc.).

For detecting YFP, we used a blue laser (488 nm) for excitation and the fluorescein isothiocyanate detection channel (FITC-H) (530/30 nm filter) for emission, with a core diameter of 7.7 µM and a PMT voltage of 600. For detecting mCherry, we used the PE-Texas Red fluorescence detection channel (615/20 nm) for emission, with a PMT voltage of 584.

The lower bound for the detection threshold in FSC-H was set to 5000 to remove interference from particles. We also removed the 1% highest FITC-H values. Further, to remove abnormal cells, we then used an iterative procedure to discard outliers, which are those points whose vertical distance from the best-fit function is larger than 1 [Bar-Even et al., 2006]. The process always converged in 1 to 2 iterations. Finally, we searched for additional abnormal measurements at the single gene level in the 3 repeats, but we did not find any.

III. Microscopy and image analysis

Cells were pelleted and re-suspended in ~100 µL of the remaining media. Three microliters of cell suspension were placed on a 2% agarose gel pad made up of M9 medium and kept in between the round microscope slide and a coverslip. It took ~5 min to move cells from the incubator to the microscope and start the observation. This includes the assembly of the microscope imaging chamber containing the slides and cells. Cells were visualized by confocal microscopy with a 100x objective. Phase-contrast images were taken by an external phase-contrast system. YFP tagged strain was visualized by a 488 nm laser and a 514/30 emission filter, while DAPI stained nucleoids were visualized by a 405 nm laser and a 447/60 emission filter. Phase contrast and confocal images were taken simultaneously.

Analysis of the microscopy images was performed using the CellAging software [Häkkinen et al., 2013], for segmenting cells from phase contrast images, and the SCIP software [Martins et al., 2018] to segment nucleoids and characterize the spatial distribution of fluorescently (YFP) tagged GyrA and RpoB (both from the YFP fusion library).

IV. Correction for cellular auto-fluorescence in Flow Cytometry data

When assessing the single-cell distributions of protein expression levels measured by flow-cytometry, we corrected for the cell auto-fluorescence [Bahrudeen et al., 2019; Galbusera et al., 2020]. For this, we first measured by flow-cytometry the auto-fluorescence of control cells (i.e., absent of YFP fusions). Next, we corrected the mean fluorescence measured by flow cytometer by applying equation IV.1 [Galbusera et al., 2020]:

$$M_p = M_T - M_{cell} \tag{IV.1}$$

Here, M_p is the mean cell fluorescence due to YFP presence alone, after subtracting the cell auto-fluorescence. Meanwhile, M_T is the mean cell fluorescence measured by flow cytometry, while M_{cell} is the mean cell auto-fluorescence. Similarly, to correct the variance σ^2, we apply equation IV.2 [Galbusera et al., 2020]:

$$\sigma_p^2 = \sigma_T^2 - \sigma_{cell}^2 \tag{IV.2}$$

From equations IV.1 and IV.2 one can derive equation IV.3, to estimate the corrected squared coefficient of variation, CV_p^2, of the single-cell distribution of protein expression levels:

$$CV_p^2 = \left(\frac{\sigma_p}{M_p} \right)^2 \quad \text{(IV.3)}$$

For correcting the skewness (S) of the distribution we apply equation (IV.4) as in [Bahrudeen et al., 2019]:

$$S_p = \frac{S_T \cdot {\sigma_T}^3 - S_{cell} \cdot {\sigma_{cell}}^3}{{\sigma_p}^3} \quad \text{(IV.4)}$$

V. Derivations

Va. Squared coefficient of variation and skewness assuming a Γ distribution of single-cell protein numbers

From [Taniguchi et al., 2010], most single-cell distributions of protein numbers in E. coli are well described by a Γ distribution. If that holds, the first three moments of single-cell distributions of protein numbers should be given by:

$$M = k\theta \quad \text{(Va.1)}$$

$$\sigma^2 = k\theta^2 \quad \text{(Va.2)}$$

$$S = \frac{2}{\sqrt{k}} \quad \text{(Va.3)}$$

where M, σ^2, S, k and θ are the mean, variance, skewness, shape parameter, and scale parameter of a Γ distribution, respectively. From equations V.1 and V.2:

$$CV^2 = \frac{\sigma^2}{M^2} = \frac{\theta}{M} \quad \text{(Va.4)}$$

This relationship was empirically validated in [Bar-Even et al., 2006][Taniguchi et al., 2010]. Finally, from equation V.1 and V.3:

$$S = \frac{2}{\sqrt{k}} = \frac{2}{\sqrt{\frac{M}{\theta}}} = \frac{2}{\sqrt{M}} \cdot \sqrt{\theta} \quad \text{(Va.5)}$$

V.b Derivation of Ω assuming the 1-step model

The steady state solutions for mean RNA and protein numbers, assuming the one step Model (reactions 1.1, 2, 3, and 4 in Figure 6), are given by, respectively [Taniguchi et al., 2010]:

$$M_{RNA} = \frac{k_{1.1}}{\lambda_1} \tag{Vb.1}$$

$$M_P = \frac{M_{RNA} \cdot k_2}{\lambda_2} = \frac{k_{1.1} \cdot k_2}{\lambda_1 \cdot \lambda_2} \tag{Vb.2}$$

Meanwhile, the variance of the single-cell protein numbers is given by [Taniguchi et al., 2010]:

$$\sigma_{\mathrm{P}}^2 = \frac{k_{1.1} \cdot k_2}{\lambda_1 \cdot \lambda_2} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right) \tag{Vb.3}$$

From (Vb.2) and (Vb.3):

$$CV_{\mathrm{P}}^2 = \frac{y_1 \cdot y_2}{k_{1.1} \cdot k_2} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right) = \frac{1}{\mathrm{M_P}} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right) \tag{Vb.4}$$

Here, we define the constant **Ω** as:

$$\Omega = \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right) \tag{Vb.6}$$

This result is in line with past results in [Bar-Even et al., 2006]. Also, given (Va.4), we expect θ = Ω. As such, we refer to this constant as Ω.

V.c Derivation of the squared coefficient of variation assuming the 1-step model

Assuming Model 1.1 in Figure 6, from (Vb.2), (Vb.4), and (V.5):

$$CV_p^{\ 2} = \frac{1}{\frac{k_1 \cdot k_2}{\lambda_1 \cdot \lambda_2}} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right) \tag{Vc.1}$$

V.d Derivation of the Squared Coefficient of Variation assuming the ON-OFF model

Assuming the ON-OFF model (Model 1.3 in Figure 6), the mean RNA and protein numbers in single cells at steady state are, respectively:

$$M_{RNA} = \frac{k_+}{k_+ + k_-} \cdot \frac{k_{1.3}}{\lambda_1} \tag{Vd.1}$$

$$M_P = \frac{M_{RNA} \cdot k_2}{\lambda_2} = \frac{k_+}{k_+ + k_-} \cdot \frac{k_{1.3} \cdot k_2}{\lambda_1 \cdot \lambda_2} \tag{Vd.2}$$

Meanwhile, the variance is [Taniguchi et al., 2010]:

$$\sigma_P^2 = \frac{k_+}{k_+ + k_-} \cdot \frac{k_{1.3} \cdot k_2}{\lambda_1 \cdot \lambda_2} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\left(1 + \left(1 - \frac{k_+}{k_+ + k_-}\right)\frac{k_{1.3}(k_+ + k_- + \lambda_1 + \lambda_2)}{(k_+ + k_- + \lambda_1)(k_+ + k_- + \lambda_2)}\right)\right) \quad \text{(Vd.3)}$$

From equations Vd.3 and Vd.3:

$$CV_P^2 = \frac{1}{M_P} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\left(1 + \left(1 - \frac{k_+}{k_+ + k_-}\right)\frac{k_{1.3}(k_+ + k_- + \lambda_1 + \lambda_2)}{(k_+ + k_- + \lambda_1)(k_+ + k_- + \lambda_2)}\right)\right) \quad \text{(Vd.4)}$$

From Vd.4 and Va.4, **Ω** should equal:

$$\Omega = \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\left(1 + \left(1 - \frac{k_+}{k_+ + k_-}\right)\frac{k_{1.3}(k_+ + k_- + \lambda_1 + \lambda_2)}{(k_+ + k_- + \lambda_1)(k_+ + k_- + \lambda_2)}\right)\right) \quad \text{(Vd.5)}$$

V.e Changes in Ω due to cold shock

To estimate the change in **Ω** due to shifting to CS we considered that, prior to cold shock, the CS responsive genes due to sensitivity to PSB are not significantly affected by locking due to PSB (since they are relatively highly expressing in optimal growth conditions). As such, their dynamics in optimal conditions should be well modeled by Model 1.1 in Fig 6A in the main manuscript. On the other hand, when subject to CS, we expect that frequent locking due to PSB will be the main responsible for their negative response. Thus, the appropriate model during CS should be Model 1.3 in Fig 6A in the main manuscript.

Thus, given the results in supplementary sections V.b and V.d, the change in **Ω** when shifting to cold shock should equal:

$$\frac{\Omega_{15°C}}{\Omega_{30°C}} = \frac{\left(1 + \frac{k_2(15°C)}{\lambda_1(15°C) + \lambda_2(15°C)}\left(1 + \left(1 - \frac{k_+}{k_+ + k_-}\right)\frac{k_1((k_+ + k_-) + \lambda_1(15°C) + \lambda_2(15°C))}{((k_+ + k_-) + \lambda_1(15°C))((k_+ + k_-) + \lambda_2(15°C))}\right)\right)}{\left(1 + \frac{k_2(30°C)}{\lambda_1(30°C) + \lambda_2(30°C)}\right)} \quad \text{(Ve.1)}$$

Next, consider that, from Table S7, $\frac{k_2}{\lambda_1 + \lambda_2}$ = 14.9, which is much larger than 1. Thus, for simplicity, we replace $\left(\frac{k_2}{\lambda_1 + \lambda_2} + 1\right)$ by $\frac{k_2}{\lambda_1 + \lambda_2}$. Further, we consider that, in *E. coli*, RNA degradation rates are much higher than protein degradation rates, i.e., $\lambda_1 >> \lambda_2$ (Table S7). Given this, we also replace $(\lambda_1 + \lambda_2)$ by λ_1. Consequently:

$$\frac{\Omega_{15°C}}{\Omega_{30°C}} = \frac{k_2(15°C)\times\lambda_1(30°C)}{k_2(30°C)\times\lambda_1(15°C)}\cdot\left(1+\left(1-\frac{k_+}{k_+ + k_-}\right)\frac{k_1((k_+ + k_-)+\lambda_1(15°C)+\lambda_2(15°C))}{((k_+ + k_-)+\lambda_1(15°C))((k_+ + k_-)+\lambda_2(15°C))}\right) \quad \text{(Ve.4)}$$

This, by derivation, can be simplified to:

$$\Leftrightarrow \frac{\Omega_{15°C}}{\Omega_{30°C}} = \frac{k_2(15°C)\times\lambda_1(30°C)}{k_2(30°C)\times\lambda_1(15°C)}\cdot\left(1+\left(\frac{k_1.k_-}{(k_+ + k_-)^2}\right)\right) \quad \text{(Ve.4)}$$

We note that we expect that the term $\left(\frac{k_1.k_-}{(k_+ + k_-)^2}\right)$ is the one containing the most temperature sensitive rate constants, given that model 1 in Figure 6 includes the other rates constants (k_2 and λ_1) and could not explain the dynamics following cold shock for the reasons listed in section 3.6 in the main manuscript.

VI: Effects of cell division on Ω

At temperatures above cold shock (30°C, 25°C and 20°C), the cells exhibited significant doubling times (Figure 2A). Meanwhile, at CS they did not divide (Figure 2A).

Cell division can increase single-cell variability in protein numbers, provided asymmetries in the partitioning of RNA and protein numbers between sister cells (for a review, see [Baptista et al., 2020]). This difference between the conditions, could affect the comparison of the contribution of noise in gene expression in optimal and CS conditions.

We thus estimated the effects of cell division on Ω at temperatures above cold shock (30°C, 25°C and 20°C) if the division rate at 30°C, 20°C and 25°C was null.

Since we are considering temperatures above cold shock (30°C, 25°C and 20°C), we assume the 1-step model (model 1.1 in Figure 6). Next, we assume that $\lambda_1 >> \lambda_2$, since, in general RNA degrades much faster than proteins [Taniguchi et al., 2010]. Given this, from Equation V.b.6 in supplementary section V.b:

$$\frac{\Omega_{30°C(\lambda_d=7\times10^{-5})}}{\Omega_{30°C(\lambda_d=0)}} = \frac{\left(1+\frac{k_2}{(\lambda_1+\lambda_d)+(\lambda_2+\lambda_d)}\right)}{\left(1+\frac{k_2}{\lambda_1+\lambda_2}\right)} = \frac{\left(\frac{k_2}{(\lambda_1+\lambda_d)+(\lambda_d)}\right)}{\left(\frac{k_2}{\lambda_1}\right)} = \frac{\lambda_1}{\lambda_1+2\lambda_d} \quad \text{(VI.1)}$$

Next, from [Bernstein et al., 2002], we assume that, on average λ_1=0.004s^{-1}. Also, we measured the mean cell division rate at 30°C-20°C to be 241 min^{-1} (Figure 2A). This inverse should correspond to the protein and RNA dilution rates due to cell division and it equals: λ_d=7×10^{-5} s^{-1}. As such:

$$\frac{\Omega_{30°C(\lambda_d=7\times10^{-5})}}{\Omega_{30°C(\lambda_d=0)}} = 0.97 \quad \text{(VI.1)}$$

Given this, if cells were not dividing in optimal conditions, Ω would be 3% higher. This is within the 90% confidence interval of the green line in Figure 4B. Thus, we do not include cell division in the models.

VII: Promoter sequence logos

Promoter sequence logos were created using WebLogo [Crooks et al., 2004]. From positions -25 to -1 of each promoter, it counts in how many promoters is each nucleotide present. Then, it piles up the nucleotides (A, C, T, G), sorted from the rarest in the bottom to the most frequent in the top.

The height of each nucleotide letter in the plot, in each position, equals the frequency multiplied by the total information at that position. That total information is quantified by the difference between the maximum uncertainty at any position ($\log_2(n)$, where n = 4 is the number of possible codons) and the uncertainty given the frequencies found, also quantified by Shannon's information: $\log_2 n - \sum_{i=1}^{4} f_i \times \log_2(f_i)$. Given this, we expect that DNA with more conserved positions will have more 'bits' (Schneider and Stephens, 1990).

VIII: RBS and start codon sequences of CSR genes

RNA translation rates are controlled by the rate at which ribosomes are recruited to the ribosome binding site (RBS) region of the RNA, along with the rate at which they then initiate translation. The recruitment rate differs with the RBS sequence [Ringquist S et al., 1992] and the genome wide consensus sequence of RBSs is "5'-AGGAGG-3'", being named the Shine-Dalgarno (SD) sequence [Saito et al., 2020].

Meanwhile, the rate of translation initiation is influenced by the start codon upstream the RBS. In *E. coli*, 83% of the start codons have the sequence AUG (3542/4284), 14% (612) the sequence GUG, 3% (103) the sequence UUG [Blattner, 1977] and a couple the sequence AUU [Sacerdot et al., 1982; Missiakas et al.,1993].

We obtained the mean and standard deviation of the distributions of the p-distances (Supplementary Section X) of the RNAs coded by CSR genes to the SD (Table S9) and to each of the 4 start codons

sequences (Table S10) and studied if they differ significantly from the mean and standard deviation of p-distances of the genome wide distribution. Finally, since the distance (in number of nucleotides) between the SD sequence and the start codon can affect translation initiation rates [Saito et al., 2020], we also compared them as above.

From Supplementary Figure S8, not only the consensus levels of the two cohorts are the same (bit values of 0.5 between positions -10 and -15 and 1 in the region -1 to -3), but the distances between them are, in both cohorts, 5 nucleotides.

To support, we also compared the sequence logos of the 25 nucleotides upstream of the start codons [Wenfa, 2019] (Supplementary Figure S9) of CSR genes and the genome wide distribution. Again, we find little differences between the logos in Fig. S9A and S9B.

IX: Estimation of the average transcription rate of CSR genes during optimal growth conditions

From [Taniguchi et al., 2010], the mean RNA numbers (as measured by FISH) of a CSR gene during optimal growth is 0.35 per cell. Given the 1-step model (model 1.1 in Figure 6), which is applied during optimal growth, the mean number of RNAs per cell in steady state is given by:

$$k_1 = M_{RNA} \times \lambda_1 \qquad \text{(VIII.1)}$$

Assuming λ_1=0.004 s^{-1} [Bersntein et al., 2002; Selinger et al., 2003], k_1 is estimated to be 1.4×10^{-3} s^{-1}.

X: P-distances

We calculated the p-distance between a promoter sequence and the consensus sequence (sequence composed of the most common nucleotide for each position of the sequence). The p-distance is the fraction of nucleotides of the promoter sequence that differ from the consensus sequence. Thus, it ranges from [0,1], where 0 corresponds to identical sequences and 1 to sequences whose nucleotides differ in every position. For genes with more than one promoter, we obtained the average of the p-distance of each promoter.

XI: Correlation between RNA and protein numbers

RNA and protein numbers are expected to be positively correlated in bacteria, since transcription and translation are mechanically bound [Dahan, O. et al., 2011; Yanofsky, C. et al., 1981; Proshkin, S. et al., 2010] and because most gene expression regulation occurs during transcription initiation [Alberts, B. et al., 2008].

To assess if this holds true during cold-shock, we searched for correlations between LFC's (Supplementary section I), as measured by RNA-seq at 20 and 80 min after the temperature shift, and

the corresponding LFC's in protein numbers, measured by flow-cytometry (Supplementary section II) at 120 min and 180 min after the temperature shift. The lag of 100 minutes between RNA and protein measurements should suffice for changes in numbers of the former to propagate to the latter. The list of genes tested is shown in Table S1. Results in Figure S11B show that changes in RNA and protein numbers are correlated during CS.

XII: Quantifying ATP using spectrophotometry

To quantify the fluorescence from GFP tagged ATP inside cells, we use the method in [Yaginuma et al., 2014]. First, the total cell fluorescence at excitation wavelength λ is given by:

$$F_\lambda = F_\lambda^m + f_\lambda^{bg} \cdot C + f_\lambda^p . C \tag{XI.1}$$

F stands for total fluorescence and C for the number of cells. Meanwhile, F^m stands for media fluorescence, f^{bg} for single-cell fluorescence background and f^p for single-cell protein fluorescence (in our case, λ equals 400 nm in one case and 494 nm in the other).

Meanwhile, the total cell fluorescence (without ATP sensors) is:

$$F_\lambda^c = F_\lambda^m + f_\lambda^{bg} \cdot C \tag{XI.3}$$

The subtraction of (XI.1) from (XI.3) corrects for media and cell background autofluorescence:

$$F_\lambda - F_\lambda^c = f_\lambda^p . C \tag{XI.5}$$

Given this, the fluorescence from ATP-GFP from a cell is estimated by:

$$\frac{f_{\lambda=494}^p(t)}{f_{\lambda=400}^p(t)} = \frac{F_{\lambda=494}(t) - F_{\lambda=494}^c(t)}{F_{\lambda=400}(t) - F_{\lambda=400}^c(t)} \tag{XI.7}$$

XIII: Analysis of the AT and CG content of the promoters

From Regulon DB, we obtained the lists of all 8791 promoters and of all 3700 transcription units (TUs) [Santos-Zavaleta et al., 2019]. We then filtered the promoter list to contain only the 2355 promoters associated to TUs, and subsequently discarded 93 promoters with unknown sequence. The resulting list was comprised of 2262 promoters, each with a sequence spanning from 60 nucleotides upstream the transcription start site (TSS) to 20 nucleotides downstream (i.e., from positions -60 to +20, with the TSS assumed to be in the position +1).

Next, for each promoter, we extracted the sequences from positions -60 to -35, positions -35 to -10, and positions -10 to +1 (the -35 to -10 being the sequence that most influence the RNAP binding

[deHaseth et al., 1998], while the others were used for comparison). Similarly, we also extracted the same sequences of 443 CSR genes with known promoter sequences. Finally, for each set of sequences, we calculated the fractions of A, C, G and T. Finally, the AT (or GC) content of each promoter was calculated by summing the fractions of A and T (or C and G, respectively).

XIV. Estimation of a lower bound of skewness

Supplementary Figure S10, informs on a lower bound for noise ($CV^2 \sim 0.38$) since noise and mean are no longer correlated below that value. Based on this lower bound, we estimated a lower bound of skewness. From the equations in Table S6, one can write:

$$S = 2 \cdot CV \qquad \text{(XIII.1)}$$

We estimate a lower bound for the skewness to equal ~1,23, in agreement with the data (Figure 5), in that below that value, the predicted and empirical skewness do not correlate.

XV. Null models and statistical tests

In general, to create null-models of how variable X affects variable Y, we performed random sampling without replacement of both *X* and *Y* datapoints. The number of samplings and the sampling size (number of samples in each sampling) are set to the maximum array size allowed by MATLAB (~45980 × 45980, 15.8GB). The number of samplings (*K*) is set to 100 and the size is set according to *Max_size*/*K* where *Max_size* = 45980/2. Next, for both *X* and *Y*, we combine the sampled datapoints in a vector (*sample_X*, *sample_Y*) and calculate the correlation between *sample_X* and *sample_Y* by linear regression fitting using Ordinary Least Squares.

We obtained a p-value of the fitted regression lines from t-tests with the null hypothesis that the line is horizontal.

We further evaluated the null hypothesis that slopes and intercepts of the best fitting lines of empirical and null-model data are equal. We performed the ANCOVA test [McDonald, 2009], which evaluates the significance of an F-test under the null hypothesis that both slopes and intercepts are equal. To correct for over-representation of datapoints in these tests, we corrected the degrees of freedom to be (*size_XY* – *C*), where *size_XY* is the number of datapoints and *C* is the number of parameters. For the linear regression fitting, *C* equals to 2 (intercept and slope of best fitting line). For the ANCOVA test, *C* equals to 4 (intercept and slope of one best fitting line and the difference of these between both best-fitting lines).

XVI. Gene ontology (GO) analysis

To study the GO representation [Ashburner et al., 2000; Gene Ontology Consortium, 2021] of CSR genes, we performed an overrepresentation test using the PANTHER Classification System [Mi et al., 2019]. This test finds statistically significant overrepresentations using Fisher's exact test, which rejects the null hypothesis that there are no associations between the genes' cohort and the corresponding GO of the biological process for p-values < 0.05. This p-value is corrected for the False Discovery Rate (FDR) using the Benjamini-Hochberg procedure [Benjamini and Hochberg, 1995].

XVII. Analysis of gene fitness

From 4133 reference bacterial DNAs with listed genes [Xavier et al., 2021], we used the Rentrez package [Winter et al., 2017] to count in how many of these genomes one finds each gene of MG1655 (GCF_000005845.2_ASM584v2). We use these numbers (divided by the total number of genomes) as a measure of the evolutionary fitness of each gene, which in bacteria can propagate in the biosphere by cell division or by horizontal gene transfer. In detail, we calculated the mean and CV^2 of gene fitness of CSR genes as well as for all genes in the genome, along with their standard errors using bootstrapping (104 resampling with replacement). For mixed genes cohort, we calculated the same along with their standard errors using bootstrapping (104 resampling with replacement) but here each sample consists of genes from different groups of sampled at different sizes.

XVIII. Short-term responses to CS cannot be explained by Transcription Factor interactions

We studied if the dynamics of CSR genes during CS is influenced by their direct input TFs. According to Regulon DB [Santos-Zavaleta A et al., 2019], the 4698 genes of *E. coli* have a total of 4590 TF interactions between them. Our RNA-seq data informs on the LFC_{CS} of most genes (4328), which have 4435 TF interactions between them. These genes include all 381 CSR genes identified. Also included are 733 TF interactions between them (Supplementary File X3).

Given the time length of protein assembly and maturation [Balleza et al., 2018; Hebisch et. al., 2013; Maurizi, 1992], we expect a time delay from the moment RNAs change until the moment the corresponding TFs change. Once this happens, we expect a short time delay until their output genes' RNA numbers to change in response [Bernstein et al., 2002]. Thereby, we searched for correlations between RNA numbers of genes coding for a TF in a given time moment, and the RNA numbers in a future moment of genes whose promoters are known outputs of that TF (information of the TF-promoter interactions was obtained from Regulon DB).

From the time-lapse RNA-seq data (Supplementary Section I), neither the mid- nor the long-term responses of CSR genes (at 80 and 180 min after CS, respectively) are correlated with the short-term changes (20 min after CS) in the numbers of the RNAs coding for their input TFs.

It is worth noting that, although CSR genes do not show influence from their TFs, there is a visible propagation of information in the TFN during CS. Namely, on average, non-CSR genes exhibited correlated dynamics with their input TFs (Supplementary Figures S17). As such, the lack of correlation during CS is a feature of CSR genes, rather than genome wide. We observed the same at optimal temperatures (Supplementary Figure S17). Potentially, these interactions will be active under conditions not studied here.

Finally, we searched for (but failed to find) correlations between the short-term responses to CS of global transcription regulators (Supplementary Table S5 and Supplementary Figure S1).

XIX. Short-term responses of CSR genes cannot be explained by the promoters' AT richness

For CSR genes to exhibit strong repression upon CS, they necessarily were strongly expressing in optimal conditions. This is confirmed in Supplementary Figures S20A (and agrees with data in [Taniguchi et al., 2010]).

Meanwhile, AT-rich promoters are more strongly expressed than GC rich promoters [Liu et al., 2004] in optimal conditions. This could imply that CSR promoters are AT-rich. We thus confronted the AT richness (Supplementary section XIII) of promoter sequences with their short-term response strengths to CS (Supplementary Figures S20B). While we find a genome-wide correlation, that correlation does not exist in the CSR gene cohort (Supplementary Figures S20B inset). We conclude that AT richness is not involved in short-term responsiveness to CS.

Supplementary Figures

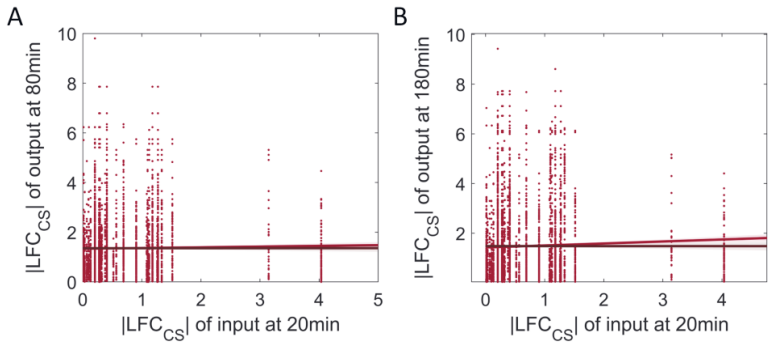

Figure S1: Signal propagation after cold shock. Correlation plots of the $|LFC_{CS}|$ between input and output genes with a transcription factor pathlength of 1. **(A)** Correlation during CS between output genes at 80min vs input genes at 20min of all pairs of genes input-output (blue dots) and of pairs where the input is a global regulator (dark red dots). **(B)** Correlation during CS between output genes at 180min vs input genes at 20 min of all pairs of genes input-output (blue dots) and of pairs where the input is a global regulator (dark red dots). To the red circles, we fitted by OLS the best fit line. Null models were generated as described in Supplementary Section XV. For each time point, we did an ANOVA test to test for the null hypothesis that the red and black line are not statistically distinguishable. P-values are presented in Supplementary Table S4. For each plot fitted line we performed a likelihood ratio test between the zero-order polynomial and higher order polynomials. P-values < 0.05 reject the null hypothesis that the best fit line is a horizontal line. Shadow areas are 68% confidence bounds.

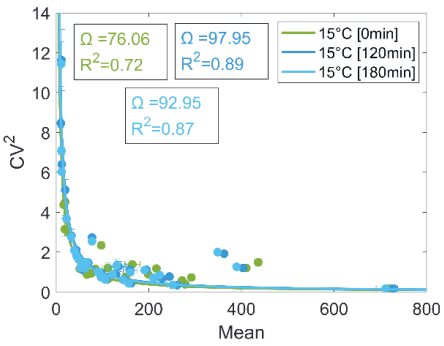

Figure S2: Squared Coefficient of variation (CV^2) as a function of Mean Protein Fluorescence level. **(A)** CV^2 versus mean protein fluorescence level measured by the FITC-H channel of the flow cytometer at 15 °C, 0 min, 120 min and 180 min. To the data we fit the best-fit function $CV^2 = \Omega / P$ [Bar-Even et al., 2006; Taniguchi et al., 2010], where Ω is a constant (values for each curve are

shown in the insets) and P is the mean protein numbers. We performed a 2-sample t-test to check for the null hypothesis that there is no difference between the Ω values at 120min and 180min. The test does not reject the null hypothesis with a p-value of 0.43.

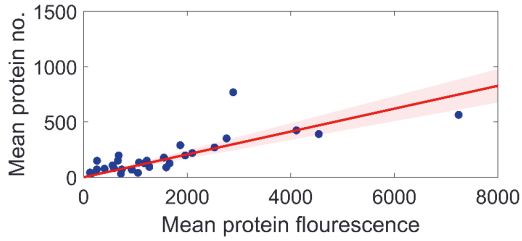

Figure S3: Our measured mean protein fluorescences (Supplementary section II) are plotted against the corresponding protein numbers reported in [Taniguchi et al., 2010] in the same growth conditions. The best fitting line of the type y=mx has a statistical significant slope m = 0.10 and its R^2=0.56. We performed a F-test on the regression model, which tests for the hypothesis that the 0 order polynomial fits significantly better than a 1st order polynomial. The test rejected the null hypothesis (p-value > 10^{-6}).

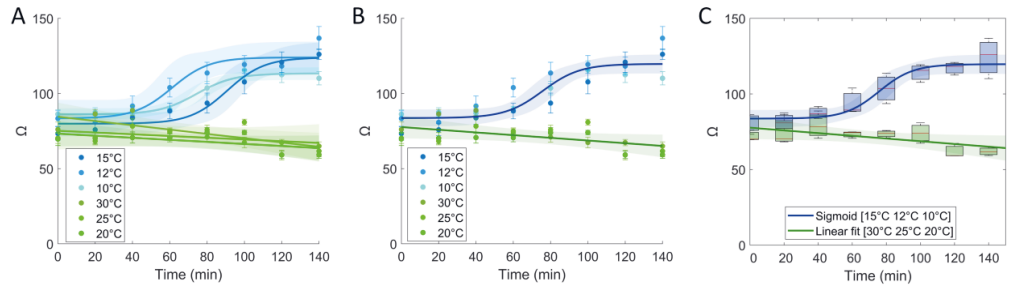

Figure S4: (A) For temperature we fit the best fitting function. Data >= 20°C is best fit by a 1st order polynomial while data < 20°C is best fit by a sigmoid function (function given by $\frac{L}{1+e^{-0.1\cdot(x-x_0)}}$, where L is the curve's maximum value and x_0 is the x value of the sigmoid midpoint). **(B)** Average best fitting functions for data below 20°C and above or equal to 20°C. The dark blue and green lines correspond to the best fit curves (that maximize R^2), for the 15°C-10°C dataset and for the 30°C - 20°C dataset, respectively. **(C)** Box plot of Ω as a function of time for temperature set (Control and CS). The red line in the box is the median and the top and bottom of the box are one STD above and below the median, respectively. For control and cold-shock temperatures, we fit the best fitting function.

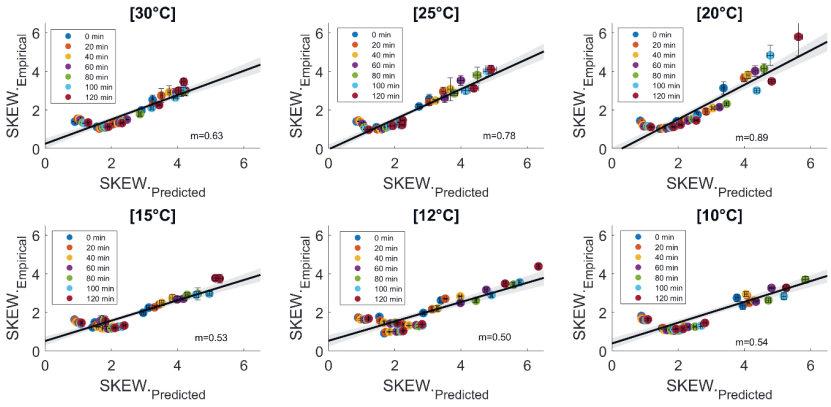

Figure S5: Correlation between empirical and predicted skewness for 30°C, 25°C, 20°C, 15°C, 12°C, 10°C. Predicted skewness is estimated from the Ω estimated from the relationship CV^2 vs M (Supplementary section XIV) from empirical Flow cytometry data while the empirical skewness is the skewness of the empirical distributions obtained by Flow cytometry. For all fittings R^2 is > 0.8.

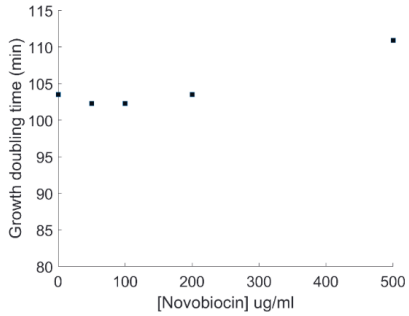

Figure S6: *E. coli* K-12 MG1655 cell growth doubling time versus various concentrations of Novobiocin in M9 medium supplemented with 0.4% glucose, amino acids, and vitamins at 30 ºC. The doubling time was measured by the initial OD_{600nm} value, the final $OD_{600\ nm}$ value and the time interval in between.

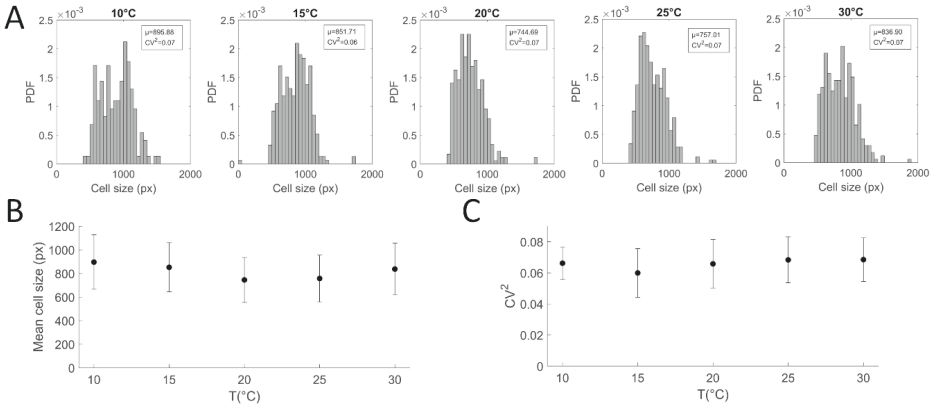

Figure S7: Microscopy measurements of cell size (px) of *E. coli* cells expressing the manY gene endogenously tagged with YFP coding sequence at different temperatures (10 °C, 15 °C, 20 °C, 25 °C and 30 °C). **(A)** Probability density function (PDF) of the distribution of cell size, 180 min after the temperature shift. **(B)** Mean cell size (px) as a function of temperature. Vertical error bars correspond to the standard deviation. **(C)** Squared coefficient of variation as a function of temperature. Vertical error bars were calculated by bootstrapping (standard deviation of 500 resamples of 20 cells each).

Figure S8: Sequence logos of the RBS sequence from the SD to the start codon for **(A)** 179 genes and **(B)** and the 31 CS genes of those 179. Data from RegulonBD. The last nucleotide of the start codon is placed in the position -1. Also, the nucleotides positions decrease in the upstream direction.

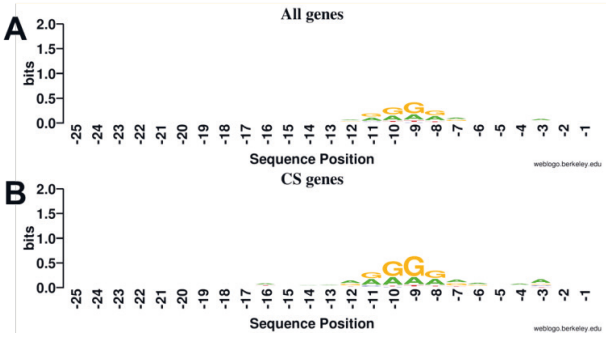

Figure S9: Sequence logo of the RBS sequence ranging from the SD sequence to the start codon for all 4357 genes **(A)** and 377 CS genes **(B)** in the other list. The nucleotide upstream of the start codon of the gene is assumed to have the position -1 and the position of a nucleotide decreases as it is located more upstream.

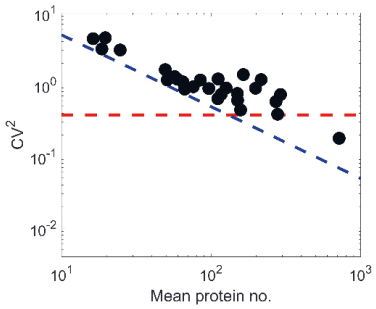

Figure S10: Squared coefficient of variation versus mean protein numbers. For mean protein number < 150 the protein expression noise is inversely proportional to the mean with a lower noise limit (blue dashed line), corresponding to intrinsic noise [Taniguchi et al., 2010]. For mean protein number >150 noise becomes independent of the mean with a lower bound of ~0.38 (red dashed line).

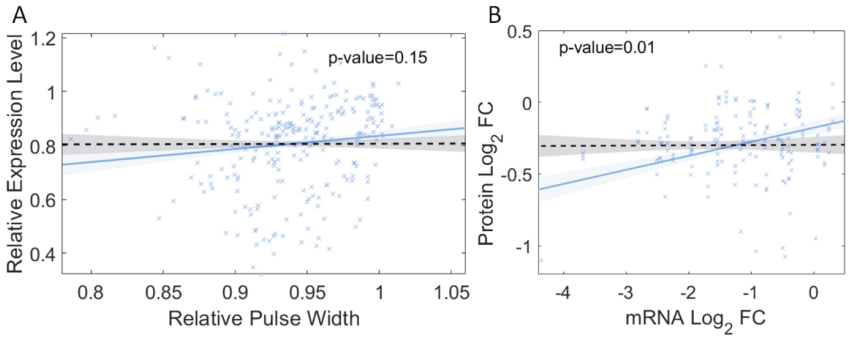

Figure S11: (A) Correlation plot between the pulse width and protein expression levels, measured by flow cytometer (Supplementary Section II). Values are relative to 0 min. Black dashed line is the null

model (Supplementary Section XV), the two lines are not statistical distinguishable (p-value = 0.15). **(B)** Correlation plot of the $|LFC_{CS}|$ of mRNAs (measured by RNAseq at 15 °C, at time 20 min and 80 min after the temperature shift) and protein expression levels (measured by flow cytometry at 15 °C, 120 min and 180 min after (with a gap of 100 min [Startceva et al., 2019]) the temperature shift). We fit the first-order polynomial to the data points by ordinary least-squares and performed an F-test on the regression model, which tests for the hypothesis that the 0 order polynomial fits significantly better than a 1st order polynomial. The test rejected the null hypothesis (p-value = 0.0002). Black dashed line is the null model, the two lines are statistical distinguishable (p-value = 0.01) (Supplementary Section XV). This is consistent with the cell division times (~150 min in M9 media). As such, we expect phenotypic consequences from the cold-shock response at the RNA level to be propagated into the proteins produced in the same period [Newman, et al., 2006; Vogel, et al., 2012; Liu, et al., 2016].

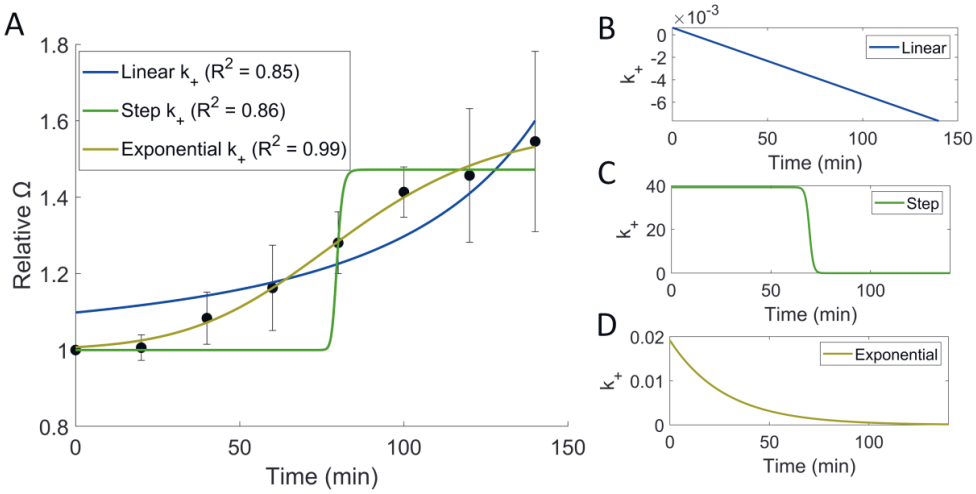

Figure S12: (A) Model fitting given a specific function for k_+ changes (linear, step and exponential function). Black dots are the empirical data for CS temperatures and error bars are the standard error. **(B-D)** Changes in k_+ over time assuming a linear, step and exponential function, respectively. In **A**, **B**, **C**, **D**, all lines are best fitting curves for respective model predictions. Table S12 shows the R^2 for each model fit and prediction. Due to low R^2 when fitting the empirical data the fitting of quadratic type function is not shown.

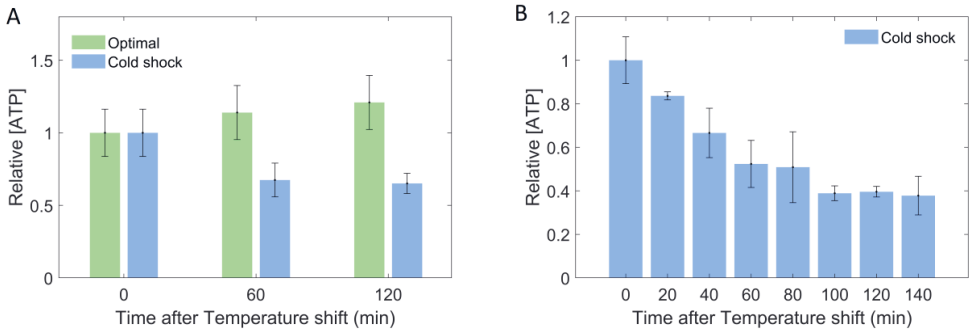

Figure S13: Relative ATP concentration measured (Methods Section 2.4) using the QUEEN-2m sensor [Yaginuma et al., 2014], with 3 replicates per condition. Detailed description of correction for autofluorescence and cell division is in (Supplementary Section XII).

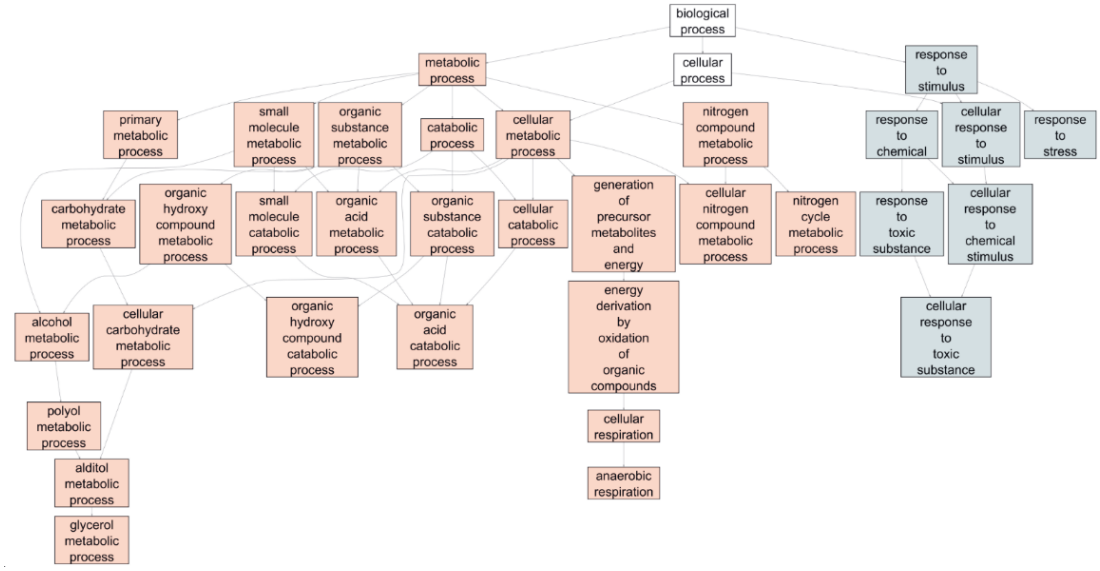

Figure S14: Graph of overrepresented gene ontology terms and their ancestors for CSR genes (Supplementary Section XVI). The more general biological processes are connected to specific ontology terms by an arrow pointing to the latter. The ontologies that have the ancestor "metabolic process" and the ones with the ancestor "response to stimulus" are marked separately.

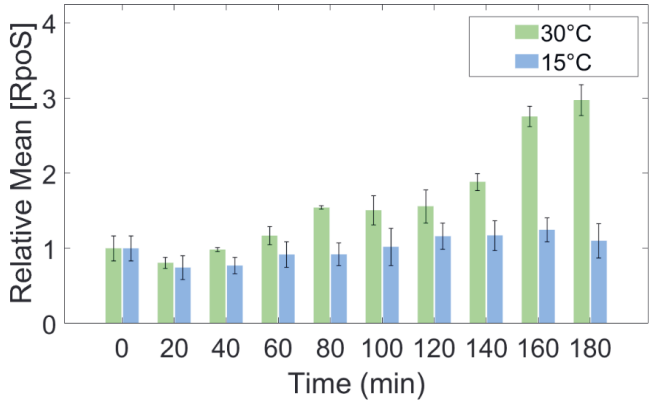

Figure S15: Mean concentration of RpoS (σ^{38}), relative to 0 min, measured every 20 min in both optimal (30 °C) and cold temperature (15 °C), by flow cytometer.

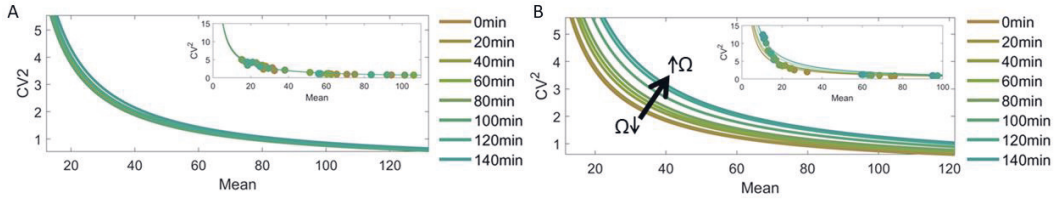

Figure S16: CV^2 versus mean protein numbers measured by flow cytometer during optimal and CS temperatures, every 20 min for 120 min, respectively. To each time moment we fit the function $CV^2 = \Omega / M$ [Bar-Even et al., 2006; Taniguchi et al., 2010]. Fittings are shown for each time moment. Each time moment data comes triplicates from 6 genes (aldA, feoA, many, ndk, pepN, tktB). **(A-B inset)**. Data points along with best-fit curves are shown.

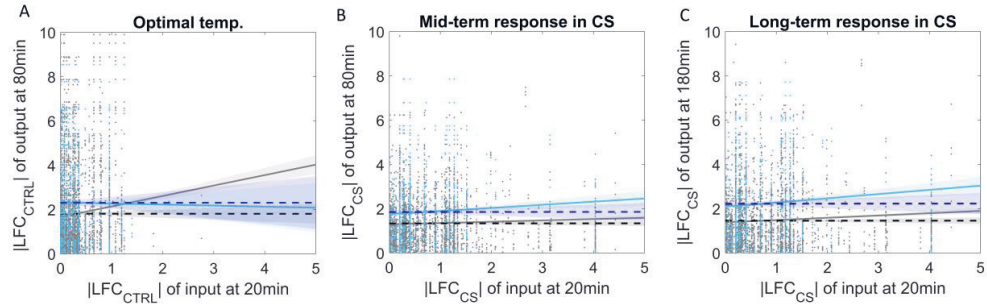

Figure S17. (A) Scatter plot between $|LFC_{CS}|$ of TF output genes at 80 min and corresponding TF input genes at 20 min in optimal conditions. **(B) and (C)** Scatter plots between $|LFC_{CS}|$ of output gene at 80 min and 180 min, respectively, and the corresponding TF input genes at 20 min during CS. Grey circles are all pairs of genes (4435 pairs) (Supplementary section XVIII). Blue circles (733) are of pairs of input-output TF genes for which the output is a CSR gene. We fitted by ordinary least-squares a best-fit line (grey and blue lines, respectively). Dashed lines are the null models (Supplementary Section XV). For each best fit line (grey and blue line) and correspondent null model (black and dark blue dashed line, respectively), we performed an ANCOVA test for the null hypothesis that the two lines are not statistical distinguishable (Supplementary Section XV). P-values < 0.05 reject the null hypothesis. P-values for each test are in Table S2 and Table S3, respectively. Shadow regions are 68% confidence bounds.

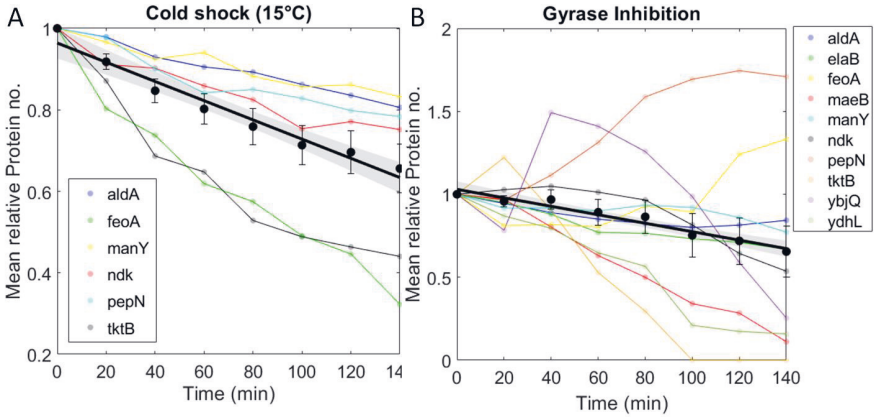

Figure S18: (A) Change in relative mean protein numbers of 6 CSR genes over time following CS. **(B)** Change in relative mean protein numbers of 10 CSR genes over time following Gyrase inhibition by Novobiocin. Related with Section 3.7 of the main manuscript.

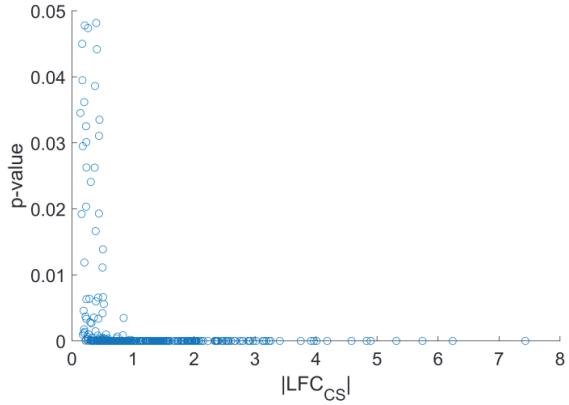

Figure S19: Scatter plot between $|LFC_{CS}|$ and p-values of the CSR genes from the RNA-seq data.

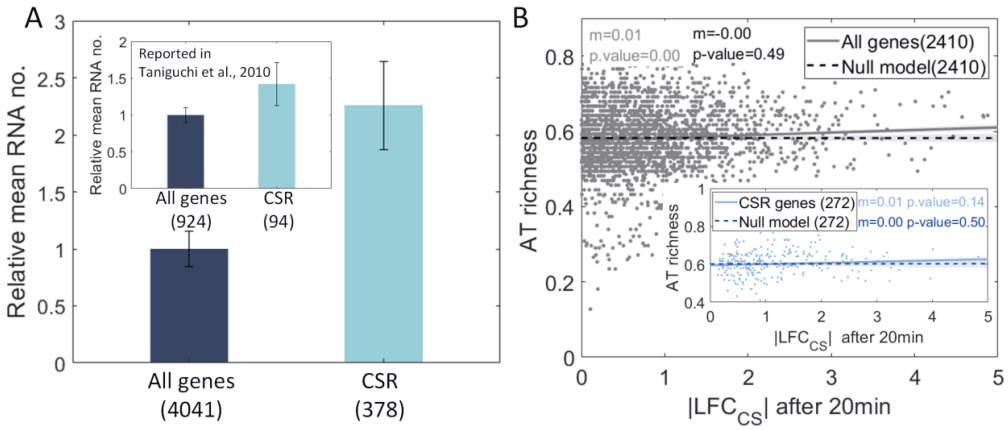

Figure S20: (A) Relative RNA expression levels of all genes in the RNAseq (4041) and of the CSR genes alone (378) at optimal temperature. Inset data from [Taniguchi et al., 2010]. **(B)** Correlation between the genes' AT richness and their response to CS ($|LFC_{CS}|$) 20 min after CS. The inset shows the same plot when considering only CSR genes.

Supplementary Tables

Table S1: List of YFP-fusion strains [Taniguchi et al., 2010] used in this study.

S. No.	Strain name	Genotype	Source
1	MG1655	λ-, rph-1	Yale CGSC (CGSC # 6300)
2	MGmCherry	Same as MG1655, with rpoS::mCherry, chromosomally integrated, replacing rpoS	Gift of James Locke [Patange 2018]
3	SX1047	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gyrA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12602)
4	SX1056	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, gyrB792-YFP(::cat)	Yale CGSC (CGSC # 12611)
5	SX1020	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], topA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12575)
6	SX1440	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], topB792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12995)
7	SX1051	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, rpoB791-YFP(::cat)	Yale CGSC (CGSC # 12606)
8	SX1397	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], clpA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12952)
9	SX1519	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], pepN794-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13074)
10	SX1812	F-, yaeH791-YFP(::cat), Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13367)
11	SX1986	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], ydfG791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13541)
12	SX1882	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yqjD791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13437)
13	SX1394	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], ndk-791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12949)
14	SX1989	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yeeX793-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13544)
15	SX1505	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yciI793-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13060)
16	SX1695	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], elaB792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13250)
17	SX1950	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857	Yale CGSC

		Δ(cro-bioA)], putP792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	(CGSC # 13505)
18	SX1550	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, glpD792-YFP(::cat), rph-1	Yale CGSC (CGSC # 13105)
19	SX1674	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gcvT792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13229)
20	SX1919	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gcvP791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13474)
21	SX1917	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gabD791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13472)
22	SX1901	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], aldA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13456)
23	SX1488	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, tnaA791-YFP(::cat)	Yale CGSC (CGSC # 13043)
24	SX1284	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], gatZ794-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12839)
25	SX1763	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], manY793-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13318)
26	SX1416	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], pgk-791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12971)
27	SX1087	F-, Δ(argF-lac)169, bolA791-YFP(::cat), gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 12642)
28	SX1526	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], katE791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13081)
29	SX1771	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], nuoE794-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13326)
30	SX1954	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], tktB792-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13509)
31	SX1859	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, yjbQ792-YFP(::cat)	Yale CGSC (CGSC # 13414)
32	SX1781	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, feoA791-YFP(::cat), rph-1	Yale CGSC (CGSC # 13336)
33	SX1718	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], wrbA791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13273)
34	SX1975	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], yccJ791-YFP(::cat), IN(rrnD-rrnE)1, rph-1	Yale CGSC (CGSC # 13530)
35	SX1085	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857	Yale CGSC

		Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, tpiA791-YFP(::cat)	(CGSC # 12640)
36	SX1349	F-, Δ(argF-lac)169, gal-490, Δ(modF-ybhJ)803, λ[cI857 Δ(cro-bioA)], IN(rrnD-rrnE)1, rph-1, pfkA791-YFP(::cat)	Yale CGSC (CGSC # 12904)
37	BW25993	pRSET B_QUEEN-2m_AMPR	Gift from Hiromi Imamura [Yaginuma et al., 2014]

Table S2: For each condition we did an ANCOVA test to test for the null hypothesis that the grey and black line are not statistically distinguishable. P-values are presented below. For conditions that did not reject the null hypothesis (p-values > 0.05).

	CS		Optimal
	80 min	180 min	80 min
20 min	0.028	0.0077	2.1×10^{-04}

Table S3: For each condition we did an ANCOVA test to test for the null hypothesis that the light blue and dark blue line are not statistically distinguishable. P-values are presented below. For conditions that did not reject the null hypothesis (p-values > 0.05).

	CS		Optimal
	80 min	180 min	80 min
20 min	0.15	0.096	0.52

Table S4 For each condition we did an ANOVA test (Supplementary Section XV) to test for the null hypothesis that the red and dark red lines are not statistically distinguishable. P-values are presented below. For conditions that did not reject the null hypothesis (p-values > 0.05).

	CS	
	80 min	180 min
20 min	0.3625	0.1817

Table S5: List of global regulators (GR) known to regulate 30 or more genes.

S.No	GR	No. Genes regulated	Genes that code the GR
1	CRP	574	crp
2	FNR	310	fnr
3	IHF	258	ihfA; ihfB
4	Fis	239	fis

5	H-NS	194	hns
6	ArcA	184	arcA
7	NarL	138	narL
8	Fur	131	fur
9	Lrp	110	lrp
10	NsrR	84	nsrR
11	Cra	82	cra
12	FlhDC	82	flhD; flhC
13	CpxR	71	cpxR
14	NarP	66	narP
15	PhoB	66	phoB
16	LexA	61	lexA
17	PhoP	59	phoP
18	NtrC	56	glnG
19	MarA	46	marA
20	ModE	46	modE
21	SoxS	42	soxS
22	PdhR	41	pdhR
23	NagC	39	nagC
24	ArgR	38	argR
25	OxyR	35	oxyR
26	SlyA	34	slyA
27	IscR	32	iscR
28	CysB	31	cysB
29	PurR	31	purR
30	FhlA	30	fhlA

Table S6: Equations of Squared Coefficient of Variation (CV^2) and skewness (S) as a function of **Ω** and the mean protein numbers (M). Also present is the derivation of CV^2 and S as a function of the model parameters. Step-by-step derivation in shown in Supplementary Section V

Variable	Function of Ω and *Mean*	1-step model	2-step model	ON-OFF model
CV^2	$CV^2 = \Omega \cdot \frac{1}{M}$	$CV^2 = \frac{1}{\frac{k_1 \cdot k_2}{\lambda_1 \cdot \lambda_2}} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right)$	$CV^2 = \frac{1}{\frac{\left(\frac{1}{k1} + \frac{1}{ki}\right)^{-1} \cdot k_2}{\lambda_1 \cdot \lambda_2}} \cdot \left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right)$	$CV^2 = \frac{1}{\frac{k_+}{k_+ + k_-} \cdot \frac{k_1 \cdot k_2}{y_1 \cdot y_2}} \cdot \left(1 + \frac{k_2}{y_1 + y_2}\left(1 + (1-P)\frac{k_1(k_0 + y_1 + y_2)}{(k_0 + y_1)(k_0 + y_2)}\right)\right)$
S	$S = \frac{2}{\sqrt{M}} \cdot \sqrt{\Omega}$	$S = \frac{2 \cdot \sqrt{\left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right)}}{\sqrt{\frac{k_1 \cdot k_2}{\lambda_1 \cdot \lambda_2}}}$	$S = \frac{2 \cdot \sqrt{\left(1 + \frac{k_2}{\lambda_1 + \lambda_2}\right)}}{\sqrt{\frac{\left(\frac{1}{k1} + \frac{1}{ki}\right)^{-1} \cdot k_2}{\lambda_1 \cdot \lambda_2}}}$	$S = \frac{2 \cdot \sqrt{\left(1 + \frac{k_2}{y_1 + y_2}\left(1 + (1-P)\frac{k_1(k_0 + y_1 + y_2)}{(k_0 + y_1)(k_0 + y_2)}\right)\right)}}{\sqrt{\frac{k_+}{k_+ + k_-} \cdot \frac{k_1 \cdot k_2}{y_1 \cdot y_2}}}$

Table S7: Reference parameter values assuming the one-step and the ON-OFF models. Given that Lac is considered to be a strong promoter, combining these parameter values should result in relatively high RNA and protein numbers.

Rate Constant	Description	Values (s^{-1})	References
k_+	Promoter unlocking	7×10^{-4}	For the native Lac promoter [Chong et al., 2014; Palma et al., 2020]
k_-	Promoter locking	0.0012	For the native Lac promoter [Stracy et al., 2019; Palma et al., 2020]
k_1	Transcription rate	8.3×10^{-4} to 0.02	Average expressing native genes [Prajapat et al., 2018]
k_2	Translation rate	0.04 to 0.06	Average expressing native genes [Bremer et al., 1996]
λ_1	RNA degradation	0.004	Average RNA degradation rates [Bernstein et al., 2002]
λ_2	Protein decay (degradation and dilution in cell division)	2.93×10^{-5}	Average protein degradation rates [Koch and Levy 1955; Taniguchi et al., 2010]

Table S8: Set of parameter values use to investigate the noise levels of each model: 1-step model (Model 1 in Figure 8), 2-step model (Model in figure S8) and ON-OFF model (Model 2 in Figure 8)

Rate Constant	One rate-limiting step model	Two rate-limiting step model	ON-OFF model
k_1 (s^{-1})	0.002	0.004	0.0057
k_+(s^{-1})			0.0018
k_-(s^{-1})			0.0033
k_i(s^{-1})		0.004	
k_2(s^{-1})	0.23	0.23	0.23
λ_1(s^{-1})	0.004	0.004	0.004
λ_2(s^{-1})	2.9×10^{-5}	2.9×10^{-5}	2.9×10^{-5}

Table S9: p-distances to the consensus SD sequence (AGGAGG) [Saito et al., 2020]. The sequences were extracted from the available ones in RegulonDB. Shown are the mean and standard deviation of the p-distances of all genes (179 genes), of the 31 genes of those 179 that were found to CSR, and of 1000 cohorts of 31 genes assembled by random selection from the 179 genes.

	Consensus SD sequence ('AGGAGG') (Mean ± STD)
Genome wide p-distance (179 genes) ± STD	0.57 ± 0.22
p-distance of CSR genes (31 genes) ± STD	0.48 ± 0.20
p-distance of randomly selected genes (31 genes)	0.54 ± 0.20

Table S10: p-distances of the start codon sequences of CSR genes to the most frequent start codon sequences (ATG; GTG; TTG; ATT; CTG) [Blattner et al., 1977; Sacerdot et al., 1982; Missiakas et al.,1993]. Sequences extracted from *Regulon DB.* Shown are the mean and standard deviation of the p-distances of all genes (179 genes), of the 31 genes of those 179 that were found to CSR, and of 1000 cohorts of 31 genes assembled by random selection from the 179 genes.

	ATG	GTG	TTG	ATT	CTG
Average start codon sequence (N=179) ± STD *(Regulon DB)*	0.49 ± 0.22	0.64 ± 0.34	0.61 ± 0.29	0.47 ± 0.18	0.63 ± 0.31
Average start codon sequence of CS genes (N=31) ± STD	0.59 ± 0.47	0.71 ± 0.34	0.66 ± 0.29	0.52 ± 0.17	0.67 ± 0.30
Average start codon sequence of random cohort (31) ± STD	0.53 ± 0.49	0.63 ± 0.35	0.59 ± 0.28	0.52 ± 0.17	0.63 ± 0.31

Table S11: Over-represented biological process according to the Gene Ontology (GO) Overrepresentation Test (Supplementary Section XVI) with Fisher's exact test with FDR correction. For each GO biological process it is shown the number of genes related to it, in the *E. coli* genome (out of the 4390 genes); the number of genes related in the CSR cohort (out of the 376 recognized genes by GO); the number of genes expected to be present in a cohort of the size of the CSR cohort; and finally, the fold-enrichment.

Biological Processes	**GO biological process complete**	**No. Genes in E. coli genome**	**No. Genes in CS cohort**	**Expected No. Genes in CS cohort**	**Fold Enrichment**
Metabolic process (GO: 0008152)	Glycerol metabolic process (GO:0006071)	20	12	1,71	7,01
	Nitrogen cycle metabolic process (GO:0071941)	25	13	2,14	6,07
	Organic hydroxy compound catabolic process (GO:1901616)	52	16	4,45	3,59
	Generation of precursor metabolites and energy (GO:0006091)	210	60	17,99	3,34
	Organic acid catabolic process (GO:0016054)	191	35	16,36	2,14
	Carbohydrate metabolic process (GO:0005975)	389	61	33,32	1,83
	Organic acid metabolic process (GO:0006082)	542	77	46,42	1,66
Response to stimulus	Response to stress (GO:0006950)	556	74	47,62	1,55

(GO: 0050896)	Cellular response to toxic substance (GO:0097237)	28	10	2,4	4,17

Table S12: Best fitting parameter values for model functions of k_+ as a function of temperature.

Function type	Equation	Estimated parameters	Average R^2
Linear	$b_2 \cdot x + b_1$	$b_1 = 6\times10^{-4}$; $b_2 = -5.9\times10^{-5}$; $k_- = 0.013$ s^{-1}	0.85
Exponential	$e^{(-b\cdot x)} \cdot a$	b = 0.036; a = 0.019; $k_- = 0.002$ s^{-1}	0.99
Quadratic	$b_3 \cdot x^2 + b_2 \cdot x + b_1$	$b_1 = 0.847$; $b_2 = 0.894$; $b_3 = -0.024$; $k_- = 1.183$ s^{-1}	< 0
Step	$\left(1 - \frac{1}{1+e^{-(x-L)}}\right) \times a$	a = 39.25; L = 69.32; $k_- = 0.003$ s^{-1}	0.86

REFERENCES

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2008) Molecular Biology of the Cell. Garland Science, New York, NY, ISBN 978-0-8153-4105-5

Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nat Genet. 25: 25–29. DOI:10.1038/75556

Bahrudeen, MNM, Chauhan, V, Palma, CSD, Oliveira, SMD, Kandavalli, VK, Ribeiro, AS (2019) Estimating RNA numbers in single cells by RNA fluorescent tagging and flow cytometry. J Microbiol Methods 166:105745. DOI: 10.1016/j.mimet.2019.105745

Balleza, E, Kim, JM, Cluzel, P. (2018) Systematic characterization of maturation time of fluorescent proteins in living cells. Nat Methods 15, 47-51. DOI: 10.1038/nmeth.4509

Baptista, ISC, Ribeiro AS (2020) Stochastic models coupling gene expression and partitioning in cell division in Escherichia coli. *BioSystems* 193–194, 104154. DOI: 10.1016/j.biosystems.2020.104154.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., Barkai, N. (2006) Noise in protein expression scales with natural protein abundance. Nat Genet. 38(6):636-43. DOI: 10.1038/ng1807.

Blattner FR, Plunkett G 3rd, Bloch CA, Perna NT, Burland V, Riley M, Collado-Vides J, Glasner JD, Rode CK, Mayhew GF, Gregor J, Davis NW, Kirkpatrick HA, Goeden MA, Rose DJ, Mau B, Shao Y. (1997) The complete genome sequence of Escherichia coli K-12. Science 277(5331):1453-62. DOI: 10.1126/science.277.5331.1453.

Benjamini, Y., Hochberg, Y. (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. Journal of the Royal Statistical Society: Series B (Methodological), 57: 289-300. https://DOI.org/10.1111/j.2517-6161.1995.tb02031.x

Bernstein, JA, Khodursky, A., Lin, P.-H, Lin-Chao, S,Cohen, SN. (2002) Global analysis of mRNA decayand abundance in Escherichia coli at single-gene resolution using two-color fluorescent DNA microarrays. Proc. Natl Acad. Sci. USA 99, 9697–9702. DOI:10.1073/pnas.112318199.

Bolger AM, Lohse M, Usadel B. (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30(15):2114-20. DOI: 10.1093/bioinformatics/btu170.

Bremer, H., and Dennis, PP. (1987). Modulation of chemical composition and other parameters of the cell by growth rate. In Escherichia coli and Salmonella typhimurium, F.C. Neidhardt et al., eds. Washington, D.C.: American Society for Microbiology, 1527–1542

Chong, S., Chen, C., Ge, H., Xie, X.S. (2014) Mechanism of transcriptional bursting in bacteria. Cell 158(2):314-326. DOI: 10.1016/j.cell.2014.05.038.

Crooks, G. et al. (2004) NCBI GenBank FTP Site\nWebLogo: a sequence logo generator, Genome research, 14, 1188–1190. DOI: 10.1101/gr.849004.1.

Dahan, O., Gingold, H., Pilpel, Y. (2011) Regulatory mechanisms and networks couple the different phases of gene expression. Trends Genet. 27, 316-322. DOI: 10.1016/j.tig.2011.05.008

deHaseth, P. L., Zupancic, M. L., Record, M. T. (1998). RNA polymerase-promoter interactions: the comings and goings of RNA polymerase. Journal of Bacteriology, 180(12), 3019–3025. DOI: 10.1128/JB.180.12.3019-3025.1998

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. (2013) STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1):15-21. DOI: 10.1093/bioinformatics/bts635.

Galbusera, L., Bellement-Theroue, G., Urchueguia, A., Julou, T., van Nimwegen, E. (2020). Using fluorescence flow cytometry data for single-cell gene expression analysis in bacteria. PLoS ONE, 15(10). DOI: 10.1371/journal.pone.0240233

Gene Ontology Consortium (2021) The Gene Ontology resource: enriching a GOld mine. Nucleic Acids Res. 49(D1):D325-D334. DOI: 10.1093/nar/gkaa1113.

Häkkinen A, Muthukrishnan AB, Mora A, Fonseca JM, Ribeiro AS. (2013) CellAging: a tool to study segregation and partitioning in division in cell lineages of Escherichia coli. Bioinformatics 29(13):1708-9. DOI: 10.1093/bioinformatics/btt194.

Hebisch,E., Knebel,J., Landsberg,J., Frey,E. and Leisner,M. (2013) High variation of fluorescence protein maturation times in closely related Escherichia coli strains. PLoS One, 8, e75991. DOI: 10.1371/journal.pone.0075991

Koch L, Levy HR (1955) Protein turnover in growing cultures of Escherichia coli. J Biol Chem, 12;1955 217:947

Langmead B, Salzberg SL. (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods. 9(4):357-9. DOI: 10.1038/nmeth.1923.

Li B, Dewey CN. (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 12:323. DOI: 10.1186/1471-2105-12-323.

Liao Y, Smyth GK, Shi W. (2019) The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 47(8):e47. DOI: 10.1093/nar/gkz114.

Liu, L.F., Wang, J.C. (1987) Supercoiling of the DNA template during transcription. Proc Natl Acad Sci of the U.S.A. 84(20):7024-7. DOI: 10.1073/pnas.84.20.7024.

Love MI, Huber W, Anders S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15(12):550. DOI: 10.1186/s13059-014-0550-8.

Martins L, Neeli-Venkata R, Oliveira SMD, Häkkinen A, Ribeiro AS, Fonseca JM. (2018) SCIP: a single-cell image processor toolbox. Bioinformatics 34(24):4318-4320. DOI: 10.1093/bioinformatics/bty505.

Maurizi, MR. (1992) Proteases and protein degradation in Escherichia coli. Experientia, 48, 178-201. DOI: 10.1007/BF01923511

McDonald, JH. (2009) Analysis of covariance in the Handbook of biological statistics, 2nd ed. (Sparkey House, Baltimore, 2009), 232–238.

Mi H, Muruganujan A, Ebert D, Huang X, Thomas PD. (2019) PANTHER version 14: more genomes, a new PANTHER GO-slim and improvements in enrichment analysis tools. Nucleic Acids Res. 47(D1):D419-D426. DOI: 10.1093/nar/gky1038.

Missiakas D, Georgopoulos C, Raina S. The Escherichia coli heat shock gene htpY: mutational analysis, cloning, sequencing, and transcriptional regulation (1993) J Bacteriol. 175(9):2613-24. DOI: 10.1128/jb.175.9.2613-2624.1993.

Newman, J. R. S., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L., and Weissman, J. S. (2006) Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature 441:7095, 441(7095), 840–846. DOI:10.1038/nature04785

Ng, W. (2019): Database of ribosome binding site of all genes in Escherichia coli K-12 MG1655. figshare. Dataset. DOI:10.6084/m9.figshare.10282685.v1

Palma CSD, Kandavalli V, Bahrudeen MNM, Minoia M, Chauhan V, Dash S, et al. Dissecting the 681 in vivo dynamics of transcription locking due to positive supercoiling buildup. Biochimica et 682 Biophysica Acta (BBA) - Gene Regulatory Mechanisms. 2020;1863: 194515. doi: 683 10.1016/j.bbagrm.2020.194515

Prajapat MK, Ribeiro AS. (2018) Added value of autoregulation and multi-step kinetics of transcription initiation. R Soc Open Sci. 28; 5(11):181170. DOI: 10.1098/rsos.181170.

Proshkin, S., Rahmouni, A.R., Mironov, A., Nudler, E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. Science 328, 504-508. DOI: 10.1126/science.1184939

Ringquist S, Shinedling S, Barrick D, Green L, Binkley J, Stormo GD, Gold L. (1992) Translation initiation in Escherichia coli: sequences within the ribosome-binding site. Mol Microbiol. 6(9):1219-29. DOI: 10.1111/j.1365-2958.1992.tb01561.x.

Sacerdot C, Fayat G, Dessen P, Springer M, Plumbridge JA, Grunberg-Manago M, Blanquet S. (1982) Sequence of a 1.26-kb DNA fragment containing the structural gene for E.coli initiation factor IF3: presence of an AUU initiator codon. EMBO J.1(3):311-5.

Saito, K., Green, R., & Buskirk, A. R. (2020) Translational initiation in E. coli occurs at the correct sites genome-wide in the absence of mRNA-rRNA base-pairing. ELife, 9. DOI: 10.7554/ELIFE.55002

Santos-Zavaleta A, et al (2019) RegulonDB v 10.5: tackling challenges to unify classic and high throughput knowledge of gene regulation in E. coli K-12. Nucleic Acids Res. 47(D1):D212-D220. DOI: 10.1093/nar/gky1077.

Schneider, TD and Stephens, RM (1990) Sequence logos: A new way to display consensus sequences, Nucleic acids research 18(20), 6097–6100. DOI: 10.1093/nar/18.20.6097.

Selinger DW, Saxena RM, Cheung KJ, Church GM, Rosenow C. (2003) Global RNA half-life analysis in Escherichia coli reveals positional patterns of transcript degradation. Genome Res. 13, 216– 223. DOI: 10.1101/gr.912603

Stracy M, Wollman AJM, Kaja E, Gapinski J, Lee J-E, Leek VA, McKie SJ, Mitchenall LA, Maxwell A, Sherratt DJ, Leake MC, Zawadzki P. (2019) Single-molecule imaging of DNA gyrase activity in living Escherichia coli, Nucleic Acids Res. 47(1) 210–220, DOI: 10.1093/nar/gky1143.

Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, Emili A, Xie XS. (2010) Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science. Jul 30;329(5991):533-8. DOI: 10.1126/science.1188308. Erratum in: Science. 2011 Oct 28;334(6055):453.

Vogel, C., Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. Nat Rev Genet. 13(4):227-32. DOI: 10.1038/nrg3185.

Winter, D. (2017) rentrez: An R package for the NCBI eUtils API. R J. 9: 520. DOI:10.32614/rj-2017-058.

Xavier JC, Gerhards RE, Wimmer JLE, Brueckner J, Tria FDK, Martin WF (2021) The metabolic network of the last bacterial common ancestor. Commun Biol. 4:413. DOI:10.1038/s42003-021-01918-4

Yanofsky, C. (1981) Attenuation in the control of expression of bacterial operons. Nature 289, 751-758. DOI: 10.1038/289751a0