

SAMUEL MATOS DA COSTA

Functionality of Genetic Circuits at Fluctuating Temperatures

SAMUEL MATOS DA COSTA

Functionality of Genetic
Circuits at Fluctuating
Temperatures

ACADEMIC DISSERTATION

To be presented, with the permission of
the Faculty Council of Biomedical Sciences and Engineering
of Tampere University of Technology,
for public discussion in the auditorium TB109
of the Tietotalo building, Korkeakoulunkatu 1, Tampere,
on 04.01.2019, at 12 o'clock.

ACADEMIC DISSERTATION

Tampere University, Faculty of Medicine and Health Technology
Finland

<i>Doctoral candidate</i>	Samuel Matos da Costa (née Samuel Matos Dias Oliveira) Tampere University Finland	
<i>Responsible Supervisor and Custos</i>	Andre S. Ribeiro, Professor Tampere University Finland	
<i>Pre-examiners</i>	Mary J. Dunlop, Assist. Professor Boston University United States of America	Tal Danino, Assist. Professor Columbia University United States of America
<i>Opponent</i>	Namiko Mitarai, Assoc.Professor University of Copenhagen Denmark	

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

Copyright ©2019 author

Cover design: Roihu Inc.

ISBN 978-952-03-0969-5 (print)
ISBN 978-952-03-0970-1 (pdf)
ISSN 2489-9860 (print)
ISSN 2490-0028 (pdf)
<http://urn.fi/URN:ISBN:978-952-03-0970-1>

PunaMusta Oy
Tampere 2019

Abstract

Temperature affects virtually all biophysical processes inside cells. Relevantly, it shifts over time, periodically and stochastically, sometimes in minutes. For these reasons, organisms have evolved complex mechanisms to cope with both temperature shifts as well as with temperature cycles.

Endothermic organisms have evolved cellular processes to generate and dissipate heat so as to regulate their body temperature. Failure in these processes can lead to death. Meanwhile, ectothermic organisms, such as *Escherichia coli*, do not have this ability. Instead, they evolved mechanisms to minimize potential harms. These are activated if temperature deviates from optimal ranges and are controlled by genes and genetic circuits. The robustness and sensitivity of genetic circuits to temperature shifts and what determines them remain largely uncharacterized.

In this thesis, we have studied the effects of temperature shifts on the dynamics of genes and small genetic circuits in *Escherichia coli* using *in vivo* single-cell, single-RNA detection techniques, microscopy and microfluidics devices, and image and signal analysis tools. Relevantly, two synthetic constructs were built specifically for these studies: i) an RNA sequence with multiple binding sequences for fluorescent probes was integrated into *E. coli*'s chromosome and, ii) the Elowitz-Leibler 'repressilator' circuit was inserted into a single-copy F-plasmid.

First, we focused on the effects of temperature shifts on the dynamics of a synthetic genetic clock, the above mentioned Elowitz-Leibler low-copy repressilator (LCR). By studying its fluorescence over time, we observed a loss of functionality (fraction of cells exhibiting oscillations) and robustness (fraction of expected oscillations that do occur) for higher-than-optimal temperatures. We hypothesized that this is due to a loss of functionality of the CI repressor, which is one of the proteins composing the structure of this circuit. To test this, we made use of a genetic switch (CI-Cro, where CI is also a component of the circuit's structure), which we subjected to the same temperature shifts. We found a behavioral change at the same temperature. Namely, as temperature is increased, at a given value, the kinetics of RNA production of the P_{RM} promoter changes from sub-Poissonian to super-Poissonian, consistent with the emergence of tangible repression by the opposing protein, Cro. These behavioral changes in the two circuits, at the same temperature ranges, are best explained by the loss of functionality of CI, which is the only component present in both circuits.

Second, we investigated how coupling within a cell between the copies of the Elowitz-Leibler synthetic genetic clock affects its efficiency in time tracking. For that, we compared the functionality and robustness of the LCR and the newly engineered single-copy repressilator (SCR) to temperature shifts and to external perturbations by introducing Isopropyl β -D-1-thiogalactopyranoside (IPTG), an inducer of the P_{L-LacO} promoter. By analyzing the mean and variability of the periods of oscillation of these two constructs we found that, surprisingly, contrary to when at

optimal conditions, the SCR is more robust in regimes of low-temperatures and low-concentrations of IPTG.

Third, we focused on how intrinsic factors to single gene expression dynamics are affected by temperature shifts. Namely, we studied the effects of temperature fluctuations on the kinetics of transcription initiation of the promoters $P_{\text{lac-ara-1}}$ and P_{tetA} . For that, time intervals between consecutive transcription events were extracted using time-lapse, single-RNA microscopy measurements in live cells. To identify which rate-limiting steps of transcription initiation were responsible for the observed differences in varying temperature conditions, detailed stochastic models were fitted to the empirical data using statistical methods. The results suggest that different genes have different rate-limiting steps patterns, i.e. in number and duration of rate-limiting steps, which may allow them to adapt their sensitivity and RNA production kinetics (mean rate and variability) to environmental fluctuations.

Next, we studied the differences in the dynamics of RNA production of a gene when integrated into the chromosome and when integrated into a single-copy plasmid, for various temperature conditions. The results showed that when chromosome-integrated, long-lasting super-coiled states affect the temperature-dependence of the kinetics of transcription initiation. This was not observed in the plasmid-borne promoter, explaining why it is less noisy at low temperatures. We expect this to occur in other genes and to depend on the location in the chromosome. If true, and assuming that the location of genes in the chromosome is subject to natural selection, it implies that cells may make use of supercoiling as a means to regulate noise in gene expression.

Finally, we focused on the effects of temperature shifts on extrinsic factors to gene expression. In particular, we focused on the temperature-dependence of the cytoplasm viscosity. For that, the spatial distribution and mobility of large synthetic and natural protein aggregates in cells were assessed. Previous studies reported that their spatial distribution, at optimal temperatures, is heavily affected by the phenomenon of nucleoid exclusion. However, when cells were subject to critically low temperatures or osmotic stress, we found that both the natural and synthetic protein aggregates became more homogeneously distributed (consistent with weaker effects from nucleoid exclusion). We showed also that this is a result of increased cytoplasm viscosity, which reduced the effects of nucleoid exclusion. Interestingly, in agreement, we observed a reduction in the degree of asymmetries in aggregate numbers between sister cells across generations, which may affect the ability of rejuvenation of cell lineages.

Overall, the results above contribute to a better understanding of the complex consequences of temperature shifts on cellular processes. By affecting intrinsic and extrinsic factors, these shifts can alter significantly the dynamics of gene networks in bacteria. This knowledge is particularly important to understand the high degree of plasticity of natural genetic circuits, which will assist the engineering of robust synthetic genetic circuits.

Preface

This Ph.D. thesis is submitted for the Doctoral Program of Biomedical Sciences and Engineering and is based on research projects carried out in the Laboratory of Biosystem Dynamics (LBD) of the BioMediTech Institute at Tampere University of Technology, under the supervision of Prof. Andre S. Ribeiro, Ph.D.

First, I would like to thank Andre for, seven years ago, giving me the opportunity to work with him and his top-level research group in Finland. Thanks for sharing and teaching me during these years everything I know about science and management in research. Surely, this knowledge will make a great difference in determining the success of my future scientific career. Finally, I would like to thank Andre for motivating and guiding me towards my MSc. and Ph.D. degrees. They would not be possible in their present form without his precious and dedicated time and effort.

Next, I am grateful for having contributed in many scientific works with former and current members of the group during this time at the LBD. In particular, I would like to thank Jarno Mäkelä, Antti Häkkinen, and Jason Lloyd-Price for the time and immense support spent with me during my Ph.D. studies. Then, I would like to thank Ramakanth Neeli, Nadia Goncalves, and Jerome Chandraseelan for their contributions and remarks in the thesis' publications. Finally, I would like to thank all current members of LBD, in particular, Vinodh Kandavalli, Mohamed Bahrudeen and Sofia Startceva, for their contributions, and valuable advice in other publications.

Finally, I would like to express my deep sense of gratitude to my dad, mom, and wife, for always encouraging me to follow my dreams, and for their immense and constant love during these 7 years abroad. Special thanks to my lovely wife for always strengthening my thoughts and believes and for helping me to achieve my goals. Thanks for always being next to me during these 16 years together, especially now, during this stimulating journey in Finland.

Tampere, November 9, 2018

Samuel Matos da Costa

Contents

Abstract	i
Preface	iii
List of Figures	vii
List of Symbols and Abbreviations	ix
List of Publications	xi
1 INTRODUCTION	1
1.1 Background and Motivation	1
1.2 Thesis Objectives	4
1.3 Thesis Outline	6
2 BIOLOGICAL BACKGROUND	7
2.1 Gene Expression in <i>Escherichia coli</i>	7
2.1.1 Transcription Initiation and its Rate-Limiting Steps	10
2.1.2 Gene Regulatory Networks and Genetic Motifs	18
2.1.3 Coupling of Genetic Circuits	26
2.1.4 Extrinsic Factors to Gene Expression Affecting the Dynamics of Genetic Circuits	27
2.2 Methods in Synthetic Biology	30
2.2.1 Engineering Principles of Synthetic Biology and Advancements in Synthetic Biology Tools	30
2.2.2 DNA Assembly and Genome Engineering Techniques	32
2.3 Methods for RNA Detection and Quantification from Single-Cells Images	35
3 THEORETICAL BACKGROUND	39

3.1	Stochastic Models of Gene Expression	39
3.1.1	Stochastic Simulation Algorithm (SSA) – Monte Carlo Simulation of the Chemical Master Equation (CME)	39
3.1.2	Delayed SSA and Stochastic Model of Transcription	42
3.1.3	Delayed Stochastic Model of Small Genetic Networks.....	45
4	RESULTS: GENETIC CONSTRUCTS, MEASUREMENTS, AND ANALYSIS	49
4.1	Genetic Constructs	49
4.1.1	Single-Copy Repressilator (SCR)	49
4.1.2	Synthetic Fluorescent Probe Integrated in <i>E. coli</i> 's Chromosome	50
4.2	Microfluidics Platform for Perturbation Assays	54
4.3	Live Cell Imaging and Data Analysis	56
4.3.1	Dynamics of Genetic Oscillators.....	56
4.3.2	Dynamics of Transcription Initiation.....	58
4.3.3	Diffusion Coefficient of Fluorescent Aggregates and Cytoplasmic Viscosity	59
4.3.4	Nucleoid and Protein Aggregates Spatial-Distribution.....	61
5	CONCLUSIONS AND DISCUSSION.....	63
	REFERENCES	71

List of Figures

Figure 2.1: The central dogma of molecular biology.....	8
Figure 2.2: Representation of the dynamic coupling between transcription and translation.	9
Figure 2.3: Proposed mechanisms of transcription initiation and promoter escape.....	12
Figure 2.4: Representation of the native <i>lac</i> operon of <i>E. coli</i>	17
Figure 2.5: Representation of the native <i>araBAD</i> operon of <i>E. coli</i>	18
Figure 2.6: A synthetic genetic AND logic gate and the dynamics of the output signal	20
Figure 2.7: The network diagrams of the ‘repressilator’	22
Figure 2.8: The network schemes of the ‘temperature-compensation oscillator’	23
Figure 2.9: The design of transcriptional logic gates.	25
Figure 2.10: The effect of host processes on the dynamics of synthetic circuits.....	29
Figure 2.11: Schematics of Gibson Assembly one-step isothermal reaction.	34
Figure 2.12: (A) Schematic representation of the genetic components	38
Figure 4.1: Linear representation of the plasmid-borne mRNA tagging system	51
Figure 4.2: Representation of the construction of the knock-in (KI) cassette	52
Figure 4.3: Sequence of BW25993 Δ lacZ::mCherry-MS2-33bs.	53
Figure 4.4: The schematic illustration of the thermal-microfluidic system.....	55
Figure 4.5: (A) Estimating period length of oscillatory signals in individual cells.	57
Figure 4.6: (A) In the top panel (a), example fluorescence microscopy images of <i>E. coli</i>	59

List of Symbols and Abbreviations

ATP	Adenosine triphosphate
BAC	Bacterial artificial chromosome
bp	Base-pairs
bs	Binding sites
cAMP	Cyclic adenosine monophosphate
CAP	Catabolite activator protein
CC	Closed complex
CME	Chemical master equation
CRP	cAMP receptor protein
CV	Coefficient of variation
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
EC	Elongation complex
FISH	Fluorescence in situ hybridization
FP	Fluorescent probe
GRN	Gene regulatory network
GFP	Green fluorescent protein
HILO	Highly inclined and laminated optical sheet
IPTG	Isopropyl- β -D-1-Thiogalactopyranoside
KDE	Kernel density estimation
KI	Knock-in
LCR	Low-copy repressilator
mRNA	messenger RNA
MSD	Mean squared displacement
NAPs	Nucleoid-associated proteins
OC	Open complex

ORF	Open reading frame
PALM	Photo-activation localization microscopy
PCA	Principal component analysis
PCR	Polymerase chain reaction
qPCR	Quantitative PCR
RBS	Ribosome binding site
RNA	Ribonucleic acid
rRNA	Ribosomal RNA
RNAp	RNA polymerase
RT-PCR	Reverse transcriptase PCR
SCR	Single-copy repressilator
SSA	Stochastic simulation algorithm
STORM	Stochastic optical reconstruction microscopy
TALE	Transcription-activator-like effector
TF	Transcription factor
TIRF	Total internal reflection fluorescence
tRNA	Transport RNA
TSS	Transcription start site

List of Publications

This thesis is a compilation of five studies. The first three and the fifth are published and referred to in the text as **Publication I, II, III, and V**, respectively. The fourth is currently under review and is thus referred to as **Study IV**. The publications are reproduced with permission from the publishers.

It is noted that in this manuscript, the author's name is **Samuel Matos Da Costa** (née **Samuel Matos Dias Oliveira**, as named in the publications).

- I. J.G. Chandraseelan, **S.M.D. Oliveira**, A. Häkkinen, H. Tran, A. Sala, I. Potapov, M. Kandhavelu, and A.S. Ribeiro, "Effects of temperature on the dynamics of the LacI-TetR-CI Repressilator", *Mol. Biosyst.*, 9:3117-23, 2013.
- II. **S.M.D. Oliveira***, J.G. Chandraseelan*, A. Häkkinen, N.S.M. Goncalves, O. Yli-Harja, S. Startceva and A.S. Ribeiro, "Single-cell kinetics of the Repressilator when inserted into a single-copy plasmid", *Mol. Biosyst.*, 11:1939-1945, 2015. *Equal contributions.
- III. **S.M.D. Oliveira***, A. Häkkinen*, J. Lloyd-Price, H. Tran, V. Kandavalli, and A.S. Ribeiro, "Temperature-Dependent Model of Multi-Step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements", *PLoS Comp. Biol.*, 12(10): e1005174, 2016. *Equal contributions.
- V. **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, and A.S. Ribeiro, "Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*", *Mol. Microbiol.*, 99(4):686-699, 2016.

Unpublished Manuscripts

- IV. N.S.M. Goncalves*, **S.M.D. Oliveira***, L. Martins, R. Neeli-Venkata, J. Reyelt, J.M. Fonseca, J. Lloyd-Price, H. Kranz, and A.S. Ribeiro, "Chromosome and plasmid-borne *lacO301* promoters differ in sensitivity to critically low temperatures", Submitted. *Equal contributions.

The author of this thesis contributed to these publications as follows.

In **Publication I**, the author performed all microscope measurements and assisted J.G. Chandraseelan with other measurements. The author performed all image analysis of the data from the repressilator and assisted H. Tran with the image analysis of the image data on the genetic switch. The author analyzed all the data along with Antti Häkkinen, and contributed to the writing of the manuscript with A.S. Ribeiro.

In **Publication II**, the author contributed to the design and construction of the SCR, along with J.G. Chandraseelan. The author also conducted measurements for validating this construction. Further, the author performed all microscope measurements, executed all image analysis, and analyzed the resulting data with Antti Häkkinen. Finally, the author participated in the discussions and contributed to the writing of the manuscript with A.S. Ribeiro.

Publications I and II have been used by J.G. Chandraseelan in his Ph.D. dissertation.

In **Publication III**, the author conducted all experimental and microscopy measurements. Second, the author, along with V. Kandavalli, conducted validation measurements by reverse transcriptase PCR (RT-PCR). Third, the author, along with J. Lloyd-Price and H. Tran, performed image analysis. Finally, the author contributed to discussions and to the writing of the manuscript with A.S. Ribeiro.

In **Study IV**, the author designed and validated the functioning of the single-RNA fluorescent probe integrated into the chromosome. Second, the author aided in the design of the experiments and assisted in microscopy measurements, along with N.S.M. Goncalves. Third, the author performed image and data analysis with L. Martins. Finally, the author analyzed the results, participated in the discussions, and contributed to the writing of the manuscript with A.S. Ribeiro.

In **Publication V**, the author conceived and conducted experimental and microscopy measurements, along with R. Neeli-Venkata and N.S.M. Goncalves. The author performed all image and data analysis. Finally, the author analyzed the results, participated in discussions, and wrote the manuscript, assisted by A.S. Ribeiro.

Publication V has been used by R. Neeli-Venkata in his Ph.D. dissertation.

1 Introduction

1.1 Background and Motivation

In all kingdoms of life, organisms have developed regulatory mechanisms that can control the activity of many genes and genetic circuits (Kannan et al. 2008; Arsène et al. 2000) in response to a wide variety of environmental conditions. Temperature is one of the most challenging environment fluctuations for cells to adapt and survive. Because temperature fluctuations can cause rapid changes, e.g. in seconds to minutes, cells have evolved complex different biological responses to cope with them. Endothermic organisms evolved mechanisms to generate and dissipate heat, regulate their body temperature. Meanwhile, ectothermic organisms, such as *Escherichia coli*, are unable to robustly control their internal temperature and, thus, instead, evolved mechanisms that assist adapting the internal processes kinetics to the environment temperature.

E. coli is a widely spread, in most cases harmless, bacteria. In the wild, they live in the intestinal system of mammals with an optimal-controlled temperature, where they assist in the breakdown of carbon compounds (Touchon et al. 2009). Given its relative simplicity, easy propagation and maintenance in the laboratory, *E. coli* has been extensively used to host synthetic genes and genetic circuits. Thus, it has been used in several studies over the years and allowed researchers to better understand the basic mechanisms of molecular genetics. In particular, in studies of environmental fluctuations, where cells must experience temperature shifts, *E. coli* cells have been used to assess the limits and sensitivity of temperature-response mechanisms, e.g. the modifications of chemical-structure or composition of proteins that alter the dynamics of gene expression and network connections (Jana et al. 1999; K. S. Koblan & Ackers 1991; K S Koblan & Ackers 1991). Furthermore, *E. coli* cells have been used in studies of how the cellular biophysical processes change with temperature (Stylianidou et al. 2015; Parry et al. 2014), and how long-term cellular biological adaptations occur, e.g. the regulation of specific genes to avoid potential harms to cells across generations (Richter et al. 2010; Arsène et al. 2000; Sabate et al. 2010).

Regarding *E. coli*'s long-term adaptations to temperature fluctuations, specific genes and gene networks are activated in order to produce heat or cold shock proteins that prevent cellular mechanisms from breaking down (Richter et al. 2010; Verghese et al. 2012). In the case of above-optimal temperatures, heat shock mechanisms provoke the transient expression of cytoprotective proteins, which first attenuate its effects by re-folding non-function proteins, and then prevent protein aggregate formation and avoid changes in cell morphology (Arsène et al. 2000; Richter et al. 2010; Verghese et al. 2012). Meanwhile, in the case of temperature downshifts, the cold shock machinery is activated. The expressed proteins enhance both transcription and translation rates, helping cells to cope with the lowering temperature (Yamanaka 1999; Phadtare 2004). Besides such adaptation mechanisms, other intracellular processes have their dynamics affected by temperature fluctuations.

Cellular processes are regulated by genes and gene networks that can perform complex tasks, such as time-counting (Ko & Takahashi 2006), state holding (Gally et al. 1993) and signal filtering (Wolf & Arkin 2003). Such natural circuits have to be sensitive and robust to environmental changes in order for cells to maintain their cellular functions efficient when subject to external perturbations (Nandagopal & Elowitz 2011; Becskei et al. 2000). For instance, circadian clocks (Dunlap 1999; Mihalcescu et al. 2004), chemotaxis (Oleksiuk et al. 2011), cell cycle oscillators (Pomerening et al. 2008), and physiological adaptations (Süel et al. 2006; Balaban et al. 2004) in bacterial cells have been tested for their stability in varying environmental conditions. Similar performance is desired for synthetic circuits designed for performing complex tasks in live cells (Gardner & Collins 2000; Khalil et al. 2012; Litcofsky et al. 2012). In this regard, since 2000, synthetic oscillatory networks have been constructed, tested, and optimized to approximate the precision of natural circuits (Potvin-Trottier et al. 2016). Further, recent studies demonstrate that they can perform time-keeping tasks at optimal (and some even at sub-optimal) environmental conditions, in organisms ranging from bacteria (Elowitz et al. 2000; Atkinson et al. 2003; Stricker et al. 2008) to mammalian cells (Fussenegger et al. 2009).

The first functional synthetic genetic clock, a ring type oscillator named 'repressilator', was built and implemented in *E. coli* cells (Elowitz et al. 2000) to generate oscillatory signals at optimal conditions. The circuit consists of a negative feedback loop of three genes, one gene repressing the activity of the next. Measurements of a fluorescence signal probing the numbers of one of the component proteins in live cells over time showed that, in optimal conditions, this artificial network oscillates (stochastically) with a period length longer than cells' lifetime (Elowitz et al. 2000). Its noisy behavior arises from fluctuations in the dynamics of its component genes (thus, it is named 'intrinsic' noise to gene expression, i.e. the kinetics of transcription and translation processes) (Leibler & Kussell 2010; Elowitz et al. 2002). However, the cell-to-cell variability in the dynamics of this circuits arises not just from intrinsic noise, but also from 'extrinsic' noise sources, such as variability in the numbers of cell components, cell health, differences in the interactions with the environment, etc.

These circuit's dynamics is temperature dependent since, among other reasons, the interactions between proteins and regulatory operator sites of genes are temperature-dependent (Goncalves et al. 2018; Tran et al. 2015). For instance, while the functionality of the wild-type CI protein has been reported to be maximized at $\sim 30^\circ\text{C}$, its activity gradually reduces ($\sim 50\%$) as temperature increases (Jana et al. 1999) due to either a CI inability to

discriminate between operator sites (K S Koblan & Ackers 1991) in these conditions, and/or to a temperature-dependence of CI 's dimerization (K. S. Koblan & Ackers 1991). In addition, the interactions between proteins and their operators can be further affected by the interactions of these proteins with other specific molecules in the cells, e.g. the regulation of the P_{araBAD} promoter promoted by the interaction between the repressor AraC and the inducer L-arabinose in the *araBAD* operon (Schleif 2002; Bondeson et al. 1993). Interestingly, the temperature-dependence of particular components of gene networks have been used to engineer synthetic circuits whose performance is, purposely environment-dependent (Isaacs et al. 2003), or to compensate for the dynamical changes caused by temperature shifts on other components (Hussain et al. 2014).

Gene expression, which converts genetic information into functional proteins, through transcription and translation reactions, is made possible by complex, multi-stepped, sequence-dependent reactions. In *E. coli*, since translation can begin prior to the completion of transcription, the two processes are dynamically coupled (McClure 1985; Saecker et al. 2011; Ramakrishnan 2002; Jones et al. 2014). Importantly, the major regulatory mechanisms of gene expression dynamics act at the transcription initiation stage. In particular, *in vitro* studies have characterized the key steps in transcription initiation processes (McClure 1980; Browning & Busby 2004; Buc & McClure 1985; Lutz et al. 2001). Recent advancements in the methodologies for live cell imaging with single molecule sensitivity have allowed the kinetics of rate-limiting steps of transcription initiation to be characterized also *in vivo* (Golding et al. 2005; Lloyd-price et al. 2016; Mäkelä et al. 2017).

In *in vivo* studies of transcription dynamics, individual RNA molecules are detected in real-time by the MS2-GFP tagging on RNA sequences (Golding et al. 2005; Xie et al. 2008). From measurements of the numbers of RNA molecules per cell over time and the times between consecutive transcription events, information was extracted on the dynamics of the underlying processes of transcription, i.e. the number and duration of rate-limiting steps in transcription. Initially, this was done by obtaining a distribution of time intervals between consecutive RNA production events in individual cells, the assumption of a multi-step model of transcription, and statistical inference of the kinetics of each step that best fits the data (Häkkinen & Ribeiro 2015). A study made use of this methodology to study how temperature fluctuations affect the dynamics of the underlying steps of transcription initiation for the promoter P_{tetA} (Muthukrishnan et al. 2014). The results showed that, for this promoter, in particular, a third rate-limiting step appears at sub-optimal temperatures, in agreement with past studies using *in vitro* techniques (Buc & McClure 1985).

Gene expression and the activity of gene networks can also be affected by 'network coupling'. Namely, when multiple copies of a circuit are present in a cell, the communication between the components of each copy of the circuit causes their behavior to differ from that of individual circuits. Theoretical studies have addressed the effects of copy-number on the network behavior (Klappenbach et al. 2000; Mileyko et al. 2008; R. Zhu et al. 2007; Zhu et al. 2007; Ribeiro 2007c). One of these studies (Mileyko et al. 2008) has suggested that changes in the copy-number of the repressilator can severely affect the behavior of the circuit, from sustained oscillations to single steady states. In addition, the communications between the components of multiple copies of the circuit may be affected by the intracellular (Klappenbach et al. 2000; Stevenson & Schmidt 2004) and

environmental contexts (Cardinale & Arkin 2012). Thus, when engineering synthetic circuits to be inserted in *E. coli* host cells, one usually needs to deal with unexpected divergences between design and actual function of these systems in real contexts. E.g., having multiple copies of a circuit in a cell has, in some cases, create an excessive burden of important biological cellular functions such as, cell growth and vitality, which leads to unwanted contextual issues (Cardinale & Arkin 2012; Dong et al. 1995).

In this regard, changes in the intracellular context can emerge, directly or indirectly, from environmental fluctuations. To address these, *E. coli* can resort to energy-consuming mechanisms (Govers et al. 2014) that attempt to avoid loss of protein functionality (Deuerling et al. 1999; Wickner et al. 1999), degrading mal-functioning proteins (Viaplana et al. 1997), or neutralize mal-functional proteins by grouping them into protein aggregates (Sabate et al. 2010; Tyedmers et al. 2010; Bednarska et al. 2013). Subsequent studies have shown that once produced, protein aggregates are segregated and accumulated at the cell poles by an energy-free biophysical mechanism (Winkler et al. 2010; Gupta et al. 2014) called ‘nucleoid occlusion’, a phenomenon reported to be temperature-dependent (Gupta et al. 2014). In that, protein aggregates tend to move away from mid-cell by a volume exclusion phenomenon (Winkler et al. 2010; Coquel et al. 2013; Stylianidou et al. 2015), due to higher density of the nucleoid when compared to the cytoplasm. This phenomenon is also believed to affect the spatial distribution of other large molecular species in the cytoplasm, such as plasmids (Vecchiarelli et al. 2012) and other large complexes (Straight et al. 2007). Relevantly, due to the aggregates being mostly at the poles, as cell division occurs, they first become asymmetrically distributed in the daughter cells. Namely, they preferentially locate at the older pole of each daughter cell. As new division events occur, only some daughter cells will carry such aggregates. This allows the generation of ‘rejuvenated’ cells, free from aggregates (Lindner et al. 2008; Govers et al. 2014).

Overall, we focused on the study of the effects of temperature shifts on gene expression and genetic circuits and, more generally, on the cell’s functioning as observed *in vivo*, at the single-cell level. The results contribute to our understanding of the robustness and sensitivity of natural circuits, and thus, to the enhancement of our ability to engineer functional synthetic circuits with pre-defined dynamics.

1.2 Thesis Objectives

We focused on studying the effects of temperature fluctuations on the dynamics of individual genes and of small genetic circuits (switches and clocks), using both stochastic models and microscopy measurements. First, we measured the robustness and sensitivity of the repressilator to temperature fluctuations and searched for possible causes for the lack of robustness observed in the repressilator dynamics, when at higher-than-optimal temperatures. Second, we implemented a single-copy plasmid repressilator from its original sequence, measured the dynamics in the same temperature range, and compared the results with those from the original low-copy repressilator, so as to investigate why single-copy and multi-copy circuits differ in behavior. After that,

we studied the temperature-dependence of difference in the dynamics of the two constructs. Third, we synthesized *de novo* an RNA sequence with multiple binding sequences as part of the construction of a fluorescent probe to be integrated both into *E. coli*'s chromosome and in a single-copy plasmid. This probe allows the detection and tracking of individual RNA molecules as they are produced in cells. We used it to measure the temperature-dependence of the dynamics of transcription initiation process and studied whether this is affected by the gene location, i.e. whether on a plasmid or on the chromosome. Finally, we observed an independent phenomenon that may also affect, indirectly, the functioning of genetic circuits. In particular, we found that the diffusion process in *E. coli*'s cytoplasm is heavily temperature dependent, in that at lower temperatures, it can affect the process of the nucleoid exclusion of aggregates from midcell. It is tangible that a similar, although weaker in effect, phenomenon occurs for some proteins, which would affect the response of some genetic circuits to temperature shifts.

In particular, our objectives are:

I. Study the robustness of the repressilator (periodicity length, amplitude, and shape) and of a genetic switch (stability of its genes kinetics over time) to various temperature conditions.

II. Study the effects of network coupling and transient perturbations, differing in duration, strength, and nature (chemical or temperature-based), on the functionality of the repressilator.

III. Characterize the temperature-dependence of transcription initiation dynamics at the level of the kinetics of the rate-limiting steps, to better understand how temperature shifts affect the functioning of circuits. For that, we perform measurements of RNA production at the single molecule level for varying temperatures and then fit detailed multi-step stochastic models to the data in maximum likelihood sense using statistical methods.

IV. Study the effect of the location of a gene on the temperature-dependence of its transcription initiation dynamics. Namely, for various temperature conditions, we compared the RNA production levels of a gene when located in *E. coli*'s chromosome with when in a single-copy plasmid.

V. Study the temperature-dependence of other cellular processes. For this, we tracked synthetic and natural aggregates in live cells, to measure the relative dynamic viscosity of the cytoplasm as a function of temperature. From this, we studied the effects of changes in this viscosity on the robustness of nucleoid occlusion and aggregate polar segregation.

Objectives I and II were complete in **Publications-I and II**, respectively. Objective III was completed in **Publication-III**. Objective IV was completed in **Study-IV**. Finally, Objective V was completed in **Publication-V**.

1.3 Thesis Outline

This thesis is organized as follows. After the introduction to the problem and objectives described in the present chapter, it follows Chapter 2 that introduces the biological background on the topics covered. Namely, it consists of an introduction of the present knowledge related to the regulatory mechanisms and dynamics of gene expression in bacteria. In addition, it discusses intrinsic and extrinsic sources of noise in gene expression along with known effects of temperature fluctuations on the dynamics of gene expression. Chapter 2 further introduces the design principles of synthetic biology and describes recent advancements in the tools used in molecular biology for DNA assembly and genome editing. At the end of Chapter 2, the concepts of fluorescence microscopy techniques for *in vivo* single-cell, single-molecule measurements are presented. Next, Chapter 3 presents concepts underlying the stochastic models of gene expression, along with the modeling strategy and the simulation algorithm used in the studies included in this thesis. Subsequently, Chapter 4 presents a summary of the main results of the research work conducted. First, it is provided a description of the design process and the results of the methods of validation of the genetic constructs engineered during the course of these research efforts. Second, it is presented the results of perturbation assays performed using a microfluidics platform tailored for keeping cells at various conditions during microscope image acquisition. Third, it is presented the results of image and data analysis methods of the images collected by microscopy. Finally, in Chapter 5, the final conclusions and a discussion on wider issues are provided.

2 BIOLOGICAL BACKGROUND

This chapter provides an introduction to our present knowledge on how the process of gene expression in *Escherichia coli* takes place, with emphasis on its kinetics. In particular, we focus on the multiple steps that occur during the process of transcription initiation. It follows a description of the mechanisms of regulation of the dynamics of transcription and of the concept of operons. Afterward, it is provided a general description of what constitutes a genetic circuit and how they operate in *E. coli*. Further, it is introduced the concept of cell-to-cell variability and what phenomena generate it. Subsequently, the design principles of synthetic biology are presented, along with the most recent molecular biology tools employed for DNA assembly and genome editing. Finally, it is provided an overview of fluorescence microscopy techniques, of in vivo single-cell, single-molecule measurement techniques, and of computational tools used for extracting and processing information from the empirical data, including microscopy images, on single-RNA measurements.

2.1 Gene Expression in *Escherichia coli*

Gene expression is the process by which genetic information, encoded in genes in the form of DNA (a deoxy-ribonucleic acid molecule), is transformed into gene products, such as functional proteins and various non-protein coding RNA sequences that are able to perform specific tasks in the cells. The process of gene expression follows the central dogma of molecular biology (presented in **Figure 2.1**) proposed by (Crick 1970). It consists of two major steps, namely transcription, and translation, which together can modulate the dynamics of expression of a gene (Alberts et al. 2002).

In detail, during transcription, genetic information encoded in genes, i.e. DNA nucleotide sequences, is transcribed into complementary mRNAs (messenger RNA, a ribonucleic acid) (McClure 1985). Subsequently, during translation, the mRNA molecule is used as a template to translate a linear amino-acid structure (Ramakrishnan 2002). This linear structure of amino-acids is then converted into a functional protein by a process named ‘protein folding’, a simpler but similar process to post-translational modifications of higher-

order organisms (Lewin 2008). In bacteria, transcription and translation are dynamically coupled, in that the latter can initiate prior to the completion of the former.

The activity of genes, measured by its mRNAs and proteins production rates, is determined by the affinity of its sequence to interact with RNA polymerases, ribosomes, and other global regulatory molecular species. Also, specific genes response to specific regulatory molecules, such as metabolites, repressors, activators, and ions (Alberts et al. 2002). All of these can change (in number, affinity, etc.), e.g., in response to changes in the extracellular environment (Alberts et al. 2002), which causes the gene's activity to change accordingly.

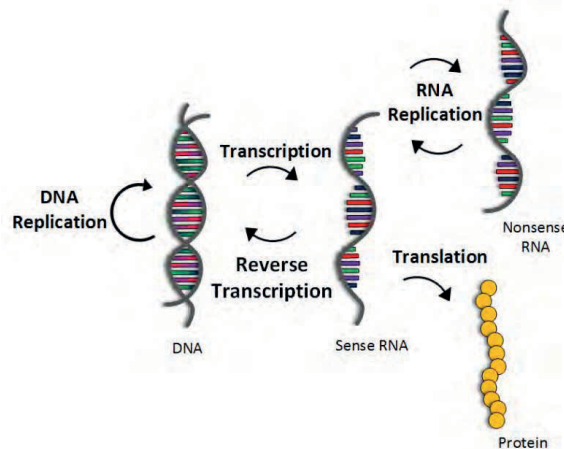


Figure 2.1: The central dogma of molecular biology. The first step of transcription is responsible for the production of an individual mRNA molecule whose sequence is determined by the genetic code in the DNA sequence which is read by RNA polymerases. The second step, translation, generates linear amino-acid sequences from the mRNA sequence, a process performed by ribosomes. Also, before cell division, all genetic information in the double-stranded DNA sequences is duplicated by the process of DNA replication (Alberts et al. 2002). In addition, in some virus species having the enzyme reverse transcriptase and lacking DNA polymerases, double-stranded DNA sequences can be produced from RNA templates by a reverse transcription process. Finally, in many viruses, the replication of the RNA material is made possible by RNA-dependent RNA polymerases (Ortín & Parra 2006).

E. coli is one the most common host organism, due to its simplicity and easy manipulation. Namely, it is used as a host for synthetic genes and genetic circuits used to study the basic mechanisms of molecular genetics, such as DNA replication, RNA degradation, and gene expression (Blattner et al. 1997; Bernstein et al. 2002; Chen et al. 2015). In this organism, many types of DNA sequences can carry genetic information, e.g. the *E. coli*'s genome, bacterial F-plasmids, cosmids, fosmids, and bacterial artificial chromosomes (BAC).

The genome of *E. coli*, in particular, consists of a double-stranded DNA sequence ranging from 4.5Mbp to 5Mbp in length, which contains ~ 4500 genes coding for structural and regulatory functional proteins (Blattner et al. 1997; Bergthorsson & Ochman 1998). Meanwhile, circular plasmids are additional double-stranded DNA sequences, physically separated from *E. coli*'s chromosome that can replicate independently. These usually range from 1kbp to over 300kbp (Thomas & Summers 2008), and usually, carry additional genes that benefit the organism's maintenance and survival (Eliasson et al. 1992). Several molecular biology techniques have

allowed researchers to isolate and engineer artificial bacterial plasmids carrying, e.g., synthetic genes encoding for antibiotic selection markers and functional heterologous proteins (Kahl & Endy 2013).

In prokaryotes, a functional regulatory gene (or multiple genes, i.e. operons) consists of three main units: a promoter region, regulatory operator sites, and one (or more, if an operon) encoding region for a structural or regulatory protein (Alberts et al. 2002). The promoter region is a highly-specific region of the DNA that has high affinity with sub-units of the RNA polymerase (RNAP) core enzyme. These regions are responsible for transcription initiation.

To regulate transcription initiation, highly-specific DNA regions named ‘operator sites’ are located around and sometimes within the promoter region, repressing or activating gene expression when bound by transcription factors (Lutz & Bujard 1997), small regulatory RNAs (Levine et al. 2007), or other regulators, which decrease/increase the binding affinity of RNAP to the promoter region (**Figure 2.2**) (Ellefson et al. 2014).

Importantly, as mentioned above, due to the lack of physical barriers between *E. coli*’s chromosome and the gene expression machinery, transcription and translation are dynamically coupled (unlike eukaryotic organisms). In particular, translation starts soon after the ribosome binding site (RBS) of the mRNA becomes exposed (**Figure 2.2**) (Miller et al. 1970). To initiate translation, a key regulatory operator site, named RBS, attracts a ribosome, which is a complex molecular machine made of small ribosomal subunits. Once this occurs, it follows the elongation process that converts the genetic information on the mRNA into a polypeptide chain by following a triplet-wise degenerated universal code (codon) of nucleotides (Alberts et al. 2002).

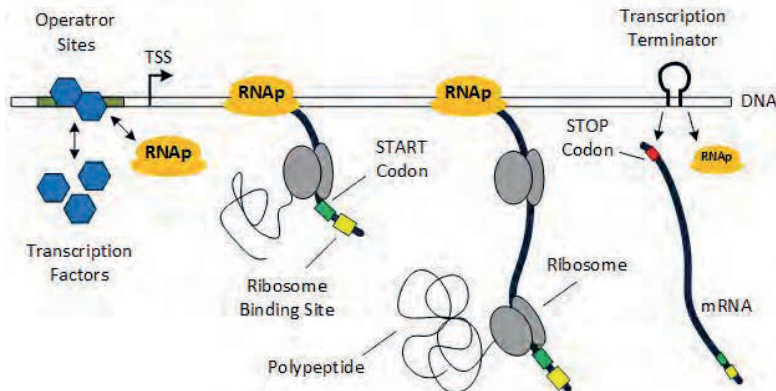


Figure 2.2: Representation of the dynamic coupling between transcription and translation in *E. coli*. RNA molecules are elongated after the RNAP escapes from the promoter region, where the transcription start site (TSS) is located, following the formation of the open complex (OC) (McClure 1980). As soon as the RBS becomes exposed in the newly produced mRNA molecule, linear-structured proteins are synthesized by ribosomes. This allows transcription and translation of a gene to occur simultaneously in *E. coli* (Miller et al. 1970).

Since the expression rate of genes and operons is strongly dependent on the binding/unbinding of regulatory molecules to the DNA, and as these usually exist in small numbers, there are significant temporal fluctuations

in the effective expression rate of gene expression, which cause significant temporal fluctuations in the numbers of gene products. This then causes significant differences in the numbers of gene products between identical cells, even when in the same environment.

2.1.1 Transcription Initiation and its Rate-Limiting Steps

In *E. coli*, the process of transcription consists of three sequential steps, namely, transcription initiation followed by elongation and then termination. These step kinetics depend on the ability of the multi-functional RNAP holoenzyme ($E\sigma$) to bind to a promoter region, to initiate transcription, and then to coordinate the synthesis of an individual mRNA molecule.

The RNAP holoenzyme ($E\sigma$) is composed of the RNAP core enzyme (i.e. a complex of multiple subunits $\beta\beta'\alpha2\omega$) (Murakami et al. 2002; Young et al. 2002) and a σ -subunit, which confers to the RNAP complex affinity to specific sequences in the promoter regions of genes, named consensus sequences, located at -10 (TATAAT) and -35 (TTGACA) positions upstream to TSS (von Hippel et al. 1984; Harley & Reynolds 1987). For example, for genes related to heat shock responses in *E. coli* to be expressed under these conditions, the RNAP core enzyme is bound by a σ^{32} unit (Alberts et al. 2002).

The formation of the RNAP holoenzyme ($E\sigma$) complex and the search and binding of this complex to specific regions of the promoter are followed by a partial unwinding of the double-stranded DNA until a transcription bubble is formed, exposing a small region in each of the strands. Consequently, the transcription bubble allows the RNAP to move along the DNA template in 3' to 5' direction, which consists of the elongation step during which an individual mRNA molecule is synthesized. Lastly, when reaching a specific DNA structure known as the transcription termination site, the newly synthesized mRNA and the RNAP are released from the DNA template (Nudler & Gottesman 2002).

Studies suggest that it is during transcription initiation that most regulatory mechanisms act on, implying also that it is the kinetics of this stage that most affects the expression kinetics of *E. coli* genes (McClure 1985; Browning & Busby 2004; Saecker et al. 2011). For that reason, the steps during transcription initiation are referred to as the rate-limiting steps of transcription. Evidence suggests that there are two main rate-limiting steps of active transcription (Walter et al. 1967; Chamberlin 1974; McClure 1985; Saecker et al. 2011), represented in Equation 2.1.



In Equation 2.1, the first step represents the binding of an RNAP holoenzyme (R) to a promoter (P), with an equilibrium constant (K_B), which allows forming the closed complex (CC). This is a reversible step, which can occur several times. When successful, it follows an isomerization reaction with a rate constant (k_i) to form the open complex (OC). At this stage, the process is generally irreversible. Once complete the OC formation, it follows the relatively fast step of promoter escape. In particular, the RNAP holoenzyme attempts to synthesize a small portion of the DNA template (~ 10 nucleotides) (Hsu 2002) via a *scrunching* mechanism that pulls the

downstream DNA into itself while keeping its position fixed on the promoter region (Revyakin et al. 2006). During this stage, depending on a 6-8 nucleotides non-consensus sequence located upstream to TSS of the promoter, transcription elongation should occur by following two possible ways, i.e. the ‘branched’ and ‘unbranched’ mechanisms (**Figure 2.3**) (Susa et al. 2006; Henderson et al. 2017).

In the ‘branched’ mechanism, based on *in vitro* studies of the dynamics of transcription initiation of the *lacT7A1* and of λ_{PR} promoters (Susa et al. 2006), transcription initiation has been considered to account for a productive and a non-productive pathway. In a productive pathway, an RNAP is capable of escaping from the promoter region and synthesizing a long mRNA sequence without releasing a short (abortive) RNA molecule. Meanwhile, in the non-productive pathway, the abortive cycling only leads to the production of abortive RNA molecules, and never moves towards the formation of EC (**Figure 2.3**).

Importantly, most studies of bacterial transcription initiation have considered the ‘unbranched’ as the most common mechanism driving the formation of the elongation complex (EC), in which all OCs undergo rapid cycles of abortive initiation, until a productive initiation and synthesis of RNA are achieved (Hsu 2002; Straney & Crothers 1987). This accumulated stress during transcription initiation is used to successfully release RNAP holoenzyme and transcription factors from the promoter region, thus making the formation of the transcription elongation complex (EC) and elongation phase to begin (Kapanidis et al. 2006). Although relevant, this process is expected to be much higher than all other rates (e.g. CC and OC), and thus assumed to be ‘negligible’ in terms of defining the rate-limiting steps of transcription initiation (Equation 2.1) (Bremer & Dennis 1996). For this reason, the rate constant of this process in Equation 2.1 is assumed to equal ‘infinite’, since its time-length is much short than that of the previous steps (Hsu 2002).

The characterization of the rate-limiting steps in *E. coli* genes has been conducted using *in vitro* and *in vivo* methods (Buc & McClure 1985; McClure et al. 1978; McClure 1980; Lutz et al. 2001; Lloyd-price et al. 2016). In the first proposed *in vitro* method, the derivation of the OC formation rate is based on the time taken for abortive initiation products to be formed until reaching a state-steady (McClure et al. 1978). In another *in vitro* method, the CC formation rate is derived from the linear-dependence of this step with varying RNAP concentrations on transcription *in vitro* assays (Buc & McClure 1985; Cech & McClure 1980). Relevantly, this direct linear relationship between the lag times for RNA production and the inverse of the reciprocal RNAP concentration can be drawn in a Lineweaver-Burk plot (Lineweaver & Burk 1934), named as ‘ τ -plot’.

From the line drawn in τ -plots, the mean time for CC formation is obtained from its slope and the mean time for OC formation from its intercept with the y-axis (McClure 1980; Patrick et al. 2015). Since the duration of such steps has been reported to range from seconds to minutes, much longer than elementary steps in enzyme-catalyzed reactions, and to be sequence-dependent as they differ between promoters (Lutz et al. 2001; Jones et al. 2014), such steps are considered to be the rate-limiting steps in transcription initiation of *E. coli* genes (Buc & McClure 1985; Saecker et al. 2011).

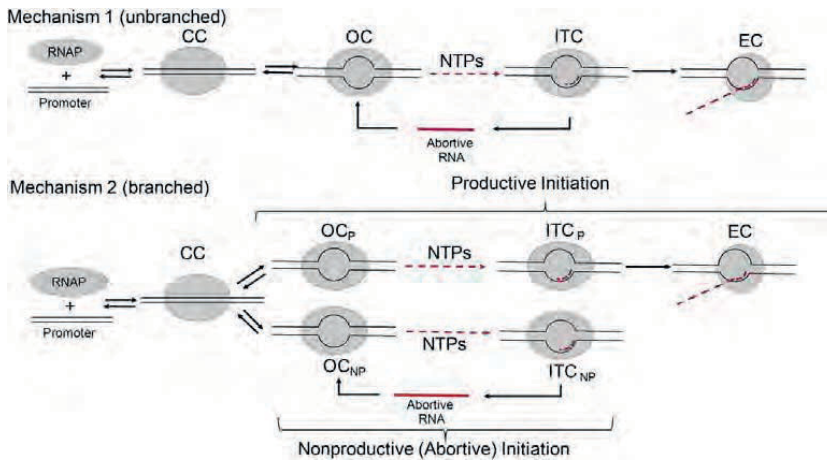


Figure 2.3: Proposed mechanisms of transcription initiation and promoter escape by bacterial RNAP. In the first mechanism, RNAP holoenzyme containing a specific σ -subunit binds to a promoter region and forms a closed complex (CC). This is followed by the formation of the open complex (OC) responsible for creating the transcription bubble, exposing a short sequence of nucleotides. After the OC is formed, abortive synthesis occurs, on the pathway to promoter escape, during the initial transcribing complex formation (ITC). Following an abortive initiation cycle, the mRNA starts to be synthesized by the transcription elongation complex (EC). In that, the RNAP holoenzyme moves along the DNA template through alternative pathways, such as spontaneous or transcription factor-mediated pausing, until finding the transcription termination site, leading to the release of the σ -subunit of the RNAP core enzyme and the complete mRNA molecule. Second, in the branched mechanism, two classes of initiating complexes are proposed: the formation of a productive complex (OC_p) that can escape from promoter without releasing any short RNA; and the formation of a nonproductive (abortive) complex (OC_{np}) that cannot escape towards the elongation complex (EC), and can only synthesize and release short RNA sequences. Adapted and reprinted with permission from (Henderson et al. 2017); Copyright 2017, PNAS.

Interestingly, using the concept underlying the *in vitro* ‘ τ -plot’, the number and duration of rate-limiting steps in transcription initiation were also characterized using *in vivo* measurements (Lloyd-price et al. 2016; Kandavalli et al. 2016; Mäkelä et al. 2017). It is important to note that, when implemented in live cells, this method can present severe limitations due to the limit at which the *in vivo* intracellular concentration of RNAP is increased while maintaining the cells’ functionality unaffected. This method is based on measuring time-intervals between consecutive RNA production events in individual cells using an RNA fluorescent probe to detect individual RNA molecules in live cells in the microscope, then applying statistical analysis from these distributions to infer the number and duration of underlying steps in transcription initiation. In recent years, this approach has been widely employed to characterize the kinetics of transcription initiation of multiple promoters in various environmental conditions. The results of these studies suggest that transcription kinetics can be explained by the existence of multiple rate-limiting steps (Muthukrishnan 2014; Mäkelä 2016), in agreement with *in vitro* characterization of the rate-limiting steps (McClure 1985; Saecker et al. 2011; Lloyd-price et al. 2016; Mäkelä et al. 2017).

Other studies have used *in vivo* single cell imaging microscopy techniques to study transcription by measuring mRNA molecules in populations of cells using fluorescence in situ hybridization (FISH) microscopy. In this regard, recent studies have shown that transcription initiation is a sequence-dependent process that dictates the

mean and variability of mRNA numbers in a population of cells (Jones et al. 2014; So et al. 2011). In general, most of the studies above suggest that the mean rate and variability of transcription process are promoter-dependent and that the regulatory molecules present in the promoter regions can accelerate/reduce the duration of one or more steps involved in transcription initiation. Further, the dynamics of these steps is not immune to environmental factors such as temperature (Muthukrishnan 2014). In this study, in particular, it has been shown that the best-fitting models explaining the empirical data contain two main rate-limiting steps that are associated with the closed and OC formation. Further, the results showed that a third step emerges and becomes also a rate-limiting step as temperature decreases from optimal (down to 24°C). This additional temperature-dependent step was hypothesized to be the isomerization step that happens immediately before the completion of the OC formation, in agreement with *in vitro* studies on the transcription dynamics of the *lacUV5* promoter (Buc & McClure 1985), and multiple failures in promoter clearance and escape (Hsu 2002). In **Publication III**, from the distributions of time-interval between RNA production events in individual cells, we assessed the temperature-dependence of the kinetics of the rate-limiting steps in transcription initiation, when subject to temperature shifts.

Finally, recent studies have shown that DNA supercoiling, which affects DNA compaction and thereby the activity of most genes in *E. coli*'s chromosome, is also affected by temperature shifts. This can occur by direct means, such as tuning the activity of different DNA topoisomerases (López-García & Forterre 2000), or by indirect means, such as varying the binding affinity of nucleoid-associated proteins (NAPs) to the chromosome (Amit et al. 2003). Thus, the temperature-dependence of the kinetics of DNA-supercoiling in chromosomal genes is assumed to be one of the most important factors regulating gene expression in *E. coli* (Jones et al. 1987). In **Study IV**, we performed a similar quantitative analysis of the temperature-dependence of the kinetics of RNA production, relative to the optimal temperature condition, when the $P_{lacO3O1}$ promoter is located in a single-copy plasmid and in *E. coli*'s chromosome while being subject to temperature fluctuations.

2.1.1.1 Transcription Regulation

To function properly, cells need to produce essential, functional and structural proteins at the proper time during their lifetime. From DNA to proteins, cells have to regulate when and how genes are expressed by tuning transcription and translation. Studies of the dynamics of transcription in *E. coli* suggested that, in prokaryotes, the kinetics of transcription initiation and its key regulatory components (e.g. RNAP core enzyme, σ factors, and transcription factors) are the major factors controlling the production of mRNA and proteins (Farewell et al. 1998; Rouvière et al. 1995; Dong & Schellhorn 2009; McClure 1985; Browning & Busby 2004), while the degradation of these products, at optimal conditions, is kept at nearly constant rate for most genes of *E. coli* (Bernstein et al. 2002; Chen et al. 2015). Therefore, *E. coli* has evolved regulatory systems that can control when, and by how much, transcription starts/stops by turning genes on/off from direct binding/unbinding reactions of transcription factors (e.g. activator and inhibitor molecules) to promoter regions (Schleif 2002; Becker et al. 2013; Browning et al. 2009).

In these regions, activator molecules up-regulate transcription by either directly interacting with RNAPs, or promoting DNA conformational changes, which expose the consensus regions of the promoter thus increasing the binding affinity of RNAP to promoters (Busby & Ebright 1999), e.g. the activation of *araBAD* promoter by AraC activators (Schleif 2010). Meanwhile, during transcription initiation, repressor molecules can down-regulate transcription by either competing with RNAP molecules to occupy the promoter region (e.g. the occupancy of O2 and O3 operator sites of the native *lac* promoter by LacI repressors (Oehler 2009)), or by preventing the steps of OC formation or promoter escape from happening (Sanchez et al. 2011). The transcription mechanism of both native *araBAD* and *lac* operons are presented (**section 2.1.1.2**).

As described, promoter sequences play an important role in determining the regulation of transcription (Jones et al. 2014). Besides, the configuration of its rate-limiting steps is also important to determine the kinetics of mRNA production. Further, transcription initiation can be tuned by the interactions between the promoter region and local and global regulatory factors that are determined by the intracellular and external contexts (Bremer & Dennis 1996), such as the activity and intracellular concentration of RNAP, of specific σ factors (Kandavalli et al. 2016) (Browning & Busby 2016; Kandavalli et al. 2016), of small-ligands (e.g. ppGpp) that regulate the synthesis machinery of ribosomes in *E. coli* (Browning & Busby 2004), among others.

Regarding σ factors, since transcription kinetics in most promoters of *E. coli* is a $\sigma 70$ -dependent process, the intracellular concentration and activity of this transcription factor are also critical factors determining the expression profile of cells. When cells undergo specific stress conditions, other σ factor units are expressed and specific cellular response mechanisms are triggered, defining global changes in the dynamics of transcription of multiple genes and gene regulatory networks (Ishihama 2000). In particular, when cells are subject to sub-optimal temperatures, cells respond and adapt to it by precisely regulating the activity of specific genes that, e.g. are responsible for heat/cold shock responses (Kannan et al. 2008; Arsène et al. 2000).

Additionally, the activity of a gene located in *E. coli*'s chromosome can be locally and globally affected by DNA compaction due to, either the activity of DNA-supercoiling (Stuger et al. 2002; Holmes & Cozzarelli 2000; Dorman 2006; Dorman & Dorman 2016), or the functionality of nucleoid-associated proteins (NAPs) and other DNA-binding proteins (Dillon & Dorman 2010). Thus, depending on the degree of compaction and the position of each gene in the chromosome, the expression of a gene can be up- or down-regulated (Dorman & Dorman 2016).

Further, *E. coli*'s chromosome has segments of topological constraints in its structure such that the effect of DNA-supercoiling on DNA compaction can be accumulated (built-up) with the increase of transcription activity in neighboring regions (Postow et al. 2004; Hardy & Cozzarelli 2005; Rovinskiy et al. 2012). Importantly, since plasmids have transient protein binding activity, they have been reported to have transient topological constraints (Leng et al. 2011; Chong et al. 2014). However, studies have shown that, plasmids encoding for membrane-associated proteins, carrying tandem copies of DNA-binding sites, and carrying the T7 promoter, in *E. coli* strains lacking *topA* gene (of topoisomerase, responsible for preventing excessive supercoiling in

DNA segments), can exhibit longer-lasting topological constraints (Boeke & Model 1982; Pruss & Drlica 1986; Fulcrand et al. 2016; Samul & Leng 2007).

Interestingly, Chong, S. *et al.* have recently shown by *in vitro* measurements that, due to the lack of topological constraints, in plasmids, transient DNA supercoiling build-ups are freely diffused in the plasmid DNA structure in opposite directions, leading to their eventual annihilation (Chong et al. 2014). Finally, *in vivo* measurements have suggested that, prior to annihilation, transient supercoiling can influence the activity of genes located in both plasmids and chromosomes (Samul & Leng 2007; Rahmouni & Wells 1992; Moulin et al. 2005). In **Study IV**, we observed the differences in the dynamics of transcription initiation of a promoter when located in a plasmid and in the chromosome of *E. coli*.

2.1.1.2 *lac* and *ara* Operons

In bacterial genetics, an operon is defined as a functional unit containing two or more structural genes regulated by the same transcription unit, with transcription regulators and a terminator. This definition and the first operon, the *lac* operon, were described by Jacob and Monod in 1961, from empirical observations of the behavior of bacterial cultures when selecting for preferable sugar sources in the media. This operon consists of three neighboring genes that encode specifically for three proteins, responsible for the absorption and metabolism of lactose in *E. coli*. The choice for lactose is driven by complex mechanisms of activation and repression of transcription in the *lac* promoter region, where interactions between transcription factors (lactose inducer and LacI repressor) and promoter-operator sites (O1, O2, and O3) occur (Jacob & Monod 1961).

The O3 operator site is located upstream to the *lac* promoter consensus region. The O1 site is located immediately downstream the promoter region, and the O2 site located far from the promoter region, within the *lacZ* gene coding sequence (**Figure 2.4**) (Schlax et al. 1995; Pruss & Drlica 1986; Riggs et al. 1970). The three structural proteins encoded by the *lac* promoter are, respectively, the β -galactosidase (LacZ), the lactose transmembrane permease (LacY), and the transacetylase (LacA). These enzymes are responsible for, respectively, the process of lactose breakage into galactose and glucose, the intake of lactose from outside into the cell, and the transferring of an acetyl group from coenzyme A (CoA) to galactosidase (Jacob & Monod 1961; Lewis 2005).

Regarding its function, while in the presence of lactose in the media, the *lac* operon modulates the absorption and metabolism of lactose. Meanwhile, in the absence of lactose, the operon remains transcribing its genes in small amounts. During the latter, cells preferably produce the catabolites to intake and process other sugars as a primary carbon source.

The down-regulation of the *lac* promoter occurs when lactose is absent and LacI monomers are able to form dimers, then dimers of dimers (i.e. homo-tetramers) (Beyreuther et al. 1973; Gilbert & Maxam 1973; Miller 1980). LacI dimers and tetramers are able to bind to promoter's operator sites, preventing RNA polymerase from binding to it. This process turns-off the production of lactose metabolites, forcing the cells to metabolize

a different carbon source (Ohshima et al. 1974; Lewis 2005). For the *lac* promoter to be fully repressed, a coordinated binding of LacI dimers, between O1 and either O2 or O3, must occur, thereby causing the formation of a DNA loop and blocking transcription initiation (Oehler et al. 1990; Becker et al. 2013).

The up-regulation of *lac* occurs when lactose is present. In that, the natural allolactose (or its artificial derivative, IPTG) act as an inducer of the promoter by binding with LacI repressors and changing its conformational structure. This reduces the affinity of LacI dimers to promoter's operator sites (Barkley et al. 1975; Lewis et al. 1996; Lewis 2005) and allows RNAP to bind to the promoter and transcribe the genes of the operon for lactose metabolism (**Figure 2.4**) (Shuman & Silhavy 2003).

Besides the regulation achieved by lactose and LacI interactions with promoter's operator sites, transcription activity of *lac* can be further regulated by the presence and absence of glucose in the media. Namely, when glucose and lactose are present, cells preferably metabolize glucose as a primary source, in detriment of lactose. Lactose is only consumed after the former is completely depleted from the media. This preference and selection for sugar source are made possible by the interaction between the secondary messenger cyclic adenosine monophosphate (cAMP) and a fourth operator site, named 'catabolite activator protein' (CAP), located upstream the promoter consensus region next to the O3 operator site (Beckwith et al. 1972; Reznikoff et al. 1974), which positively regulates the transcription of the *lac* operon (Emmer et al. 1970). In that, when glucose is consumed by cells via the glucose metabolism, causing the intracellular concentration of glucose to be reduced, the intracellular levels of cAMP increase, leading to the formation of a protein complex, named cAMP receptor protein (cAMP-CRP), which acts as an activator of the *lac* operon by increasing the affinity of the RNAP to the promoter region (Zubay et al. 1970).

Similar to the mechanism employed by the *lac* operon to control sugar preference and metabolism, another well-known operon of *E. coli*, named *araBAD*, can control the selection and breakdown of the sugar L-arabinose when glucose is not present in the media (Helling & Weinberg 1963; Englesberg et al. 1965; Schleif 2000b). Namely, depending on the metabolic need of the cells, and the concentration of L-arabinose in the intracellular environment, the operon can decide which task to perform, e.g. activation or repression of L-arabinose metabolites. It controls the expression of three structural genes encoding for the enzymes required for L-arabinose metabolism into D-xylose-5-phosphate for the pentose phosphate pathway (**Figure 2.5**) (Schleif 2000b).

Prior to the metabolism of L-arabinose, two different transport mechanisms have been reported to support *araBAD*, i.e. low- and high-affinity transport systems that are able to control the uptake of L-arabinose (Lee et al. 1981; Hogg & Englesberg 1969; Schleif 2010). In the first, once *araE* gene is active, the L-arabinose transporter protein (AraE) binds to the inner membrane, transporting L-arabinose molecules from outside to cell cytoplasm by electrochemical potential driving forces (Lee et al. 1981). In the second, once *araFGH* operon is active, the high-affinity protein transporter complex (ABC transporter) uptakes L-arabinose by specifically binding to it via an active transport mechanism, which uses energy from the hydrolysis of Adenosine Triphosphate (ATP) molecules (Hogg & Englesberg 1969; Schleif 2010).

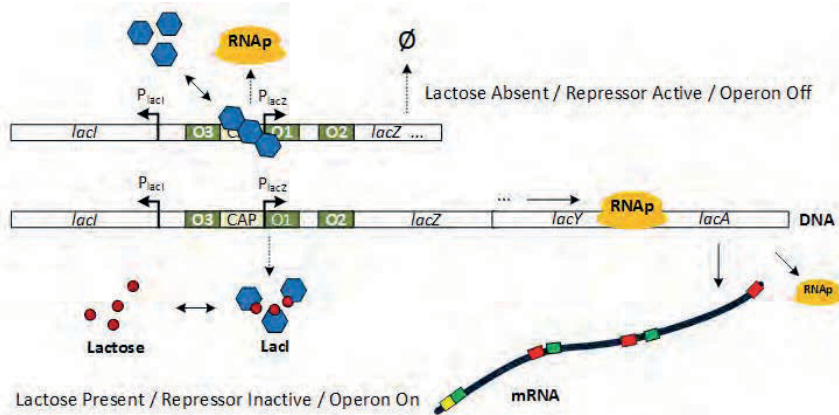


Figure 2.4: Representation of the native *lac* operon of *E. coli*. The *lac* operon contains the *lacZ*p1 promoter region with the operator sites for LacI repressor binding, along with the three genes (*lacZ*, *lacY*, and *lacA*) responsible for lactose metabolism and the terminator site for transcription. Also shown from up- to downstream the *lac* promoter region are the locations of the three *lac* operator sites (O₃, O₁, and O₂), the catabolite activator protein (CAP) binding site, and the TSS. Further, the *lacI* gene responsible for LacI repressor production is constitutively produced by its own native promoter unit located upstream to the *lac* operon in *E. coli*'s chromosome (Shuman & Silhavy 2003).

After the concentration of L-arabinose crosses a certain threshold inside the cell, the activation of the P_{araBAD} promoter is achieved by the interaction of L-arabinose with AraC dimers, which forms an activator complex that is able to bind to I1 and I2 half-sites and increase the affinity of RNAP to bind to the promoter consensus regions. This process allows the initiation of transcription in the P_{araBAD} promoter and the two promoters responsible for L-arabinose intake (Lee et al. 1981; Hogg & Englesberg 1969; Schleif 2010).

The P_{araBAD} promoter is repressed when in the absence of L-arabinose. Consequently, P_{araC}, the promoter responsible for AraC production is also repressed (Figure 2.5). In that, AraC dimers remain bound to the half-operator I1 and O2 site, forming a DNA loop that prevents RNAP from binding to the promoter region and initiating transcription (Figure 2.5) (Lobell & Schleif 1990; Johnson & Schleif 1995; Schleif 2000a; Schleif 2010).

The engineering of synthetic promoters, e.g. variations of *lac* and *araBAD* promoters, and their dynamical studies have been performed as a means to understand how natural systems work, and to increase the degree of predictability in the design of artificial genetic systems (Studier & Moffatt 1986; Elowitz et al. 2000; Golding & Cox 2004). In the past decades, multiple combinations of the native *lac* operator sites have been engineered in various synthetic promoters, from harboring only one operator to containing all three operators with no CAP binding site, to demonstrate how a variety of transcription responses can be achieved when compared to the native expression levels (Oehler et al. 1990; Oehler et al. 1994). Making use of this knowledge, Lutz et al. created an artificial promoter, named 'P_{lac-ara-1} promoter', which is a combination of promoter elements extracted from the both P_{lac} and P_{araBAD} native sequences, point-mutations and synthetic fragments (Lutz & Bujard 1997) to study the kinetics of transcription regulation in *E. coli*.

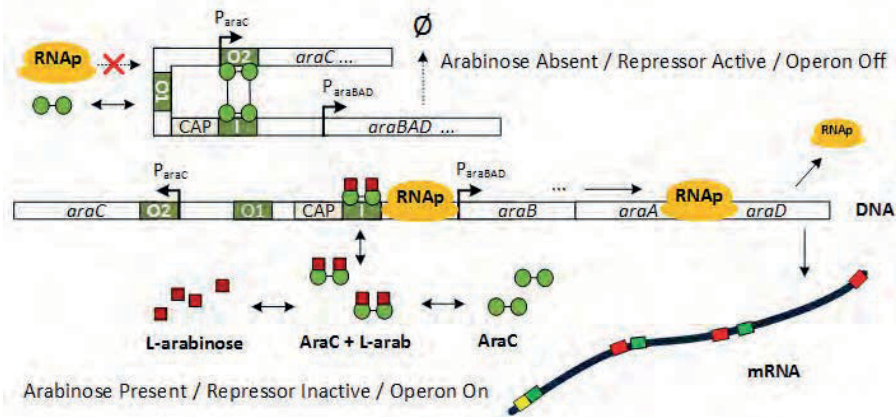


Figure 2.5: Representation of the native *araBAD* operon of *E. coli*. The *araBAD* operon contains the bidirectional P_{araC} - P_{araBAD} promoter and operator sites for dimeric AraC activator/repressor binding. On one side, the P_{araBAD} promoter region is followed by the three genes (*araB*, *araA*, and *araD*) responsible for L-arabinose metabolism and the terminator site for transcription. On the other side, the P_{araC} promoter is followed by the *araC* gene responsible for AraC repressor expression. Also shown are the operator sites of both promoters, along with the CAP binding site and TSSs. In addition, the operator sites I1 and I2 are named as half-sites when bound by a single subunit of AraC. The O1 operator site consists of two-half-sites and regulates the activity of the P_{araC} promoter to express AraC. The O2 site is a single half-site and regulates the activity of P_{araBAD} promoter (Schleif 2010).

In **Study IV**, we made use of a derivative *lac* promoter carrying two of the three native operator sites (O3 and O1). This promoter has lower repression strength than the native promoter. We use it to study the temperature-dependence of the rate-limiting steps in transcription initiation due to its location in the chromosomal or a plasmid DNA. In **Publications III**, we made use of the promoter $P_{lac-ara-1}$, mentioned above, to measure the temperature dependence of the kinetics of the rate-limiting steps in transcription initiation in varying induction schemes. We then made use of the same system to study the temperature-dependence of the biophysical properties of cell cytoplasm at fluctuating temperatures in **Study IV**.

2.1.2 Gene Regulatory Networks and Genetic Motifs

In the previous section, from two examples of native operons in *E. coli*, we showed how a single promoter can coordinate the expression of two or more genes. These genes, while having a completely different structure and biological function, can work together to achieve a desired task inside the cells. In this section, we describe how genes can coordinate their functions into creating complex tasks, not possible by individual genes independently.

Individual genes have a limited set of dynamical functions, such as the graded production and the degradation of specific mRNAs and proteins. In a system of interactive genes, or gene regulatory networks (GRNs), connections among the genes can govern the communication and coordination of more complex dynamics. These connections are able to coordinate the interdependent expression of multiple mRNAs and proteins, and are regulated by molecular regulators, such as DNA, RNA, transcription factors, protein complexes, and other

molecules from intra- and extra-cellular contexts, to control complex biological functions such as cell division (Mashaghi & Dekker 2014), sugar metabolism (Ozbudak et al. 2004), and amino-acids and structural protein synthesis (Chubukov et al. 2014). These connections, established between a molecular regulator and a specific DNA binding sequence (recognition motifs), form the physical architecture or topology of GRNs.

Natural motifs are sets of genes with highly specific connections among them that are usually responsible for executing complex behavioral patterns, such as amplitude or frequency filtering, time tracking, decision making, etc. (Wolf & Arkin 2003). Each of these tasks requires a specific topology. E.g. decision making is made possible by pairs of genes repressing one another, as in the bi-stable switch of (Gardner et al. 2000), while time tracking is made possible by having three genes repressing each other in a loop formation, as in the repressilator circuit of (Elowitz et al. 2000).

This can also be achieved by self-regulating individual genes, via either ‘positive’ or ‘negative’ feedback loops, which allow an individual gene to, respectively, self-activate or self-repress its activity. Such simple circuits allow performing non-linear tasks, such as signal pulsing (Levine et al. 2013). The characterization of various motifs has provided a better understanding of the functionality of regulatory mechanisms in live cells. Synthetic circuits have also been created to explore the potential use of the motifs to externally control internal cellular processes. E.g. there have been designed toggle genetic switches, that can switch between two steady-states by external signaling, with each state being stable across generations (Gardner et al. 2000), and genetic oscillators that create periodic signals (Elowitz et al. 2000).

Importantly, due to the stochastic nature of the interactions between genes, and the small number of molecular species involved in these interactions, in some motifs small fluctuations can ‘propagate’ between the component genes, thus influencing the stability of the corresponding biological function (Paulsson 2005a; Elowitz et al. 2002).

In this regard, there are genetic motifs that can suppress/amplify fluctuations, thereby, attenuating/increasing cell-to-cell heterogeneity (Paulsson & Ehrenberg 2001; Hilfinger & Paulsson 2011; Paulsson 2004). Regarding the latter, a study reported the construction of an artificial genetic AND logic gate, which can act as an amplitude filter inside cells. In that, the expression of an output gene is triggered only within a certain magnitude range of an input signal (near-digital AND gate behavior), which is regulated depending on the heterogeneity of the signal observed at each state (media conditions) (**Figure 2.6**) (Anderson et al. 2007). Relevantly, most natural and artificial GRNs can act as low-pass frequency filters (Samoilov et al. 2002) given the long-lasting duration of transcription rate-limiting steps (Lloyd-price et al. 2016) and the short time taken by mRNA and protein degradation (Taniguchi et al. 2010).

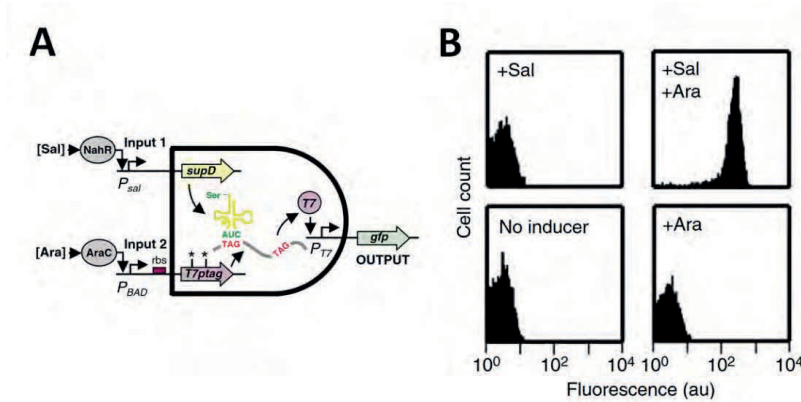


Figure 2.6: A synthetic genetic AND logic gate and the dynamics of the output signal measured from single-cell measurements. (A) The schematic representation of the AND logic gate constructed with two promoters as inputs and one promoter as the output of the gate. The first P_{sal} promoter, regulated by the inducer salicylate and the TF NahR, controls the transcription of the nonsense suppressor transport RNA (tRNA), *supD*, which enables the translation of polymerase, while the second P_{araBAD} promoter (regulated by the inducer L-arabinose and the repressor AraC) controls the transcription of T7-RNA polymerase. T7 polymerases are only expressed when SupD is present in the cells. This enables the two stop codons (T7ptag, shown in 'red'), present in the T7-polymerase RNA sequence, to encode for serine amino-acid, allowing translation to generate a full-length functional T7-polymerase. From this, the output, a P_{T7} promoter, controls the expression of the fast-degrading green fluorescent protein. (B) To determine the system behavior at the population level, fluorescence signal in individual cells were measured by a flow cytometer. When both arabinose and salicylate are present, the entire population of cells is turned 'on' (top-right corner). Meanwhile, when only one of the inducers is present, the entire population is turned 'off'. Adapted and reprinted with permission from (Anderson et al. 2007); Copyright 2007, EMBO and Nature Publishing Group.

In *E. coli*'s chromosome, thousands of genes are linked through regulatory molecules, e.g. DNA-binding proteins (Tamsir et al. 2011) and DNA recombinases (Bonnet et al. 2013), to form GRNs (as e.g. the *E. coli*'s operons previously described) able to perform specific tasks (Martinez & Walhout 2009; Martínez-Antonio et al. 2012). Based on this knowledge, researchers have been able to engineer synthetic circuits with pre-defined structure and dynamics, such as the logic gate described (Figure 2.9) (Brophy & Voigt 2014; Siuti et al. 2013).

2.1.2.1 Natural and Synthetic Small Genetic Circuits

Bacterial cells have evolved robust small genetic circuits that play important roles in determining cell-to-cell heterogeneity and cell fate determination (Kaern et al. 2005; Norman et al. 2015). One essential task for cells is the ability to make decisions. This task can be achieved by natural toggle switch systems, where two operons mutually inhibit one another. In that, repression occurs at the transcription unit level, and a gene produces the repressor of the other. An example of a natural genetic switch is the lysis-lysogenic regulation switch, which is introduced in *E. coli* following λ bacteriophage infection (Lederberg & Lederberg 1953).

This system has been widely studied, theoretically and empirically, to investigate mechanisms of cell fate determination (Friedman & Gottesman 1983; Santillán & Mackey 2004; Zong et al. 2010). In one of these

studies, the results showed that infected cells follow either lysogenic or lytic pathways according to the intracellular fluctuations in the concentration of one of the regulatory proteins of the system, originated from fluctuations in its production (Arkin et al. 1998). Other recent studies have used similar synthetic versions of the genetic switch to studying the dynamics of cell states and fate determination in bacteria (Gardner et al. 2000) and in mammalian cells (Kramer et al. 2004).

Another crucial cellular task is the ability to control temporal events. This function is performed naturally by internal clocks that regulate, periodically, the activity of cellular processes, such as the circadian clock (Dunlap 1999), the cell cycle (Tyson et al. 2001; Pomerening et al. 2005), and the mechanism that defines cellular division point along the cell major axis (Shih et al. 2003). Relevantly, natural oscillators have been also reported in more complex organisms, such as *Sacharomyces cerevisiae* (Mihalcescu et al. 2004), and *Drosophila melanogaster* (Hardin et al. 1990; Glossop et al. 1999). Similar to natural genetic switches, a number of theoretical and experimental studies have provided important insights on the dynamics of natural systems that aid the engineering of synthetic oscillators (Gardner et al. 2000; Elowitz et al. 2000; Fung et al. 2005).

The main goals of the studies aiming at engineering synthetic circuits have been: (i) to test the design principles of synthetic biology (described in chapter 2.2.1); (ii) to explore potential applications of these circuits; and (ii) to understand the complex architecture, regulatory interactions and behavior of natural GRNs.

The rational combination of *in silico* predictions and *in vivo* validations of the dynamics of parts and systems compose a bottom-up approach to engineer more complex artificial GRNs with predictable behavior. In particular, in what regards the design of such circuits, one should take into account the information on the structural and dynamical profile of each bio-parts to be used (e.g. transcription units, transcription factors, network motifs, transcription terminators, RBSs, etc.). Computational tools have aided in the organization and assembly of bio-parts into genetic devices, as building blocks (Clancy & Voigt 2010; Rodrigo et al. 2012), then the implementation of synthetic systems into live organisms (Weiss et al. 2003; Voigt 2006; Mutalik et al. 2013).

An early synthetic circuit implemented in *E. coli* cells was a toggle switch, consisting of two genes, carried by a plasmid, with one repressing the activity of the other, and vice-versa. When implemented in live cells, as previously described, this system is able to switch between two stable states (Gardner et al. 2000). Namely, for a switch of two genes, A and B, one state being ‘A-On and B-Off’ or another state being ‘A-Off and B-On’. Further, due to being biased to favor one state over the other as a function of media conditions, this system has been shown to be capable of making decisions following environmental changes (Gardner et al. 2000).

In 2000, the first functional synthetic oscillator, named as ‘repressilator’, was implemented in *E. coli* cells, carried by plasmids. The ‘repressilator’ consists of a small network of three transcription units in a negative feedback loop formation, where the activity of one gene represses the expression of the subsequent gene in the loop (**Figure 2.7-A**) (Elowitz et al. 2000). To report the dynamics of the system, a second plasmid, carrying one of the three promoters present in the repressilator, $P_{L-tetO1}$, controls the expression of a green fluorescence

protein (GFP). Interestingly, from the fluorescence signals observed by time-lapse images, the repressilator presented noisy GFP oscillations with periods of hours, longer than cell lifetime.

Next, in 2003, the first artificial oscillator implemented into *E. coli*'s chromosome was constructed with altered connectivity in that, depending on the interactions established between a 'repressor' module and an 'activator' module, the genetic system is able to escape from damped oscillations regime to a toggle-switch behavior (Atkinson et al. 2003).

Following the same concept explored by (Atkinson et al. 2003), Stricker, J. *et al.* engineered a robust oscillator, named as relaxation oscillator, implemented in multi-copy plasmids in *E. coli* cells. This system is also based on the interactions between an activator and a repressor module. It is able to interactively exhibit oscillatory signals in periods of oscillations that vary depending on the context where cells are inserted into, e.g. by changes in inducer levels, temperature and media richness (**Figure 2.7-B**) (Stricker et al. 2008).

Next, an artificial gene-metabolic oscillator (the 'metabolator') was implemented in plasmids and into *E. coli*'s chromosome with transcription and metabolic integration characteristics of natural GRNs. This synthetic network uses the natural glycolytic metabolic flux, and external stimuli via carbon sources, to control oscillations through the native metabolite acetyl phosphate (Fung et al. 2005). Following the same logic of integrating the dynamics of a synthetic construct with the functioning of an endogenous cellular process, another study made use of the natural process of competitive protein degradation to create rapid and tunable coupled oscillators at both intracellular and colony levels (Prindle et al. 2014).

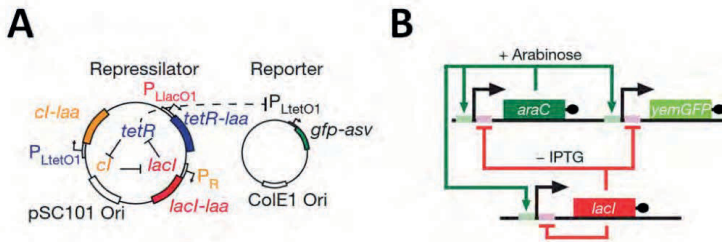


Figure 2.7: The network diagrams of the 'repressilator' designed by (Elowitz et al. 2000) then modified by (Potvin-Trottier et al. 2016), and of the relaxation oscillator designed by (Stricker et al. 2008), respectively. (A) The 'repressilator' system is a ring type oscillator with three genes in a negative feedback loop repressing one another. Oscillations in GFP levels are generated following the changes in the dynamics of the promoter (P_{LtetO1}), present in both repressilator and reporter systems. (B) In the relaxation oscillator, a repressor module (red) interacts with an activator module (green) creating an oscillatory signal (output) by the expression of green fluorescence protein (yellow). Adapted and reprinted with permission from (Stricker et al. 2008; Potvin-Trottier et al. 2016); respectively, Copyright 2016 & 2008, Macmillan Publishers Limited.

Another factor that influences the behavior of both natural and synthetic systems is the environment temperature. While natural networks have been shown to be robust to temperature changes (Tyson et al. 2001; Dunlap 1999), synthetic oscillators have been shown to have their dynamics significantly changed as a function of temperature (Stricker et al. 2008).

In all biochemical reactions, when temperature increases, the rate of these reactions increases, which makes most dynamical processes in cells faster (also known as ‘Arrhenius scaling’) (Segel 1975). As observed in the study of (Stricker et al. 2008), an increase in the environment temperature is followed by a drastic reduction in the oscillation period. Meanwhile, naturally occurring oscillatory systems, such as circadian clocks can maintain their period of oscillations almost constant under extreme conditions (Dunlap 1999; Hastings & Sweeney 1957). To create a synthetic oscillator that is robust temperature changes, Hussain, F. *et al.* engineered an oscillator with a structural temperature compensation module. For that, the authors followed the design concept presented in (Stricker et al. 2008) of an oscillator with dual-feedback reactions and added a special module, a temperature-sensitive lactose repressor mutant, which is incapable of repressing its target promoter at high temperatures (Hussain et al. 2014). By making the period lengths to increase as temperature increases, the circuit is able to compensate the decrease expected from the overall increased in reactions rates (**Figure 2.8-A**) (Hussain et al. 2014; Segel 1975).

Another promising means to make synthetic networks robust to environmental changes is by integrating genetic circuits with key elements of quorum-sensing machineries. In a recent study, based on the quorum-sensing mechanisms of the organisms *Vibrio fischeri* and *Bacillus Thuringensis*, Danino, T. *et al.* combined these key elements with elements of the ‘relaxation-oscillator’ (Stricker et al. 2008) to construct a ‘quorum-sensing’ oscillator that is able to couple genetic clocks from multiple individual cells of population, and generate stable synchronized oscillations at the colony level (Danino et al. 2010). Interestingly, the cells were able to sustain communication and stable synchronized oscillations even when subject to varying environmental conditions, provided that certain intracellular concentration of AHL (required for the activation of the genes and needed for both intracellular oscillations and intercellular coupling) was achieved in individual cells (**Figure 2.8-B**) (Danino et al. 2010).

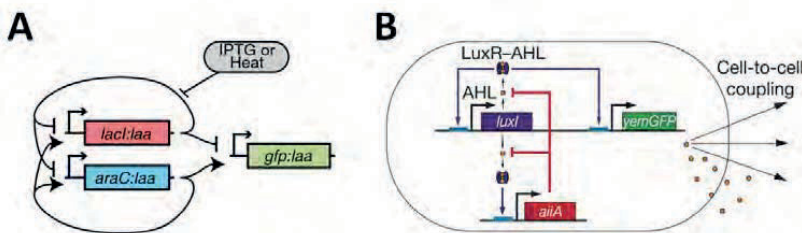


Figure 2.8: The network schemes of the ‘temperature-compensation oscillator’ designed by (Hussain et al. 2014), and of the ‘quorum-sensing oscillator’ engineered by (Danino et al. 2010), respectively. (A) The ‘temperature-compensation’ oscillator is based on the synthetic dual-feedback type of oscillator (Stricker et al. 2008), with both activator and repressor modules carrying copies of the $P_{lac-ara-1}$ promoter encoding, separately, the AraC activator, the wild-type or the temperature-sensitive LacI repressor, and GFP to monitor the dynamics of the system. All three proteins were tagged with an *ssrA* degradation tag to enhance their degradation rates. (B) The ‘quorum-sensing’ oscillator consists of a synthetic dual-feedback type of oscillator containing three genes controlled by three copies of the P_{luxI} promoter. In the first module, this promoter controls the expression of the enzyme LuxI, and a second copy of the promoter encodes for *yemGFP*. In the second module, the third copy of this promoter drives the expression of the repressor AiiA. The protein LuxI enzymatically produces a small molecule AHL (an activator of the P_{luxI} promoter) that diffuses through the cell membrane to outside, reaching other cells in the colony. Once in the cytoplasm, AHL activates the promoter. The repressor AiiA, indirectly,

negatively regulates the network by degrading the AHL molecules in the cytoplasm. Adapted and reprinted with permission from (Hussain et al. 2014; Danino et al. 2010); respectively, Copyright 2014, National Academy of Sciences, and Copyright 2010, Macmillan Publishers Limited.

Many studies have used the principles of stochasticity in biochemical reactions in their theoretical interpretations of genetic systems, to aid the rational design of synthetic circuits (Paulsson 2005a; Ribeiro 2007a). A recent study, for instance, has engineered several synthetic oscillators, from the noisy low-copy repressilator by (Elowitz et al. 2000), to test which internal features of the repressilator system are more/less related to generating the noisy behavior observed in its original version (Potvin-Trottier et al. 2016). Then, according to mathematical predictions, observations of the dynamics of multiple repressilator modifications revealed that a reduction in the copy-number of the reporter plasmid caused a reduction in the relative standard deviation (i.e. fluctuations) in the amplitude of the signal, making the repressilator system to oscillate more regularly with less noise. Interestingly, Potvin-Trottier, L. *et al.* further showed that this system was robust to varying environmental conditions, able to reduce error propagation across generations, could display macroscopic, population-scale oscillations with no cell-to-cell communication mechanism (Potvin-Trottier et al. 2016).

The functioning of these systems is highly determined by the kinetics of transcription. Namely, depending on the circuit topology, the kinetics of reactions involved in the activity of the genes, such as the binding/unbinding of regulator molecules and transcription factors to promoter sequences, may highly influence the circuit dynamics. Taking this into account, synthetic biologists have used the knowledge on the various classes of transcription regulatory species (e.g. DNA-binding proteins (**Figure 2.9-A**) (Tamsir et al. 2011) and DNA recombinases (**Figure 2.9-B**) (Bonnet et al. 2013) to design and construct complex transcriptional circuits that are able to perform advanced computational tasks, such as memory storage and logical operations. In particular, multiple digital genetic circuits (with two or more inputs returning one output) have been constructed *in vivo* following the Boolean logic in multiple logic gates configurations (e.g. AND, NOR, OR, or NOT gates). These systems were shown to be able to turn the expression of fluorescent proteins ‘ON’ and ‘OFF’ controlled by an output promoter, by following pre-defined logic gate topologies by using multiple transcription factor and regulation mechanisms of input-promoters (**Figure 2.9**) (Siuti et al. 2013; Wang et al. 2011; Moon et al. 2012; Anderson et al. 2007; Brophy & Voigt 2014).

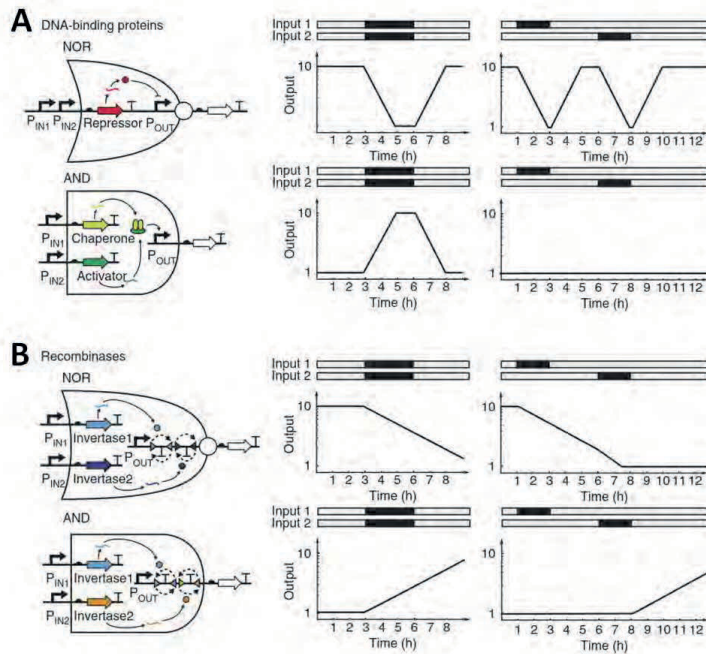


Figure 2.9: The design of transcriptional logic gates, based on different transcription regulation strategies, with two input promoters (P_{IN1} and P_{IN2}) and one output promoter (P_{OUT}). The graphs show how the gates respond to inputs introduced at the same time (center) or sequentially (right). In all panels, the ON state is assumed to generate tenfold higher response than the OFF state. (A) Top, NOR gate based on a repressor that binds DNA (Tamsir et al. 2011). The response curves are based on measured induction ($\tau_{1/2} \approx 36$ min) and relaxation ($\tau_{1/2} \approx 35$ min) half-lives (Lentini et al. 2013). Bottom, AND gate based on an activator that requires a second protein to be active (Moon et al. 2012). The responses are based on measured induction ($\tau_{1/2} \approx 36$ min) and approximate relaxation ($\tau_{1/2} \approx 35$ min) half-lives (Moon et al. 2012). (B) Top, NOR gate based on integrases that flip two terminators to turn off the output (Bonnet et al. 2013; Siuti et al. 2013). The responses are based on a rate of 1.8 h (Bonnet et al. 2013; Moon et al. 2011; Ham et al. 2006). Bottom, AND gate based on integrases (Bonnet et al. 2013). Adapted with permission from (Brophy & Voigt 2014); Copyright 2014, Nature America, Inc.

Nevertheless, synthetic networks can further make use of natural ‘post-transcriptional’ regulatory elements, such as the binding of non-coding RNA sequences to DNA, proteins or amino-acids (Isaacs et al. 2004; Isaacs 2012), to achieve a more complex integrated dynamical behavior. Due to their ability to control and sense various key signals inside cells, non-coding RNAs have become novel regulatory mechanisms of transcription and post-transcription processes (Isaacs et al. 2004; Mutalik et al. 2012) in synthetic constructs and applications using prokaryotic and eukaryotic organisms. Further, due to the reduced stress and burden caused to the cells in comparison with transcriptional circuits, when large amount of synthetic circuits are engineered in eukaryotic organisms, for instance, post-transcriptional regulation, through protein-protein interactions and allosteric systems is preferably chosen, allowing direct spatial-temporal regulation of protein functionality (Olson & Tabor 2012). Interestingly, in a recent study, based on transcription-activator-like effector (TALE) repressors competing with activators in an additional positive feedback loop connection for the same promoter

region, bi-stable switches were engineered using programmable pairs of DNA-binding domains to attain epigenetic bi-stability in cells across generations (Lebar et al. 2014).

In **Publications I and II**, we made use of the transcriptional circuit the low-copy repressilator (LCR) circuit, originally engineered by (Elowitz et al. 2000), to engineer the first functional, single-copy repressilator (SCR) circuit and study the effects of internal perturbations, e.g. networking coupling due to copy-number, and external perturbations, such as temperature shifts.

2.1.3 Coupling of Genetic Circuits

In live cells, the dynamics of GRNs are determined by the kinetics of their individual genes and their interaction with the complex intracellular context that is environment-dependent. The synthetic biology field has moved towards advancing the knowledge for the rational design of larger, more complex synthetic circuits with multiple components, modules, and devices, with pre-defined intra- and inter-connection behavior when implemented in live cells. In particular, researchers have tried to engineer synthetic oscillators that, once implemented in live cells, are able to achieve and sustain a robust behavior under environmental condition changes (Stricker et al. 2008), as in natural oscillators (Dunlap 1999).

One of the factors that may interfere in the dynamics of GRN's components is 'network coupling'. In that, due to the gene dosage inside cells, more than one copy of the component of a circuit may interact among them, thus affecting the functioning of one another (Brophy & Voigt 2014). The dosage of a particular gene, or an entire circuit, can be altered by either changing the copy-number of the plasmid, where the sequence is located or by increasing the number of motifs and transcription units within a single plasmid.

Several studies have evaluated the degree at which gene dosage can aid in achieving desired expression levels, while not affecting the cell's behavior (Molin et al. 1980; Jones et al. 2000; Chew & Tacon 1990; Togna et al. 1993). However, while these strategies can be beneficial in terms of increasing the expression levels of heterologous genes, depending on the number of plasmid types and their copy-numbers within the cells, those can also be not beneficial for the cells in a long-term. It has been reported that the increase in gene dosage, by increasing plasmid copy-number is associated with an increased burden effect on cell healthiness, as it decreases cell growth rate and promotes morphological changes (Chew & Tacon 1990; Togna et al. 1993). In this regard, natural cellular mechanisms, such as plasmid segregation (Ebersbach & Gerdes 2005; Million-Weaver & Camps 2014), become a problem for synthetic biologists. In that, a balanced point between heterologous gene expression and sustainable cell metabolism needs to be determined, prior to implementing high-copy number plasmids into cells (Bailey et al. 1986). Relevantly, for various synthetic oscillators implemented in *E. coli* cells, this balanced point has been reported to be different in terms of copy-number, i.e. from bacterial chromosome-integrated (Atkinson et al. 2003), to low-copy plasmid-borne (Elowitz et al. 2000), to medium-copy plasmid-borne (Stricker et al. 2008; Hussain et al. 2014; Danino et al. 2010). Finally, providing a support for researchers aiming at changing plasmid copy-number in *E. coli* studies, Kittleson, J.T. *et al.* have presented

an *in vivo* methodology to test the effects of gene and DNA cassette dosage on the kinetics of a gene when inserted in plasmids or integrated in *E. coli*'s chromosome (Kittleson et al. 2011).

Theoretical studies on gene dosage have further contributed to the understanding of gene dosage as a relevant parameter to fine-tune the expression of coupling elements, thus altering the circuit's behavior. Some studies have used modeling and mathematical techniques to aid synthetic biologists in assessing which properties and parameters should synthetic oscillatory systems have to attain a desired dynamics (Elowitz et al. 2000; Hasty et al. 2002; Garcia-Ojalvo et al. 2004; Danino et al. 2010; Stricker et al. 2008). For instance, in the case of the 'repressilator', small changes in copy-number can lead to a qualitative shift in the system dynamics among oscillatory and equilibrium (Mileyko et al. 2008). More recently, Brophy et al. have used a mathematical model of the 'relaxation' oscillator (Stricker et al. 2008) to test the effect of copy-number on the dynamics of the circuit. The results showed that, by increasing (or reducing) gene dosage of the original circuit by tenfold (as in a high-copy plasmid) and single-copy (as inserted in the chromosome), respectively, a substantial increase (or reduction) emerged in the amplitude of oscillations (Brophy & Voigt 2014).

In **Publication II**, we used the original sequence of the LCR circuit to construct the first functional SCR implemented in a full F-plasmid in *E. coli* cells. Then, we studied the effect of intracellular coupling between the genetic components in LCR and SCR cells on the temperature-dependence of the dynamics of the repressilator circuit (Elowitz et al. 2000). We investigated the effects of networking coupling on the dynamics of the repressilator (e.g. oscillation period length, signal amplitude, and circuit's functionality) when under temperature and inducer IPTG perturbations. For that, the kinetics of one of the three promoters of the network, i.e. the $P_{L-tetO1}$ promoter, was monitored by single cell confocal microscopy measurements of GFP levels over time. From the observed differences in period of oscillations, we found that when under temperature fluctuations, decreasing the copy-number of the repressilator circuit increased the mean and variability in oscillation periods, while keeping the dynamics of both circuits robust and sensitive to temperature shifts. Finally, when varying the concentrations of IPTG (the inducer of $P_{L-lacO1}$, one of the promoters of the repressilator circuit), we found that decreasing the copy-number provide cells higher robustness to external perturbation, suggesting that the oscillatory behavior of the LCR can be disrupted by only a few of the copies of the circuit being affected in the cells.

2.1.4 Extrinsic Factors to Gene Expression Affecting the Dynamics of Genetic Circuits

Genetic circuits provide cells with the ability to regulate key intracellular processes, such as transcription, translation, DNA replication, and cellular metabolites. Despite the possibility of a rational design of circuits with pre-defined dynamics, the availability of their key regulatory molecules, within the intracellular context, and their responses to environmental changes can heavily affect the functioning of such circuits (Brophy & Voigt 2014; Cardinale & Arkin 2012).

The changes in the intracellular context may be one of the main reasons why failures in the functioning of the first synthetic circuits were observed when implemented in live cells (Elowitz et al. 2000; Gardner et al. 2000).

For instance, when studying the synchronization of oscillations in a population of cells containing the ‘quorum-sensing’ oscillator, Danino, T. *et al.* have shown that the oscillator has its function hampered when the context in the environment changes such, that the concentration of AHL per cell (i.e. the molecule responsible for intra- and inter-connection among the circuits of the cell population) varies due to a change in the cells’ density (Danino et al. 2010). Further, synthetic circuits can also create a metabolic loading effect in the cells by monopolizing the host resources, which can eventually decrease cell growth rate and cell’s vitality, thus changing the intracellular concentrations of RNAP and ribosomes required for optimal circuit functionality (Dong et al. 1995).

However, Atkinson, M. R. *et al.* and Fung, E. *et al.* presented two successful implementations of synthetic constructs in *E. coli*’s intracellular context. They integrated a relaxation oscillator and a gene-metabolic oscillator into *E. coli*’s chromosome, respectively, and showed that these circuits were able to carry out desired functions by exploiting the host machinery and metabolism (Nandagopal & Elowitz 2011). Since the dynamics of such a system heavily depends on the amount of key molecular species available in each individual cell, being able to measure these single molecules, at the single-cell level, has become crucial when determining the sources of variability in the functioning of individual genes and circuits in live cells (Golding et al. 2005; So et al. 2011; Yu et al. 2006; Hensel et al. 2012).

The variability in the phenotype of genetically identical cells can be generated by both intrinsic and extrinsic noise sources (Elowitz et al. 2002; Taniguchi et al. 2010). Intrinsic noise is linked to the nature of the kinetics of transcription and translation kinetics, i.e. from the random encounters between molecules that act independently on individual genes with the same cell (Taniguchi et al. 2010; Yu et al. 2006; Mäkelä et al. 2017; Lloyd-price et al. 2016). Meanwhile, extrinsic noise can be generated from multiple factors related to how changes in the intracellular context of individual cells vary from cell-to-cell (due to e.g. changes in environmental conditions). These factors can, independently or combined, affect differently the expression of genes and the functioning of genetic circuits in individual cells of a population (**Figure 2.10**) (Brophy & Voigt 2014).

Cell-to-cell variability can further arise from extrinsic factors such as fluctuations in protein numbers and RNA numbers in individual cells (Paulsson 2005a; Paulsson 2005b). Further, they can emerge from variations in global regulators of transcription, e.g. σ -factors, transcription factors, ribosomes, and RNAP (Yang et al. 2014; Bakshi et al. 2012), due to differences in cell volume (Becskei et al. 2005; Newman et al. 2006; Stewart-ornstein et al. 2012) and cell cycle position (Zopf et al. 2013).

In recent years, other mechanisms, such as DNA replication, negative DNA supercoiling, DNA condensation by nucleoid proteins (Chong et al. 2014; Sanchez & Golding 2013; So et al. 2011) have been shown to contribute to cell-to-cell variability. In addition, changes in the spatial distribution, mobility (Stylianidou et al. 2015), and asymmetries in the partitioning of intracellular molecules during cell division (Huh & Paulsson 2011) associated to the nucleoid exclusion phenomenon (Winkler et al. 2010; Coquel et al. 2013; Gupta et al. 2014), have been shown additional contributors. Finally, changes in biophysical properties of cells, such as cytoplasm viscosity (Van Den Bogaart et al. 2007; Konopka et al. 2009; Mika et al. 2010), in the cellular

metabolism (Klumpp et al. 2009; Parry et al. 2014), in the expression of specific endogenous genes (**Figure 2.10-A**) (Cardinale et al. 2013), or in the growth media (**Figure 2.10-B**) (Moser et al. 2012), can strongly alter the dynamics of synthetic genetic circuits by affecting the activity of their components within the cells.

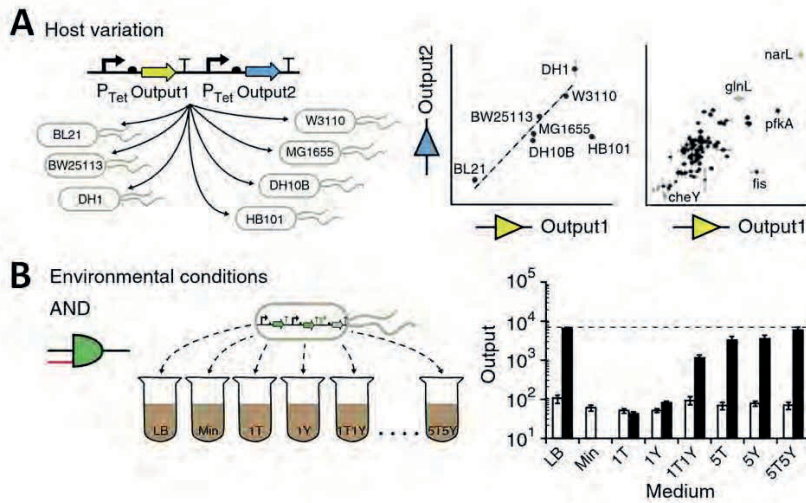


Figure 2.10: The effect of host processes on the dynamics of synthetic circuits in *E. coli* cells. (A) Multiple assays of bacterial transformation of a plasmid containing two reporter proteins, and the analysis of host interference on the two read-outs dynamics (shown as output1 and output2). The relation between the two outputs, measured by their expression ratios (shown in the plot on the right), by varying strains (with wild-type *E. coli* strain shown as MG1655, on the left plot), and by varying multiple KEIO collection knockouts (on the right plot) (Cardinale et al. 2013). (B) The effect of media changes on the *in vivo* performance of the AND logic gate designed by (Anderson et al. 2007). Cells containing the AND gate circuit were grown in various minimal media compositions in shake flask experiments using the lab strain *E. coli* DH10B, in the absence (white) and presence (black) of both inputs (inducers) of the logic gate. In the *x*-axis of the plot, minimal medium is shown as Min, and the media with input variations, i.e. by tryptone and/or yeast extract concentrations, are presented as #T and/or #Y, where # indicates their amounts supplemented in the media (in grams). Induced AND gates never grew on minimal media without any supplement and were active in varying supplemented media. Adapted and reprinted with permission from (Brophy & Voigt 2014) and (Moser et al. 2012); respectively, Copyright 2014, Nature America, Inc., Copyright 2012 American Chemical Society.

Furthermore, the kinetics of rate-limiting steps can be regulated by external factors such as environmental conditions and stress-related processes (Muthukrishnan et al. 2014; Browning & Busby 2004; Browning & Busby 2016). Since transcription kinetics, in most promoters of *E. coli*, is a σ^{70} -dependent process, the intracellular concentration and activity of σ^{70} are critical factors determining the expression profile of individual cells. When cells undergo specific stress conditions, other σ factor units are expressed and specific cellular responses triggered to define global changes in the dynamics of transcription of multiple genes and gene regulatory networks (Ishihama 2000). In particular, when cells are subject to sub-optimal temperatures, cells respond and adapt to it by precisely regulating the activity of specific genes that, e.g. are responsible for heat/cold shock responses (Kannan et al. 2008; Arsène et al. 2000).

The characterization of the most relevant extrinsic factors to gene expression that can generate cell-to-cell variability, along with the understanding of methods used by natural systems to mitigate them, can contribute to the knowledge of which design principles and techniques for rational engineering of synthetic circuits should be taken into account. For instance, the design of synthetic circuits with orthogonal, i.e. isolated expression systems not coupled with cellular regulation, can be used as an alternative to avoid extrinsic sources interference in the dynamics of genes or circuits dynamics (Rao 2012). However, this requires the complete isolation of gene expression pathways, or expression systems, to reduce any interaction between exogenous and endogenous reactions, which is a not trivial task. Finally, to aid in these designs, several theoretical studies have investigated the effects of internal and external perturbations on the dynamics of natural, artificial, and theoretical GRNs using several modeling and mathematical approaches (Ribeiro et al. 2006; Ribeiro 2007c; Kauffman 1969; Veliz-Cuba et al. 2015), as the examples described in the next chapter.

In **Publications IV**, we studied the effects of gene location on the temperature-dependence of the kinetics of rate-limiting steps in transcription initiation, an important regulatory mechanism of the components that compose synthetic circuit. In the efforts to characterize the intracellular context where genes and genetic circuits are inserted, the temperature-dependence of the cytoplasm viscosity, and correlation of this process with other relevant cellular processes, such as nucleoid exclusion phenomenon and polar localization of protein aggregates, were investigated in **Publication V**.

2.2 Methods in Synthetic Biology

2.2.1 Engineering Principles of Synthetic Biology and Advancements in Synthetic Biology Tools

The discovery of restriction enzymes in the 1970's paved the way for important advancements in synthetic biology, e.g. the ability to write artificial genetic information on the DNA of natural organisms (Roberts & Murray 1976). Techniques for molecular cloning and DNA recombination allowed the production of heterologous proteins, such as human insulin, in engineered *E. coli* (Goeddel et al. 1979). At the beginning of the 21st century, with the advent of DNA sequencing techniques, scientists became able to partially 'read' and 'write' the genetic information in the genome of small and higher-order organisms, e.g. humans (Fleischmann et al. 1995; Venter et al. 2001). During the recent years, both capabilities have been continuously improved. Synthetic biologists, in particular, have focused their efforts in making the 'writing' of genetic information of newly designed cells through a more rational process, i.e. with an engineering, controlled and predictable framework, due to many advancements in synthetic biology techniques (Endy 2005).

These frameworks commonly take into account *in silico* simulations and mathematical models of an expected network and computational tools to measure the best-fitted circuit that can perform this specific task *in vivo*.

In that, from the model's specifications, one could rationally, physically construct genetic systems and implement them in real organisms. For that to be rational, synthetic sequences, named as 'bio-parts', can be designed, constructed, and empirically assembled into genetic circuits, 'bio-devices', by following the pre-defined dynamics of each component-gene from *in silico* analysis. Next, for the validation of the construct, the dynamics of a genetic circuit is measured and compared with those from model predictions. In the end, a genetic circuit with specific 'components' (i.e. DNA, RNA, and proteins), 'topology' (e.g. protein-DNA, protein-protein, or RNA-DNA interactions), and components' interactions that are 'context-dependent' (e.g. RNAP, ribosomes and sigma factors), can be engineered (Beal et al. 2012; Haseltine & Arnold 2007).

Importantly, the knowledge from electrical engineering and components have aided synthetic biologists in the engineering of genetic circuits. Namely, the design principles for genetic engineering have been proposed by (Miyamoto et al. 2013) according to the similarity between electrical and genetic functions. The first principle consists of creating standard, specific, and modular components that can be assembled into devices and systems with compatibility and simplicity (Marchisio & Stelling 2009). This way, the design of a genetic circuit can be achieved from the characterization of their components' input-output relationships (Weiss et al. 2003). Second, genetic circuits can be engineered by assembling these characterized, modular component-genes with independent dynamics from most cellular processes (Sprinzak & Elowitz 2005). Finally, fluctuations in the components' behaviors, if existing, have to be also characterized in order to avoid changes in the dynamics of the assembled system. Relevantly, studies on synthetic logic gates in *E. coli*, in particular, are analogous to electrical circuits, and have contributed significantly to the understanding and construction of improved auto-regulation, switches, and genetic oscillators, or even more complex devices, such as circuits based on quorum-sensing communication between different *E. coli* cell types (Bulter et al. 2004).

Efforts in rationalizing this construction framework have provided new insights into how a synthetic construct should be designed, built, and its functioning validated in the cells. Following the design principles of synthetic biology, i.e. capability, functionality, compatibility, modularity and predictability, forward engineering workflows have been proposed in order to combine specific tools, from the varying and increasing pool of possible tools available, in a rational, precise, and efficient way, such as *SynBioSS* (Hill et al. 2008), *Clotho* (Xia et al. 2011), *Cello* (Nielsen et al. 2016), among others.

Besides following these principles and selecting the best-suited molecular biology methodologies, the structure and dynamics of each genetic part must be thought, independently, in terms of its: transcription unit, input signal, transcription regulators, and role in a network topology. As described in detail in Chapter 2, the transcription unit is the promoter region of a gene that controls the expression of the sequence encoding for a protein. Promoter sequences are usually extracted from natural systems (Nandagopal & Elowitz 2011), but can also be synthetically modified in the search for a better-fitted dynamics, when compared to the original (Lutz & Bujard 1997; Lutz et al. 2001; Jones et al. 2014). Then, to measure the dynamics of transcription units, researchers have engineered synthetic probes that, once inserted in the host strain *E. coli*, can detect and track

single copies of RNA in individual cells soon they are produced (Gasnier et al. 2013; Golding et al. 2005; Peabody & Lim 1996).

Following transcription and translation processes, regulatory proteins or non-coding mRNAs (Qi & Arkin 2014) are produced and, once functional, can send an activation/repression signals to one or more genes of the network (Lucks et al. 2011). This communication depends on the association of modulators (i.e. regulatory operator site) for this protein on the promoter region of other genes. In that, operator sites can be occupied by regulatory proteins (activators/repressors), influencing the rate of production of gene products, mRNA, and proteins. Additionally, transcription units can be affected by input signals from the environmental context. In the case of external signals, cells have complex cellular processes, such as membrane-bound receptors, signal transducers, and trans-membrane transporters, which are responsible for sensing, communicating, and physically incorporating such signals into the cells, respectively (Hoch 2000). Relevantly, Danino, T. *et al.* engineered an inter-cellular communication genetic clock with increased sensitivity and robustness to environmental perturbations, indicating that quorum sensing is an important property of cells when designing genetic systems (Danino et al. 2010).

Studies on the dynamics of synthetic circuits whose components, structure, and interactions with intracellular and environmental contexts are known *a priori* have effectively allowed researchers to have a better understanding of how natural systems work (Gardner et al. 2000; Elowitz et al. 2000; Çağatay et al. 2009). In 2000, Elowitz and Leibler built the first functional synthetic genetic clock (the ‘repressilator’, **Figure 2.7**) in individual *E. coli* cells, which periodically activates the expression of a green fluorescent protein whose levels inform on the circuit’s state (Elowitz et al. 2000). In the same year, Gardner *et al.* constructed the first synthetic genetic switch (the ‘toggle switch’) able to generate two-stable states based on the expression of two repressor proteins, one repressing the activity of the other at the transcription level. After that, complex regulatory networks, containing multiple genes connected by protein-DNA repression/activation reactions and positive/negative feedback loops, were built to perform logic-gates functions (see e.g. **Figure 2.9**) (Siuti et al. 2013; Chen et al. 2012). Finally, regarding the role of each component in the network topology, *in silico* and *in vivo* studies have investigated how regulatory networks with varying topologies can generate similar system dynamics, and why particular circuit-architectures are preferred by cells to accomplish specific cellular tasks (Çağatay et al. 2009).

2.2.2 DNA Assembly and Genome Engineering Techniques

More recently, advances in DNA assembly and *de novo* synthesis technologies (Gibson et al. 2009) have allowed researchers to construct the first functional artificial genome of *Mycoplasma genitalium* (Gibson et al. 2008). In that, multiple small fragments of DNA have been, separately, synthesized and assembled into larger pieces of DNA, then finally transferred into a genome-free host cell. The unification of these pieces into an artificial genome occurs when these fragments recombine in specific order inside the organism itself (Dymond et al. 2011; Daniel G. Gibson et al. 2010; Daniel G Gibson et al. 2010). Lastly, these advancements have

contributed to the efforts in engineering artificial organisms with a completely programmable genome, with pre-defined specifications and functions.

The DNA assembly methods are able to create genetic parts with standardized and modular specifications, which are necessary steps for the rational engineering of complex systems. During the past decade, several fast, cheap, reliable DNA assembly methods have been reported. A survey conducted by (Kahl & Endy 2013) revealed the most used methods for *de novo* DNA synthesis of small fragments and physical assembly of DNA fragments by recombination techniques. An expressive reduction in the costs of *de novo* DNA synthesis, along with the ability to design and predict the functioning of genetic constructs, has allowed synthetic biologists to move from traditional methodologies, such as molecular cloning using traditional restriction enzymes, to modern protocols and frameworks for the design of artificial systems in prokaryotes (Nielsen et al. 2016) and eukaryotes (Khalil et al. 2012). Meanwhile, advancements in the precision and cost of DNA recombination techniques have allowed scientists to assemble DNA fragments, e.g. small gene cassettes, into much large final fragments, or to implement modifications in parts of the genome (i.e. biochemical pathway or large cassettes) to create completely new synthetic genomes.

Up to date, many innovative recombination techniques have been developed, and widely implemented by researchers for DNA assembly tasks (Ellis et al. 2011; Engler et al. 2008; Beal et al. 2012). In particular, the ‘Gibson Assembly’ method introduced in 2009 by (Gibson et al. 2009), very suited to constructions of multiple (and long) DNA fragments that require simultaneous ligation forming large DNA products. This technique has proven its value when employed in projects aiming at synthesizing complete functional genomes, such as the *Mycoplasma genitalium* (Gibson et al. 2008) and; mouse mitochondria (Daniel G Gibson et al. 2010), and then implementing them into host cells, e.g. in *M. capricolum* recipient cells (Daniel G. Gibson et al. 2010).

The ‘Gibson Assembly’ methodology consists of a master mix reaction based on the ‘one-step isothermal reaction’ strategy connects multiple DNA fragments by using specific overlapping sequences (overhangs between two or more fragments to be ligated) that are incorporated into each fragment. After preparing the fragments with specific overhangs by polymerase chain reaction (PCR), these fragments sharing complementary regions are then ligated using a ‘one-step PCR reaction mix’ containing the enzymes T5 exonuclease and DNA polymerase (**Figure 2.11**).

When considering DNA assembly methods in an engineering framework, an important aspect of the method that needs to be taken into account is its scalability, i.e. its ability to rapidly and modularly produce standard genetic parts that can be easily implemented into multiple large constructions, with no need for further modifications in the future. Because ‘Gibson’ assembly demands the creation of overhangs between fragments to be ligated by PCR or *de novo* DNA synthesis, every time a new ligation is required, new fragment combinations with new overhangs need to be produced, a protocol that cannot be used in a rapid forward engineering manner. The forward engineering platforms require a great scalability of all methods involved in the framework so that the production of solutions can be automatized.

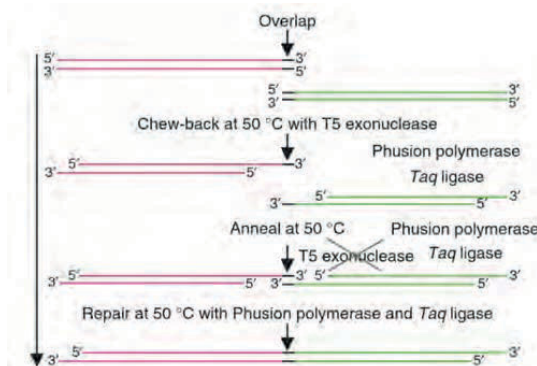


Figure 2.11: Schematics of Gibson Assembly one-step isothermal reaction. The PCR reaction occurs isothermally at 50°C with a reaction mix containing the DNA fragments to be ligated (harboring overhangs, i.e. points of ligation), the T5 exonuclease, the DNA polymerase (here referred to as ‘*Phusion* polymerase’, and the DNA ligase (referred to as ‘*Taq* ligase’). In this scheme, two target DNA fragments are treated (*de novo* synthesized or amplified by PCR) so as to present a complementary region (Overlap, in black), and new overhangs in the 5’ ends of the fragments are treated by the exonuclease. Next, the targets annealed in their 3’ overhangs, and are covalently sealed by the activity of both the DNA polymerase and DNA ligase filling the gaps of the final ligated DNA product. Adapted and reprinted with permission from (Gibson et al. 2009); Copyright 2009, Macmillan Publishers Limited.

To keep the assembly of standard parts fast and automated, synthetic biologists have most commonly used a hierarchical design strategy and DNA assembly methods that are based on the use of unique restriction enzymes. In that, depending on the hierarchy of the construction step and the function of each DNA piece to be ligated, the DNA fragment is flanked with a different restriction enzyme type. That is, if a particular fragment is in a lower hierarchical level (e.g. promoter sequence, RBS, open reading frames), then this first fragment receives enzyme X, while if another fragment is in a higher hierarchical level (e.g. individual gene), then it harbors an enzyme Y, and so on (Speer & Richard 2011). Although being beneficial in terms of modularity, compatibility, and scalability, these methods can be somewhat time-consuming due to their dependency on restriction-enzymes molecular cloning protocols. Then, most recently, a more modern, innovative, and high-throughput assembly methods, such as MoClo (Weber et al. 2011), and Gateway® cloning BP and LR reaction (Guye et al. 2013), have been proposed. They are based on the same principle of hierarchical design (with the use of non-conventional restriction enzymes) and the cornerstones of an alternative sub-cloning method called Golden Gate (Engler et al. 2008). In detail, Engler, C. *et al.* have used type II restriction sites to ligate multiple DNA targets to prove its high-throughput capability in transferring any DNA fragment of interest from a plasmid into an expression vector, without leaving any sequence fragments behind.

Type II restriction endonucleases can cleave the DNA during ligation at defined distances relative to their non-palindromic asymmetric recognition sites. Interestingly, after ligation, these enzymes can remove their restriction sites completely out of the final product, while keeping intact the desired sequences flanking the location of the restriction sites in the target fragments (Pingoud & Jeltsch 2001). This allows for sequential sub-cloning assembly set-ups of multiple sequences with no interruption of the process when physically obtaining these fragments from a library. However, when an entry clone fragment is to be ligated to an expression

vector by e.g. BsaI (a type II enzyme) and the activity of T4 DNA ligase, non-desired intermediate products can be also created and need to be removed from the final product. For that, since only the desired product (i.e. a restriction-site-free construct) is stable in the final solution, this can select during downstream selection treatments, such as the clones selected based on their kanamycin resistance and galactosidase activity (Engler et al. 2008).

Moreover, when instead of constructing synthetic systems, one is interested in modifying the information written in the chromosome of an organism, synthetic biologists have used genome editing techniques to reprogram key regulatory pathways in several organisms, including *E. coli*. Most of these studies focused on the metabolic engineering of organisms aiming to improve cellular processes, cellular behavior, enhanced conversion of substrates to products, and to develop improved artificial systems for novel products production in e.g. bacteria and yeast (Martin et al. 2003; Paddon et al. 2013). In the latter, in particular, recombination-mediated genetic engineering techniques, based on homologous recombination systems, such as the bacteriophage λ -Red system (Zhang et al. 1998), are commonly used for bacterial and yeast manipulation (Baba et al. 2006; Baudin et al. 1993). Interestingly, The re-writing of the stop codons to a different triplet sequence in the whole genome of *E. coli* was used to investigate the functionality of biological processes in recoded organisms (Lajoie et al. 2013). To date, maybe the most promising and revolutionary method for chromosome editing is named CRISPR. This method is based on the natural prokaryotic immune response to foreign genetic elements, and it has been reported to effectively and specifically make modifications in genes within multiple organisms, providing an important opportunity to advance genetic-dependent therapies and diagnosis (Sander & Joung 2014; Zhang et al. 2014).

In **Publications II and IV**, using computer-aided design, DNA assembly, and genetic engineering tools, two constructs were engineered aiming at performing specific tasks in *E. coli*, respectively, a single-copy F-plasmid repressilator (SCR), and a DNA fragment containing multiple repeats for individual mRNA detection. The latter was further integrated into *E. coli*'s chromosome and in a single-copy plasmid. The construction plan and validation of these constructs are presented in **section 4.1**.

2.3 Methods for RNA Detection and Quantification from Single-Cells Images

Biological systems, including monoclonal cell populations, exhibit diversity of phenotypes (Elowitz et al. 2002). One cause is noise in gene expression. Classical bulk measurements of gene expression, such as quantitative PCR (qPCR) and RNA sequencing (RNAseq), while of use to analyze the mean behavior of cell populations, are not able to capture the cell-to-cell variability (Snijder et al. 2009). To address this, quantitative measurement techniques have been developed for analyzing the dynamics of such systems at the single-cell, single-molecule level, using live cell fluorescence imaging methods (Golding & Cox 2004; Mäkelä 2016; Golding et al. 2005). The ability to detect fluorescent molecules in biological samples is inversely correlated

to the difficulty that a fluorescent measurement approach (i.e. fluorescent probe selected and fluorescent technology available) has in distinguishing two features located closer than approximately half of the wavelength of the illumination light, defined by Rayleigh's resolution limit (Born & Wolf 1999). This criterion, and other aspects of the illumination (e.g. the number of molecules occupying the same region) and the noise-to-signal ratio (affected, e.g., by cell background fluorescence), need to be considered when choosing the fluorescent probe and fluorescent microscopy technology to employ.

Fluorescent proteins have been isolated from organisms, such as GFP from the jellyfish *Aequorea victoria* (Prasher et al. 1992), and used as labels of target proteins in live cells (Tsien 1998), becoming useful tools in biotechnological applications. These applications follow the advances in molecular biology techniques and synthetic biology that allow, e.g. the engineering of DNA probes *de novo* and their implementation in live cells, such as *E. coli* (Yu et al. 2006; Golding et al. 2005). Recent advances in the properties of fluorescent proteins, allow for most of the visible spectrum of light to be sampled by unique fluorescent proteins (Shaner et al. 2004; Day & Davidson 2009), and new properties, such as photo-activation (turn protein On/Off) and photo-conversion by exciting them with lights of specific wavelengths, can now be used as applications in biological sciences (Day & Davidson 2009; Wu et al. 2011), including advanced imaging techniques, such as super-resolution microscopy (Huang et al. 2009). Furthermore, fluorescent probes (FP) have been structurally improved into modified structures that can last longer in live cells, as e.g. the mutated GFP fluorescent probe used for individual mRNA detection (Golding & Cox 2004).

The development of fluorescence microscope technologies towards attaining increased resolution and contrast in fluorescence images have followed the advances in fluorescent labeling techniques, allowing the progress of live-cell studies and measurements of biological processes in their native environments. Methodologies of super-resolution microscopy can attain higher image resolution (i.e. the ability to distinguish between two objects in an image) than the diffraction limit in fluorescent images (Neice 2010). Two of these methodologies, i.e. the stochastic optical reconstruction microscopy (STORM) and the photo-activation localization microscopy (PALM), can achieve up to 10 nm spatial resolution (out of the usually ~200-300 nm defined), even when a great number of target molecules are present in the system (Huang et al. 2009; Walter et al. 2008).

To attain higher contrast (i.e. the distinction between objects and image background), the methodologies, regardless if they can achieve super-resolution, must consider the physical reduction of noise that is unavoidably created by the measurement technique used during image acquisition. In the case of confocal microscopy technology, it requires the use of a pinhole (light blocker) that reduces the volume of light from out-of-focus regions (Pawley 2006). Other microscopy technologies, such as the total internal reflection fluorescence (TIRF) microscopy (Axelrod 1981) and the highly inclined and laminated optical sheet (HILO) microscopy (Konopka & Bednarek 2008) have proposed the increase of contrast by illumination of a thin region, in between the cover glass, the gel-pad, and the micro-slide where light penetrates (~150 nm from the surface in the case of TIRF), due to the inclination of the laser beam, thus avoiding light from out-of-focus regions. Differently from the above, in the conventional wide-field epi-illumination fluorescence microscope, due to its lack of pinhole (or

any exit slit) in the light path, the light emitted from out-of-focus regions is usually not easily distinguished from those of focused regions, which increases the background fluorescence in the images.

FPs can be used to probe RNA numbers in live cells by different means (Pitchiaya et al. 2014). Among the most relevant methods to achieve single RNA sensitivity, the most common methods are based on two strategies. First, one can use the direct binding of multiple FPs to structural motifs in the RNA (Golding & Cox 2004; Fusco et al. 2003), where an RNA fragment spontaneously forms stem-loop secondary structures that are chemically reactive to specific protein motifs (Keryer-Bibens et al. 2008). Second, one can use methods based on the sequence-based complementary hybridization of RNA labels carrying a fluorescent protein (e.g. FISH) with multiple specific RNA motifs (Raj et al. 2008; Levsky 2003). These methods allow measurements of cell-to-cell variability in endogenous RNAs, not possible by ensemble methods such as qPCR and RNAseq methods (Raj & Oudenaarden 2008). In the former, in particular, a DNA fragment containing multiple stem-loop RNA structures are designed in the DNA template of a target gene, so that multiple tagging proteins, the protein with affinity to bind to stem-loop RNA structure fused with a FP, can bind to the same nascent RNA molecule, allowing the detection and tracking of single mRNA in live cells (shown as a spot of light in fluorescence microscope images, e.g. **Figure 2.12-B**) (Fusco et al. 2003; Golding et al. 2005; Coulon et al. 2013; Larson et al. 2011; Hocine et al. 2013). To achieve a good contrast to differentiate spots of light (corresponding to the appearance of mRNA molecules) from the cell background fluorescence, the high production of fluorescence tagging proteins not yet bound to an mRNA molecule contributes to increasing cell background fluorescence, and thereby, entails the need for higher number of binding sites in the target RNA.

In this approach, to allow the detection of a target RNA as soon as they are produced in live cells, fused FPs (part of the reporter system) should be produced independently and extensively prior to the target promoter activation (part of the target system) (Golding et al. 2005; Mäkelä 2016). Relevantly, the tagging of MS2-GFP fused proteins provides the mRNA with a long lifetime and constant fluorescence beyond the observation time of the measurement (Tran et al. 2015). The two systems can be genetically designed and integrated into plasmids or *E. coli*'s chromosome, as the MS2-GFP RNA tagging system implemented in all studies of the publications of the thesis. In **Publications III and IV**, MS2-GFP tagging system was used to detect mRNA molecules as soon as they are produced by the target promoter $P_{lac-ara-1}$ (**Figure 2.12-B**) and $P_{lacO3O1}$, contributing for the better understanding of the cell-to-cell variability of gene expression in live *E. coli* cells.

The use of multiple binding sites also provides higher precision in detecting individual RNAs, by reducing the effect of undesired photo-bleaching during image acquisition. This, however, comes with several problems in the RNA quantification during image analysis: (i) the amount of fluorescence per tagged RNA fluctuates due to the incomplete and heterogeneous binding of tagging proteins to the stem-loop RNA structures (Fusco et al. 2003; Wu et al. 2012); (ii) the functioning of target RNAs can be affected by the binding of multiple tagging proteins (Wu et al. 2012), becoming e.g. safe from natural degradation (Golding & Cox 2004; Tran et al. 2015). One must account for these facts when designing new RNA tagging systems.

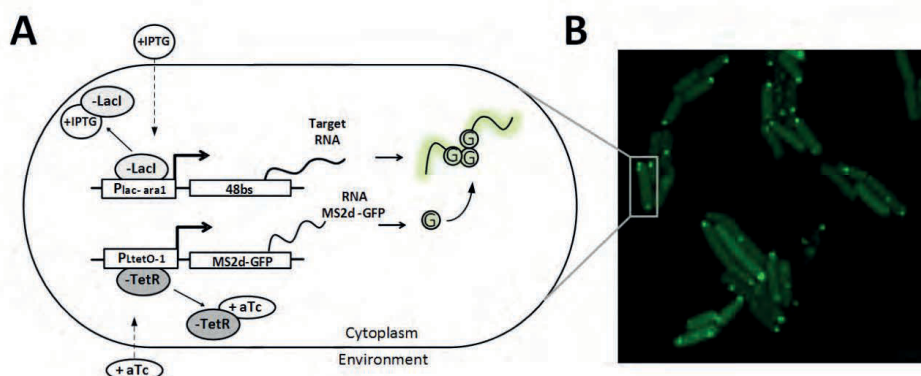


Figure 2.12: (A) Schematic representation of the genetic components present in the engineered MS2-GFP RNA tagging system used for live imaging of *E. coli* cells of the DH5 α -PRO strain. In detail, the RNA tagging system consists of a ‘target gene’ located in a single-copy F-plasmid and induced by IPTG, taken by cells from the media. When in the cytoplasm, IPTG neutralizes the overexpressed LacI repressors by forming inducer-repressor complexes (LacI-IPTG), allowing the promoter P_{lac-ara1} to express RNA molecules carrying an array of 48 MS2-binding sites. Meanwhile, the expression of MS2d-GFP tagging proteins is controlled by the activity of the P_{tetO-1} promoter regulated by the activity of TetR repressor (constitutively produced by *E. coli*’s chromosome) and the inducer anhydrotetracycline (aTc), also obtained externally. The reporter system is located in a multi-copy plasmid to allow the abundant production of MS2-GFP fused proteins prior to the activation of the target. In this system, once an RNA molecule is produced, multiple MS2d-GFP tagging proteins (here, referred to as ‘G’) rapidly binding to the RNA stem-loop structures, forming a visible bright spot under confocal microscope. (B) Example of confocal microscopy image of *E. coli* cells expressing the MS2-GFP proteins and target RNAs. The uniform background of the cell is visible due to abundant MS2-GFP production, due to the unbound MS2-GFP freely diffusing in the cell cytoplasm, while individual RNA molecules appear as fluorescent bright spots. Image in (B) obtained and adapted from **Publication III**.

Finally, the advent of fluorescence probing methodologies for studying the dynamics of cellular processes in live cells has made possible the use of such methods for many biological studies, including of the biophysical properties of cells. In **Publication V**, for example, we measured the mean displacement of MS2-GFP RNA complexes, considered as synthetic protein aggregates, within the cytoplasm in live *E. coli* cells, to evaluate the temperature-dependence of the biophysical properties of cell cytoplasm, in this case, the cytoplasmic viscosity. This study required selecting an adequate fluorescence probe and fluorescence microscopy technology, developing tailored image and data analysis software tools as well as inference methods of data-driven mathematical models (Häkkinen & Ribeiro 2015), and the use of microscope platforms mimicking natural environments for the cells while under observation (presented in **section 4.2**).

3 THEORETICAL BACKGROUND

This chapter gives an overview of the theoretical concepts of the modeling of biological systems. It also presents the modeling techniques and the simulation algorithm used in this thesis.

3.1 Stochastic Models of Gene Expression

3.1.1 Stochastic Simulation Algorithm (SSA) – Monte Carlo Simulation of the Chemical Master Equation (CME)

A chemical reaction is the instantaneous transformation process of types and quantities of chemical substances, named as ‘reactants’, into other substances different in properties, named as ‘products’. The transformation process of molecules when occurring in live organisms, or in cell-free systems (e.g. myTXTL[®]), mediated by a biocatalyst that can alter its rate and specificity, is called biochemical reaction. The collection of biochemical reactions that occur inside an organism is known as metabolism and is responsible for keeping cells alive at all times. The occurrence of these events inside cells is determined by the presence, position, and velocity of each individual molecule necessary for each biochemical reaction at each time moment. To understand and predict the behavior of these dynamical systems in live cells, multiple theoretical methods and models have been proposed, most of which are mathematical formalizations of the biological processes.

In deterministic methods, the models can provide qualitative information on the long-term behavior of the system with no room for random variation, informing on the dynamics of the system by population concentrations alone in a single trajectory through the state space. This, however, is not valid to describe the quantities and the timing of the interactions between species in biochemical reactions. On this regard, given that biochemical species appear in very low copy numbers in cells (Taniguchi et al. 2010) and that the time evolution of their interactions is stochastic (Gillespie 2007), realistic modeling approaches can provide a qualitative understanding of the dynamics of a system. In that, the probability distribution of the states that a system

occupies at a given time and the transition between these states as time evolves are described by the stochastic chemical kinetics (Gillespie 2007).

The stochastic approach considers the following definitions: namely, (i) N chemical species exist in the quantities $\{S_1, \dots, S_N\}$ in a well-stirred system; (ii) the system has constant volume (V) when in thermal equilibrium at constant a temperature; (iii) the biochemical species $\{S_1, \dots, S_N\}$ can interact through M chemical reactions $\{R_1, \dots, R_M\}$ resulting in a variation in their quantities; (iv) the number of molecules of species S_i in the system at a given time t is represented as $X_i(t)$; and finally (v) the state vector $X(t) = (X_1(t), \dots, X_N(t))$ can be estimated, given an initial condition assuming that the system was in state $X_{t_0} = X_0$ at the initial time t_0 (Gillespie 2007). Since in stochastic implementations systems are assumed to be well-stirred and to have constant volumes system, the position and velocities of individual molecules during all collisions and non-reactive collisions (i.e. not resulting in changes in species quantities) can be neglected (Daniel T. Gillespie 1977). Thus, only molecular collisions involved in reactive reactions, which change the numbers of chemical species, are considered (Gillespie 2007).

In the stochastic chemical kinetics, the changes in biochemical species amount caused by a reactive reaction (R_μ) are determined by two quantities. First, a state-change vector $v \equiv (v_{1\mu}, \dots, v_{N\mu})$, where $v_{1\mu}$ is the change in the S_i species population. Second, a propensity function (α_μ) that specifies the probability of reaction R_μ to happen in the next infinitesimal time interval $[t, t + dt)$ in the volume V , defined as follows (Gillespie 2007):

$$\text{Given } X(t) = x, \alpha_\mu(x) = c_\mu \times h(x) \quad 3.1$$

In Equation 3.1, $h(x)$ is a vector with all possible combinations of reactants of the reaction R_μ at a given time t . Derivations of the propensity function α_μ can be written depending on the underlying physics of the reactions that determine variations of $h(x)$. For instance, for a first order reaction ($S_1 \rightarrow \text{Product}$), the constant c_μ is the probability that any molecule of species S_1 will react in the next infinitesimal time interval dt . The higher the number of molecules of S_1 in the system (defined by the vector x_1), following the laws of probability, the higher is the probability that a particular molecule will react ($\alpha_\mu(x) = c_\mu \times x_1$). Meanwhile, for second order reactions ($S_1 + S_2 \rightarrow \text{Product}$), the propensity function is written following the law probability of two species to react given their numbers in the system at a particular time ($\alpha_\mu(x) = c_\mu \times x_1 \times x_2$) (Gillespie 1976).

The changes in species population defined by biochemical reactions, at any particular time, can be considered as a Markov process (satisfying the Markov property of a ‘memoryless’ system), and the propensity function α_μ solely dependent on its current state (i.e. independently from its history of past states) (Gillespie 2007). Then, assuming that $P(x, t \mid x_0, t_0)$ is the state probability of the complete system at time t , the probability of the system to be in state x at time $t + dt$ can be written as follows:

$$P(x, t + dt | x_0, t_0) = P(x, t | x_0, t_0) \times P(\text{no reaction in } [t, t + dt]) + \sum_{\mu=1}^M (x - v_{\mu}, t | x_0, t_0) \times P(\text{one reaction } R_{\mu} \text{ in } [t, t + dt]) \quad 3.2$$

The temporal evolution of the probability density function of $X(t)$ can be deterministically derived from equation 3.1 and from the laws of probability in 3.2 and described as a partial differential equation, i.e. the Chemical Master Equation (CME) (Gillespie 1976), defined as:

$$\frac{\partial P(x, t | x_0, t_0)}{\partial t} = \sum_{\mu=1}^M [\alpha_{\mu}(x - v_{\mu})P(x - v_{\mu}, t | x_0, t_0) - \alpha_{\mu}(x)P(x, t | x_0, t_0)] \quad 3.3$$

The CME in 3.3 then determines a set of coupled ordinary differential equations that inform on the probability that a system will have a particular molecular composition, i.e. a combination of biochemical species amounts, given an initial condition, at a particular time in the future. Then, depending on the different types of biochemical species, the amounts of each species in the system, and the number of reactions of the system, the space of all possible states created by the CME can be virtually impossible to be solved analytically. To avoid this problem and calculate the dynamics of these systems, the use of a Monte Carlo approach has been proposed to construct a set of simulated trajectories of random samples of the probability density function of $X(t)$ as a function of t , using a stochastic simulation algorithm (SSA) (Gillespie 1976; Gillespie 1977).

The simulation of individual trajectories using SSA is based on a new probability function $p(\tau, \mu | x, t)dt$ that, given the state of the system $X(t) = x$, defines the next reaction R_{μ} will occur in the infinitesimal time interval $[t + \tau, t + \tau + dt)$ (Gillespie 2007). This probability function $p(\tau, \mu | x, t)dt$ describes the joint probability density function of two random variables: (i) the time when the next reaction will occur (τ); and (ii) the index of the next reaction (μ) that will take place. An exact formulation of $p(\tau, \mu | x, t)dt$ can be derived from (3.1) and written as follows (Daniel T. Gillespie 1977; Gillespie 2007):

$$p(\tau, \mu | x, t) = \alpha_{\mu}(x) \exp(-\alpha_0(x)\tau), \quad 3.4$$

where,

$$\alpha_0(x) = \sum_{j=1}^M \alpha_j(x). \quad 3.5$$

From the mathematical basis of SSA described in equations (3.4) and (3.5) [69], τ is an exponential random variable with a mean of $1 / \alpha_0(x)$, and μ is a statistical independent random variable with point probabilities $\alpha_{\mu}(x) / \alpha_0(x)$. The calculation of τ and μ can be achieved by the original Monte Carlo algorithm named ‘direct method’ that accounts for the standard inversion generating method of Monte Carlo theory, following their uniform distributions (Gillespie 2007).

Finally, by running a simulation of the SSA in time with a probability density suited for shorter time and lower data storage consumption, users can create multiple trajectories of the system, i.e. multiple realizations in the

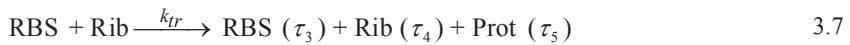
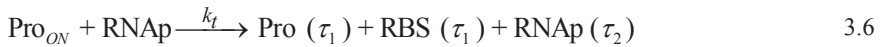
possible state-space described by the CME, thus yielding an exact estimation of the details of the dynamics of the system with certain precision.

3.1.2 Delayed SSA and Stochastic Model of Transcription

In general, stochastic models of gene expression assume that the events involved in gene expression are instantaneous reactions, following the occurrence cascade represented by the central dogma of molecular biology (namely, *gene activation* \rightarrow $[mRNA]$ \rightarrow $[protein]$). In that, the processes of transcription, translation, and of mRNA and protein degradation are considered stochastic reactions. Therefore, the time interval between occurrences per unit time of these processes follow exponential distributions and are modeled as independent Markov processes (Paulsson 2005a) using the SSA approach described (Gillespie 2007). However, other cellular processes involved in gene expression, such as transcription elongation (Greive & von Hippel 2005) and protein maturation (Megerle et al. 2008), can last for minutes before completion. Since this non-negligible lag time for completion cannot be simulated in the original SSA, a new version of the SSA approach was proposed so that ‘delayed reactions’ are introduced (Gibson & Bruck 2000). In that, although the knowledge of all possible steps of a process might not exist, models including delays can be supported by measurements of the overall duration of the process or following arbitrary distributions of such events. Overall, in a delayed reaction, when a reactant is consumed/removed from the system, the product is stored in a wait list and only released to take part into other reactions after a certain lag time.

The use of the delayed SSA has allowed the implementation of multiple delays for each reaction in delayed stochastic models (Roussel & Zhu 2006). Studies using delayed stochastic models have then investigated the potential effects of delays on the dynamics of transcription initiation (Ribeiro 2007a) and of small gene regulatory networks (Ribeiro 2010). In delayed stochastic models of gene expression, the following concepts must be taken into account in order to realistically model biological processes. First, time delays τ_i can be used to represent the lag time that mRNAs and proteins take to be produced from multiple steps in transcription and translation processes, respectively (Roussel & Zhu 2006; Ribeiro 2010; Bratsun et al. 2005; Ribeiro et al. 2006).

Second, since in live cells transcription and translation are independent from one another and the translation of proteins from single mRNA molecules is a noisy process (Zhu et al. 2007), they must be modeled as two independent stochastic reactions. Finally, since the mean and noise of mRNA and protein numbers are affected by their both production and degradation processes in live cells (Paulsson 2005b), the temporal dynamics of RNA and protein numbers must consider the following reactions (Paulsson 2005a):



In this model, the processes of transcription and translation of a constitutive gene (i.e. not subject to repression) are described in (3.6) and (3.7), and their stochastic rate constants represented as k_t and k_{tr} , respectively. The species involved in these reactions are: (i) the promoter region (Pro_{ON}) of the gene in the ‘ON’ state (i.e. available for RNAP to bind); (ii) RNA polymerases (RNAP); (iii) ribosomes (Rib); (iv) the RBS of a transcribed mRNA; and (v) a translated protein (Prot).

The time delay (τ_i) of appearance of a product from a multi-step reaction (e.g. 3.6) is a random variable determined by the single probability distribution that best approximates the overall dynamics of this complex reaction (Ribeiro et al. 2006). Namely, τ_1 determines the time taken for by RNAP to release the promoter region after the steps in transcription initiation, prior to elongation complex formation; τ_2 is the time taken by a ribosome to elongate an mRNA molecule depending on the length of the gene’s open reading frame (ORF); τ_3 represents the time that a RBS stays occupied by a ribosome; τ_4 is the time taken by a ribosome to complete the translation of a protein and get released from mRNA molecule; and finally, τ_5 is the time taken by proteins to become functional (i.e. protein folding). Finally, the processes of mRNAs and proteins degradation occur at d_{rna} and d_{prot} constant rates and are described in reactions (3.8) and (3.9), respectively. Importantly, these two reactions have been shown to exhibit exponential-like behavior in prokaryotes (Taniguchi et al. 2010; Bernstein et al. 2002), and to be crucial regulatory factors of the dynamics of small genetic circuits (Cameron & Collins 2014).

Inducible promoters, on the other hand, are transcribed only when induced, due to having their transcription activity regulated by transcription factors (TF), such as repressor (Rep) molecules, which can down-regulate the activity of the promoter by an occlusion mechanism (Garcia et al. 2010) that prevents RNAP from binding. One example of such a mechanism in prokaryotes is the *lac* operon regulation, where LacI dimers can occupy specific operator sites of the native *lac* promoter, which prevents transcription initiation (Oehler et al. 1994). Such repression mechanism can be modeled as follows (Ribeiro & Kauffman 2007):

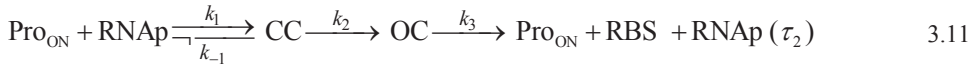


In reaction (3.10), a repressor (Rep) molecule binds to a promoter in ‘ON’ state (Pro_{ON}) at k_{rep} constant rate and forms a promoter-repressor complex that prevents the binding of RNAP to the promoter region and sets the promoter in ‘OFF’ state (Pro_{OFF}). The unbinding of the repressor from the promoter region sets the promoter to ON state and occurs at k_{unrep} constant rate.

The values set for the parameters of the above reactions can be obtained from empirical data of several studies that are based on the genome-wide heterogeneity in RNA numbers (Bernstein et al. 2002; Taniguchi et al. 2010), the kinetics of transcription initiation of individual genes (Lloyd-price et al. 2016), the dynamics of translation process at the single protein level (Mitarai et al. 2008; Bremer & Dennis 1996; Kennel & Riezman

1977), the properties of protein folding and activation in live cells (Cormack et al. 1996), and the behavior of natural and synthetic genetic circuits (Arkin et al. 1998).

Relevantly, the knowledge from the dynamics of gene expression in live cells by single-RNA measurements have allowed the multi-step transcription process of an active promoter P_{ON} , represented by τ_1 in reaction 3.6, to be alternatively model by a sequence of rate-limiting steps in transcription initiation activity, as follows (McClure 1980):



In equation (3.11) and **Figure 2.3**, the transcription rate k_t and the time delay τ_1 are replaced by the following reactions: (i) the CC formation after RNAP finds and binds to an active (free) promoter (Pro_{ON}) at the rate k_1 , and k_{-1} is the rate of reversibility of CC (Chamberlin 1974); (ii) the OC formation at the rate k_2 (Saecker et al. 2011; Chamberlin 1974); and (iii) the promoter clearance and elongation complex formation at the rate k_3 , after OCs undergo rapid cycles of abortive initiation until a productive initiation (Hsu 2002; Straney & Crothers 1987). In the end, an mRNA is synthesized, still represented by the time delay τ_2 for RNAP to complete elongation and release the mRNA. Relevantly, due their much shorter time-length when compared to the time taken for transcription initiation, when using this type of model simplification, both the rate constant k_3 and the time delay τ_2 can be considered negligible (Lloyd-price et al. 2016).

Finally, the set of reactions and possible derivations presented here can be used to model the dynamics of gene expression and regulatory networks, e.g. genetic switches and repressilator circuits, to support empirical studies in many ways, from engineering *in vitro* and *in vivo* implementations of these systems to generating hypothesis that can be empirically tested, to fitting distributions of empirical data of biological systems aiming at characterizing their underlying mechanisms. In **Publication III**, we used a model of transcription initiation (Häkkinen & Ribeiro 2015) that accounts for the active-inactive promoter model (Peccoud 1995) and the sequential model of transcription initiation (McClure 1985; Saecker et al. 2011), as the one described in equation 3.11, where a promoter can switch between being active and inactive (on/off) for transcription to initiate, depending on the binding and unbinding of regulatory molecules (Golding & Cox 2006; Lutz & Bujard 1997). Importantly, the “full commitment” of the system to produce an RNA molecule only occurs once an individual RNAP reaches the OC formation step, which, differently from the CC step (Kontur et al. 2008; Roe et al. 1985), makes highly unlikely that the system returns back to the previous states (Kontur et al. 2008). By fitting this stochastic model to the empirical data, we detected which rate-limiting steps during RNA production were most responsible for the observed differences in transcription dynamics between environmental conditions. In the end, the derived models were used to explore a possible cause for why the identified steps were identified as the main cause for behavior modifications with changing temperature.

In **Study IV**, based on models and parameter values derived from studies of transcription initiation, including genome-wide studies of variability in transcript counts (Taniguchi et al. 2010; Bernstein et al. 2002), transcription dynamics of *lac*-derivative promoters and mean RNAP intracellular concentrations (Lloyd-price et al. 2016; Kandavalli et al. 2016), and long-lasting super-coiled states driving stochastic bursts in bacterial genes (Chong et al. 2014). In that, assuming that the promoter is integrated in the chromosome, we performed 500 instances of 75 minutes-long simulations of a stochastic model of transcription initiation for differing rates of promoter escape from a supercoiled state, with high values corresponding to high temperatures (30 °C), as reported in (Chong et al. 2014). We extracted the expected mean relative RNA numbers in individual cells for every 15 minutes, as in the empirical data, and found that the differences in mean RNA production rate over time between cells with the chromosome-integrated promoter at critically low (10 °C) and at high (30 °C) can be explained by a positive supercoiling buildup on the DNA segment that eventually stops transcription initiation of chromosomal genes, when at critically low temperatures.

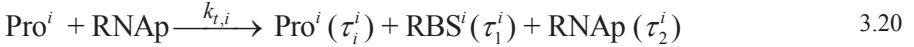
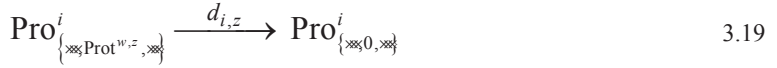
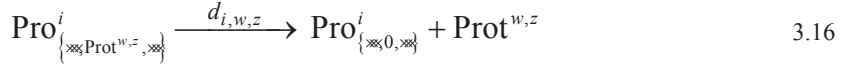
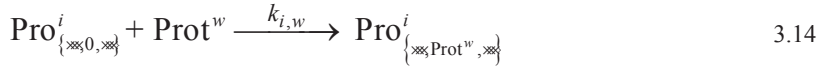
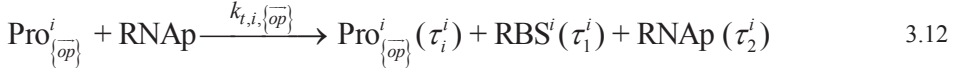
3.1.3 Delayed Stochastic Model of Small Genetic Networks

Several modeling strategies have been employed to investigate *in silico* the dynamics of both natural and abstract genetic circuits, as a means to obtain comprehensive understanding of the building blocks of biology and the principles, not possible through only experimental approaches, upon which new, simpler, and more general *in vivo* synthetic systems with pre-defined kinetics should be based (Hasty et al. 2001). Researchers must define which parameters and features are the most relevant for a particular modeling strategy. In the case of oscillatory systems, e.g. the ‘repressilator’ (**Figure 2.7**), to simulate the dynamics of this network, authors have presented a model strategy that accounts for the stochasticity of coupled reactions using a SSA (Daniel T Gillespie 1977) due to the stochastic nature of the reactions involved in the system.

Lipshtat, A. *et al.* used a stochastic model combining network structures and key biological features of this system to demonstrate, from *in silico* analysis of the dynamics of genetic toggle switch for a range of biologically relevant conditions, that an appropriate combination of the stochasticity of the interactions and the network structure is sufficient to give rise to bi-stability, even in the absence of cooperative binding of transcription factors to promoters (Lipshtat et al. 2006). However, the interaction of these regulatory molecules in the promoter region has been shown to be essential in determining the kinetics of transcription initiation (Mäkelä 2016; Golding & Cox 2004), thus the behavior of the gene networks where individual genes are present. On this regard, using a similar modeling strategy of gene expression and transcription regulation to the set of reactions described in the previous section, Ribeiro, A.S. *et al.* studied the dynamics of small gene regulatory networks by employing the delayed SSA approach and stochastic models of a couple of genes that can interact among them via the binding/unbinding activity of transcription factors produced by a gene to their respective operator sites located in the promoter region of another (Ribeiro 2010; Ribeiro et al. 2006).

Aside from the activity of regulatory molecules, many other parameters, such as promoter strength, mRNA and protein degradation rates, and gene copy-number, can be used as a means to alter the dynamics of small

genetic circuits (Brophy & Voigt 2014). Delayed stochastic models have been widely used to accurately describe the dynamics of varying small genetic circuits, e.g. genetic switches and oscillatory systems, as well as to predict the effects of perturbation in the intracellular and extracellular contexts (Rajala et al. 2010; Zhu et al. 2007; Ribeiro 2007a). For such studies, the general modeling strategy of small genetic circuits was proposed in (Ribeiro et al. 2006), including the following set of delayed reactions:



Similar to the set of reactions previously described, reaction (3.12) describes the transcription process, where $\text{Pro}_{\{\overline{op}\}}^i$ denotes the promoter of gene i in the system, and the array $\{\overline{op}\}$ is the set of operator sites and their activity states (i.e. whether bound or not by repressors), which thereby defines the state of the promoter $\{\overline{op}\}$ as either repressed $\{\dots, \text{Prot}^{w,z}, \dots\}$ or activate $\{\dots, 0, \dots\}$, respectively. The translation of the proteins of gene i from their individual mRNA molecules (RBS^i) is described in reaction (3.13).

Next, in reaction (3.14), structural variations of the repressor monomer (Prot^w) produced by gene w , such as dimers and tetramers (Rutkauskas et al. 2009), are represented as Prot^z , and their binding affinities to gene i are denoted as $k_{i,w,z}$. The general reaction of cooperative binding/unbinding of repressors to promoter operator sites of gene i ($\text{Pro}_{\{\dots, \text{Prot}^w, \dots\}}^i$) are represented in reactions (3.15) and (3.16), respectively. As such, the state of promoter resulted from the cooperative binding/unbinding reactions are denoted as $\{\dots, \text{Prot}^{w,z}, \dots\}$ and $\{\dots, 0, \dots\}$, respectively. Nevertheless, similarly to reactions (3.8 and 3.9), the degradation processes of

mRNAs (RBS^i) and proteins ($Prot^w$) are described in (3.17) and (3.18), respectively. Additionally, the degradation of repressors ($Prot^{w,z}$) can occur when bound to the promoter, as described in reaction (3.19), thus making the promoter available. Finally, reaction (3.20) represents the possibility for leaking events in transcription in each gene i (Ribeiro 2010).

Among the most common and widely studied types of small genetic circuits are the genetic toggle switch (R. Zhu et al. 2007; Atkinson et al. 2003; Roberts et al. 2011), and the repressilator and relaxation oscillator systems (**Figure 2.7**) (Stricker et al. 2008; Elowitz et al. 2000; Hilborn & Erwin 2008; McMillen et al. 2002; Pomerening et al. 2005; Ribeiro 2007a; Rajala et al. 2010). Regarding their mathematical methodologies, Elowitz and Leibler studied their repressilator system by stochastic models and SSA (Elowitz et al. 2000) and showed the importance of time delays in negative feedback reactions as the key regulatory parameters to control oscillation period lengths. Further, Zhu, R. *et al.* used the mean-field deterministic approach to theoretically study a repressilator model with time delays and showed that oscillation period lengths can be tuned by decreases in protein degradation (R. Zhu et al. 2007). Relevantly, (Tuttle et al. 2005) used a hybrid stochastic discrete-continuous algorithm to investigate in a system with no time delays which conditions of the repressilator dynamics sustainable oscillations can theoretically exist. Nevertheless, again by stochastic simulations, Loinger, A. *et al.* have shown that structural changes on the repressor proteins, e.g. their ability to cooperatively bind to operator sites, can differ the kinetics of oscillations (Loinger et al. 2007).

Other theoretical studies have performed simulations of stochastic models using the delayed SSA approach to investigate how internal and external perturbations in the dynamics of a network component affect the dynamics of the repressilator and toggle switch, namely, the period of oscillations and the existence of multiple steady-states, respectively, due to their interaction with intracellular context of the host strain, among others (Ribeiro et al. 2006; R. Zhu et al. 2007; Ribeiro 2007a; Samad et al. 2005). These studies can provide valuable insights into the underlying dynamics of genetic parts, units, and modules and how their kinetics can affect the behavior of more complex systems. This knowledge can aid the physical construction of their corresponding networks by predicting or avoiding possible problems, such as those caused by intracellular and intercellular communications (Hasty et al. 2002).

In **Publication II**, based on a model proposed by (R. Zhu et al. 2007) and parameter values used in (Elowitz et al. 2000), we implemented delayed stochastic models of the SCR and the LCR, and simulated 1000 instances (cells) of each. We extracted the functionality of each cell before and after perturbing the two repressilator system with the inducer of one of the promoters. Results showed that the LCR functionality is more drastically reduced after perturbation than in the case of SCR, according to the empirical results, suggesting that the SCR is more robust to environmental changes than the original LCR.

4 RESULTS: GENETIC CONSTRUCTS, MEASUREMENTS, AND ANALYSIS

This chapter presents a summary of the results of the studies conducted that are included in this thesis, along with the description of the new methods and tools developed to obtain these results. Namely, first, it is described the strategy used to engineer the two synthetic constructs used in the studies, followed by the report of the results of experimental validations of their composition and functionality. Next, it is described the microfluidics platform that was here designed and employed in multiple perturbation assays (i.e. various temperature and induction schemes). Finally, there is a brief description of the computational methods of image and data analysis here tailored to extract information from the measurements.

4.1 Genetic Constructs

4.1.1 Single-Copy Repressilator (SCR)

Most studies on how the dosage of genes and genetic circuits, i.e. their copy-number in the cells, affect their own dynamics have so far been based on theoretical models and predictions (Mileyko & Weitz 2010; Klappenbach et al. 2000; R. Zhu et al. 2007; Ribeiro 2007c; Ribeiro 2007b). To conduct an empirical analysis of the behavioral differences in the *in vivo* dynamics of the repressilator circuit due to copy-number differences, we engineered a F-plasmid SCR from the original Elowitz-Leibler LCR, using conventional molecular cloning protocols based on restriction enzymes.

In particular, the functional repressilator cassette from the original plasmid (pZS1-ITlrLLtCL) was amplified by PCR, and then inserted into a single-copy F-plasmid (pTRUEBLUEScript), known to exist as a single-copy and with high stability inside the cells (Gordon et al. 1997; Ogura & Hiraga 1983). The sequences to be ligated, i.e. the repressilator cassette from pZS1-ITlrLLtCL and the fragment from the pTB-BAC2 F-plasmid backbone containing the single-copy origin of replication and the chloramphenicol resistance were amplified by PCR

with *Sma*I flanking regions. The construction plan of the SCR (pBAC2-ITlrLLtCL) along with amplification and ligation steps of these DNA fragments are presented in **Publication II**. *E. coli* Δ lacI MC4100 strain cells containing no plasmids and cells of the same strain containing both the original repressilator plasmid and the reporter plasmid (pZE21-GFPaav) were generously provided by M. Elowitz (Elowitz et al. 2000) (California Institute of Technology, USA). Plasmid-free *E. coli* Δ lacI MC4100 cells were transformed with the newly constructed SCR and the reporter plasmid. The oscillatory dynamics was confirmed by qPCR and live cell microscope imaging.

In **Publication II**, to validate the construction of the SCR system, first, gel assays were performed confirming the sizes of PCR fragments and plasmids in the various steps of the construction pipeline. Second, the ligation regions between the repressilator cassette and the single-copy plasmid were confirmed by sequencing from the final SCR construct. To test the expression levels of LacI, TetR, and CI present in the SCR, and compare them with those from the LCR, qPCR measurements were performed using 16S ribosomal RNA (rRNA) expression levels as a reference (Livak & Schmittgen 2001). Green fluorescence protein (GFP) expression levels from the SCR were assessed by inspection under fluorescence microscopy. These measurements showed that the fluorescent signal produced by cells over time is similar to that of the original LCR. Finally, the functionality, sensitivity, and robustness of the SCR were measured at optimal and sub-optimal temperatures and then compared with those of the LCR.

4.1.2 Synthetic Fluorescent Probe Integrated in *E. coli*'s Chromosome

The MS2 tagging system for mRNA detection in live cells uses an RNA coding sequence of multiple binding sites as a 'target gene' for the MS2 viral coat protein fused with a green fluorescent protein (MS2-GFP), produced by a 'reporter gene' (Peabody 1993; Golding et al. 2005). Using confocal microscopy imaging, when MS2-GFP proteins bind to the MS2 binding sites of the target RNA, this mRNA appears as a fluorescent 'spot' (usually in less than 1 minute following the appearance of the mRNA in the cell), provided sufficient MS2-GFP proteins in the cell (**Figure 2.12-B**). This system allows temporal (and spatial) analyses of mRNA production in individual cells. Such information has been used to characterize the dynamics of transcription initiation of the promoter of interest. By sensing the integer-valued RNA numbers over time, intervals between consecutive RNA production events can be empirically measured in live individual *E. coli* cells (Muthukrishnan 2014; Mäkelä 2016; Häkkinen, Tran, et al. 2013; Lloyd-price et al. 2016; Mäkelä et al. 2017). In addition, distributions of RNA numbers in individual cells can be measured, and their mean (M) and coefficient of variation (CV) calculated (Goncalves et al. 2018; Kandavalli et al. 2016). The corresponding protein numbers can also be assessed since, the prior to the MS2 binding sites, the target RNA codes for a red fluorescent protein (Golding et al. 2005).

To enhance its efficiency, in **Study IV**, a new fluorescent MS2 tagging system was designed, constructed, and integrated into a single-copy plasmid, and then into the chromosome of *E. coli*. For this, first, a DNA fragment sequence was designed and synthesized *de novo* to contain the promoter $P_{lacO3O1}$ controlling the expression of a 'target gene' consisting of a red fluorescent protein (mCherry) followed by an RNA sequence with 48 binding

sites of the MS2 viral coat protein (mCherry-48BSxMS2) (**Figure 4.1**). In addition, in this new design, the RNA tagging system harbors unique restriction enzymes (pre-defined ‘cutting points’) to allow future manipulations, such as replacing the promoter or RBS sequences, as well as the integration/replacement of sequences in regions between arrays of 12 MS2 binding sites (out of the 48), and the reduction of the number of binding sites. This DNA fragment was inserted into a single-copy F-plasmid and the construct was validated by both sequencing and gel assays (GenScript, USA).

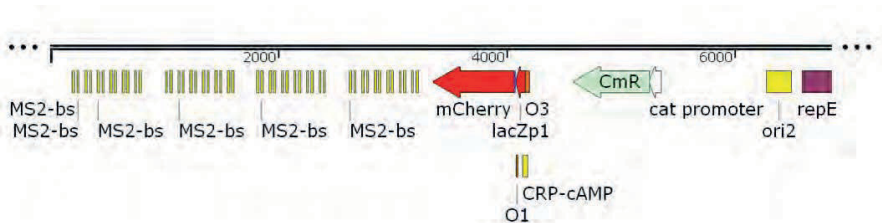


Figure 4.1: Linear representation of the plasmid-borne mRNA tagging system engineered using *de novo* DNA synthesis and GenEZ™ molecular cloning. The single-RNA, single-protein fluorescent probe ‘mCherry-48BSxMS2’ under the control of $P_{lacO3O1}$ promoter in a single-copy plasmid. The $P_{lacO3O1}$ promoter region, containing the O3, O1 operators, and cAMP-CRP binding site, is followed by a coding region for mCherry and a DNA sequence of multiple binding sites for MS2-GFP proteins, which were independently constructed by *de novo* DNA synthesis, inserted into multi-copy plasmids, and then cloned together into a single-copy F-plasmid (pBELO) using GenEZ™ molecular cloning, until forming the final sequence of the ‘target gene’ (11686bp) (GeneScript, USA). The chloramphenicol resistant marker (*CmR*) and the single-copy origin of replication (*ori2*) of the plasmid backbone (pBELO) are shown at their exact location in the final product. This scheme was adapted from the construction history generated by SnapGene® 1.5.2).

Second, the newly designed ‘target gene’ (**Figure 4.1**) was used as DNA template to create an integration cassette (mCherry-48BSxMS2) to be inserted in the genomic *lac* locus of *E. coli* strain BW25993 by multi-staged Red/ET[®] mediated recombination strategy (Zhang et al. 1998). The DNA cassette, part of the developed RNA tagging system, consists of the ORF of mCherry followed by 48 repeats of a 19bp long binding motive (MS2-BS) interspaced by random DNA sequences of about 20-200bp. Due to the lack of a promoter sequence in the DNA cassette, the transcription unit of the *lac* locus, i.e. the native *lac*-promoter (*lacZp1*), was kept intact (**Figure 4.3**). The multiple steps for DNA cassette construction, the integration of this sequence into *E. coli*’s chromosome, and the final selection of the best-clone were validated by sequencing and gel assays (**Figure 4.2**) (GeneBridges, Germany).

Before the integration of the DNA cassette into *E. coli*’s chromosome, a modification in the original cassette was performed to reduce the probability that MS2 repeat sequences are lost during Red/ET[®] recombination in future steps. Namely, to avoid undesired deletions during homologous recombination, especially in this sensitive region, a selectable marker gene was inserted in the middle of the 48 repeats sequence. In that, first, a bilaterally FRT-flanked KanR cassette (with resistance for the antibiotic kanamycin) was created by PCR and introduced into the DNA cassette by a Red/ET-mediated modification, in between the 24th and 25th MS2-bs repeats of the 48BSx sequence. For that, *E. coli* DH5a cells were transformed with the original pBELO-(*lacZ*-mCherry-MS2-48bs) single-copy plasmid and pRed/ET plasmid for the Red/ET-mediated integration of the

FRT-KanR-FRT cassette in the middle of MS2 repeats, creating a more stable intermediate cassette ‘mCherry-48BSx-MS2-KanR’ (Figure 4.2).

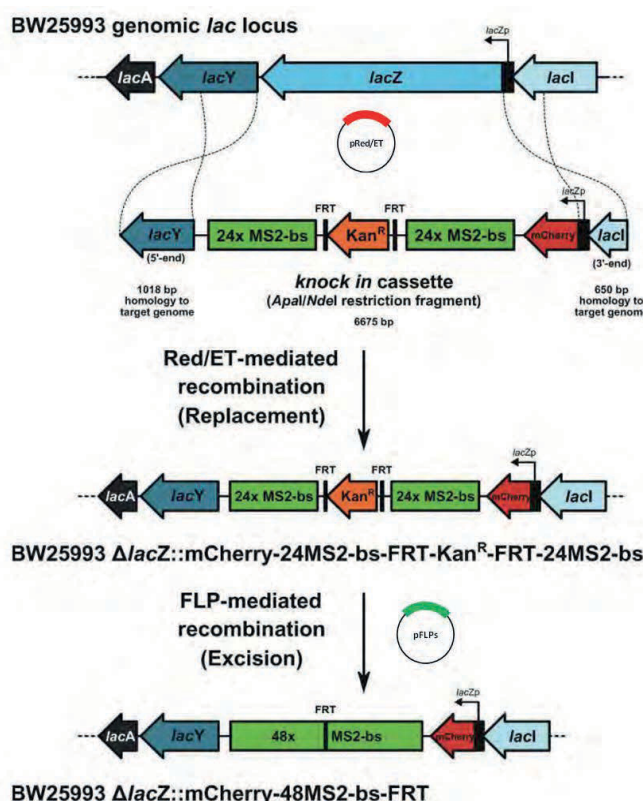


Figure 4.2: Representation of the construction of the knock-in (KI) cassette (mCherry-48BSxMS2) into *E. coli*'s chromosome by multi-staged Red/ET[®] mediated recombination. The DNA fragment mCherry-Ms2-bs-FRT-KanR-FRT knock-in cassette (6675bp, featuring two long terminal homology arms, 1018bp, and 650bp) was generated by NdeI/ApaI double restriction digestion, to guide the cassette insertion into the desired *lac* locus of the target genome. Next, the homology arms of the KI cassette recombine with the corresponding regions of the *lac* locus of *E. coli* BW25993 by Red/ET-mediated recombination. Second, successfully recombined clones containing Kanamycin-resistance are obtained. Finally, cells are transformed with specific plasmids for FLP-mediated recombinase expression to perform the excision of the region in between FRT sites, thus removing kanamycin resistance in the final clone. Scheme adapted from the construction history generated by SnapGene[®] 1.5.2).

After adding the *kanR* resistance gene into the 48BS region, a second strategy was performed to improve the stability of the DNA cassette, prior to the recombination step for chromosomal integration. Given that repetitive regions in linear DNA fragments gain stability when flanked by non-repetitive sequences, the use of long homology regions of the chromosomal locus in the final knock-in cassette can increase the stability of this fragment for genomic integration (Gray & Honigberg 2001). For that, first, the desired genomic integration site, i.e. the *lac* locus of *E. coli* BW25993's genome, was entirely sub-cloned to an intermediate R6K-ZeoR vector backbone (7340bp in size, from *lacI*, *lacZ*, *lacY*, to *lacA* genes). Second, this construct was used as

DNA template to create long homologous regions (~50bp) from the *lacZ* promoter (*lacZp1*) and *lacI* gene in one side, and *lacA-lacY* stretch in the other side, to guide the integration of the knock-in (KI) cassette into *E. coli*'s chromosome. Both homologous sequences contain *HindIII/MfeI* restriction fragments for the subsequent one-step ligation with the previously prepared DNA cassette 'mCherry-48BSx-MS2-KanR', in a so-called 'linear-linear' recombination, producing the circular knock-in (KI) cassette with the homologous regions for later genomic integration. After verifying the sequence of the circular KI cassette by a set of control PCRs and sequencing, the final linear knock-in cassette is generated by *NdeI/ApaI* double digestion. In this step, large homologous arms were left in both sides of the final linear KI cassette (~1kb at the 5'-end of the *lacY* gene, and ~650bp at the 3'-end of the *lacI* gene).

Next, BW25993 cells were transformed with the KI cassette along with the pRed/ET plasmid. The interactions between KI cassette and genomic loci are mediated by the Red/ET system using the 50bp terminal homologies that the DNA fragments have in common. In that, the *lac* locus in the genome is replaced by the knock-in cassette (**Figure 4.2**). Successfully recombined BW25993 clones were subsequently selected on LB agar plates supplemented with Kanamycin, along with X-Gal and IPTG following the galactosidase blue/white screening (Ullmann et al. 1967).

Finally, a verified knock-in clone was transformed with an FLP-recombinase expression construct for site-directed recombination reaction with the FRT sequences present in the clone genome, to eliminate the kanamycin resistance gene from the genome (Sadowski et al. 1995). During this last step, sequencing results of the multiple MS2-bs revealed that a couple of repeats linked with the removed *KanR* cassette were lost. Two massive deletions were detected: one deletion of 109bp length (comprising 2 MS2-bs), and a second deletion of 853bp length (comprising 13 MS2-bs) (**Figure 4.3**).

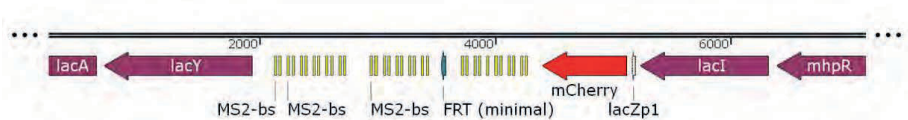


Figure 4.3: Sequence of BW25993 Δ *lacZ*::mCherry-MS2-33bs. To make a marker-less *E. coli* cell, i.e. with no Kanamycin resistance, the gene *kanR* was eliminated by FLP-recombinase activity, leaving a single FRT-site in the middle of the MS2-bs repeat region (green). After this procedure, the selected clone harbored 15 repeats less than the former 48 expected by unpredictable but expectable rearrangement events (two sites of deletions, presented in pink boxes). The final clone carries 33 MS2 binding sites (yellow boxes) in total. This scheme was adapted from the construction history generated by SnapGene® 1.5.2).

The two probes, one in a single-copy plasmid and another in the chromosome, along with the activity of the reporter gene, a multi-copy plasmid carrying the $P_{L-tetO1}$ promoter controlling the expression of the fused fluorescent protein 'MS2-GFP', can inform on the temporal and spatial dynamics of RNA molecules at the single cell level. The empirical and computational single-cell approaches used for such analysis in **Publications III, IV and V** are presented in **section 4.3**.

4.2 Microfluidics Platform for Perturbation Assays

A thermo-chiller chamber equipped with a thermos-chiller device (ranging from 5°C to 50°C \pm 0.2°C) and a micro-perfusion pump was engineered to assist the studies described in all publications of this thesis. For instance, in **Publications III and IV**, we studied the temperature-dependence of the kinetics of gene expression of multiple promoters under various induction schemes at the single-cell, single-molecule level, by observing mRNA molecules in live cells using the MS2-GFP-RNA tagging system and a tailored microfluidics platform able to keep temperature constant assays under time-lapse microscopy.

Furthermore, in **Publications I and II**, we also studied the temperature-dependence and the effects of external perturbations to the original and the single-copy version of the Elowitz-Leibler repressilator (Elowitz et al. 2000), by respectively altering the environment temperature and adding inducers in the growth media during cell growth. In the former, we expected that changes in temperature would cause variations in the oscillatory behavior of the circuit, i.e. in the stability of period of oscillations, due to changes in the rates of biochemical reactions from the Arrhenius scaling (Segel 1975), as observed from the drastic reduction in the oscillation period in (Stricker et al. 2008). In the latter, we expected that, when the promoter $P_{L-lacOI}$ is induced by IPTG, cells presenting oscillatory behavior have their GFP signal disrupted (close-to-null) in LCR cells containing more than one copy of the repressilator. Namely, we hypothesized that a failure in the functioning of one of the copies would hamper the activity of the others due to the existence of network coupling between the multiple copies of the circuit in individual cells. In agreement, we showed that the SCR strain responds more robustly to weak perturbations than the LCR strain, suggesting that a system with a single copy of a functional repressilator is more robust to transient, weak, external perturbations than a multi-copy one.

Also, a microfluidic platform containing a thermal chamber (**Figure 4.4**) was used to assist the studies described in all publications of this thesis to ensure a desired temperature of the environment and the continuous flow of fresh media supplemented with desired chemicals during microscope image acquisition. As shown in **Figure 4.4-C**, while under the microscope, cells are kept in an agarose gel-pad of fresh growth media, located between the micro-aqueduct slide and a coverslip inside the optical cavity of the chamber, allowing a uniform and rapid exchange of media with the cell population due to the laminar flow of fluids in this microenvironment. Importantly, while the gel-pad contains the nutrients for cell growth and inducers for the activity of both target and reporter genes, additional fresh media, inducers and, particularly, chemicals related to perturbation assays can be pumped into the optical cavity, throughout the duration of the experiment, by a micro-perfusion peristaltic pump, allowing steady-state growth for several hours under the microscope.

In **Publication I**, the set-up (**Figure 4.4**) was used to maintain the environment temperature (28°C, 30°C, 33°C, or 37°C) of cells containing the original repressilator system, and to keep the temperature range (24°C, 27°C, 30°C, 33°C, or 37°C) used for the studies of cells containing the genetic switch. Meanwhile, in **Publication III**, the thermal chamber was used to keep the temperature of multiple experiments constant (24°C, 27°C, 30°C, 34°C, or 37°C), while providing cells with fresh media containing the appropriate concentrations of the inducer

of the reporter and target promoters. This approach allowed the study of the temperature-dependence of the kinetics of transcription initiation of the promoter $P_{lac-ara-1}$ in varying induction schemes. Relevantly, in **Publication II**, using the thermal chamber equipped with a peristaltic pump (**Figure 4.4**), pre-warmed fresh M63 media containing 50 μM or 1000 μM IPTG (the inducer of the promoter $P_{L-lacO1}$, a component of the repressilator circuit) was continuously pumped to the cells after 2-3 hours of microscope image acquisition (to allow at least one oscillation cycle). This set-up was used to test for the response of both LCR and SCR systems to external perturbations, while keeping cells at the optimal temperature for repressilator functionality (30°C).

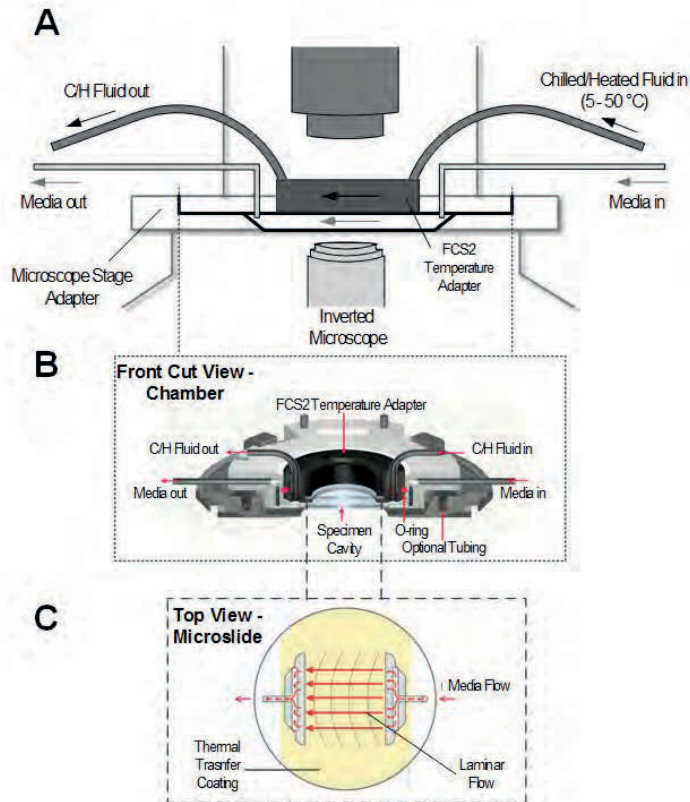


Figure 4.4: The schematic illustration of the thermal-microfluidic system. (A) The representation of both CFCS2 microfluidics and the temperature control system (Biopetech, CFCS2, Pennsylvania, USA) mounted on the stage of the inverted microscope for the culturing of cells under microscope observation. In this system, to control the temperature of the system (5°C to 50°C \pm 0.2°C), i.e. the temperature of the metal chamber and the optical cavity (where cells are placed), a thermo-chiller device (not shown) is connected to two inlets and outlets of the CFCS2 chamber for the flow of heat/chilled fluids. Further, to continually perfuse cells with fresh media and chemicals required for cell growth, a micro-perfusion pump (not shown) is connected to one inlet and one outlet of the CFCS2. (B) The illustration of the front-cut-view of the cooled FCS2 adapter (CFCS2) showing the optical cavity where cells are placed. In addition, to increase/decrease the temperature of both the metal chamber and an optical cavity, an independent tubing system facilitates the circulation of a heat/chilled fluid. (C) Finally, the representation of the top view of the micro-aqueduct slide that is placed inside the optical cavity, and of the laminar flow of fluids allowing a uniform and rapid exchange of media and cell population in the gel pad. Adapted and reprinted with permission from (Goncalves et al. 2018); Copyright 2018, IOP Publishing Ltd.

Finally, a study on the effect of osmotic stress on cells' morphology reported that a rapid osmotic upshift can be reached by increasing the concentration of sodium chloride (NaCl) in the media, thus causing cytoplasm plasmolysis of *E. coli* cells (Van Den Bogaart et al. 2007). From this, in **Publication V**, the same set-up was used to expose cells to osmotic stress during time-lapse microscopy, at 37°C. In that, the growth media was supplemented with 300mM of NaCl, approximately $\sim 0.68\text{Osm}$ (Konopka et al. 2009), and pumped to cells into the thermal chamber for 1 hour. In these cells, since high osmolality levels ($> 0.15\text{Osm}$) cause the diffusion coefficient of GFP to be heavily reduced, we combined the ability to track of synthetic protein aggregates (MS2-GFP system) in individual cells with the fact that osmotic stress can change the viscosity of cell cytoplasm, to validate the changes observed in the dynamics of nucleoid occlusion phenomenon due to an increase in cellular cytoplasmic viscosity, when cells are subject to critically-low temperatures.

4.3 Live Cell Imaging and Data Analysis

To measure the mean fluorescence intensity inside cells and to estimate the number of RNA molecules in each cell at a given time, from either a single population or multiple time-lapse fluorescence images, we made use of tailored image and signal processing tools. The implementation of these tools and the results extracted from the measurements are presented in the following sections.

4.3.1 Dynamics of Genetic Oscillators

In **Publications I and II**, from time-lapse images, we used a tailored semi-automatic method for cell segmentation and principal component analysis (PCA) to obtain dimensions and orientations of the cells from fluorescence images of cells containing either the LCR or the SCR systems. Then, the total fluorescence intensity in each cell was extracted to compute the mean pixel intensity of each cell over time. The same analysis pipeline was implemented at each time moment to study cells containing the CI-Cro genetic switch (**Publication I**).

In **Publications I and II**, the 'raw' total fluorescence intensity overtime of LCR and SCR cells was measured to assess the functionality of the cells using the methods described in (Elowitz et al. 2000). In that, only cells exhibiting any fluorescence throughout the entire time-series, by following the Fourier criterion (i.e. the power spectra bandwidth of cells exhibiting peaks higher than the background), were considered for image analysis. After considering only functional cells, each fluorescence signal was treated as follows: detected-trend subtraction by fitting a quadratic curve in the least-squares sense then subtracting the measured intensities from it (ii) normalization of the residual to unit power; and (iii) auto-correlation function calculation. From the treated signal, periods of oscillations of each cell were computed from the distance between the zero-crossing points of the autocorrelation function (**Figure 4.5-A**). Thus, distributions of period lengths were computed for LCR and SCR cells for each temperature condition. Interestingly, at temperatures higher than 30 °C, we observed that a few cells exhibit an oscillatory signal but also a few brief periods of no activity. These failures can be

due to either no oscillation occurring in the repressilator system or because the reporter system (see **Figure 2.7**) did not track properly the repressilator dynamics.

Since the methods for signal treatment and period estimation rely on robust periodic behavior, they cannot detect failed oscillations, but instead, assume a longer ‘apparent’ period length that combines ‘true’ and ‘failed’ oscillations, which are double the length of ‘true’ oscillations. To address this, we quantified failed oscillations from the ratio between ‘true’ and ‘apparent’ cycles in functional cells, in each condition. The mean and standard deviation of the two types of cycles are determined by the maximum likelihood estimates for the mean and variance of the second are double than that of the first of a single Gaussian or a mixture of two Gaussians (with mean and variance of the second being the double of the first) to the measured periods, using an iterative expectation maximization algorithm (Dempster et al., 1977). (**Figure 4.5-B**).

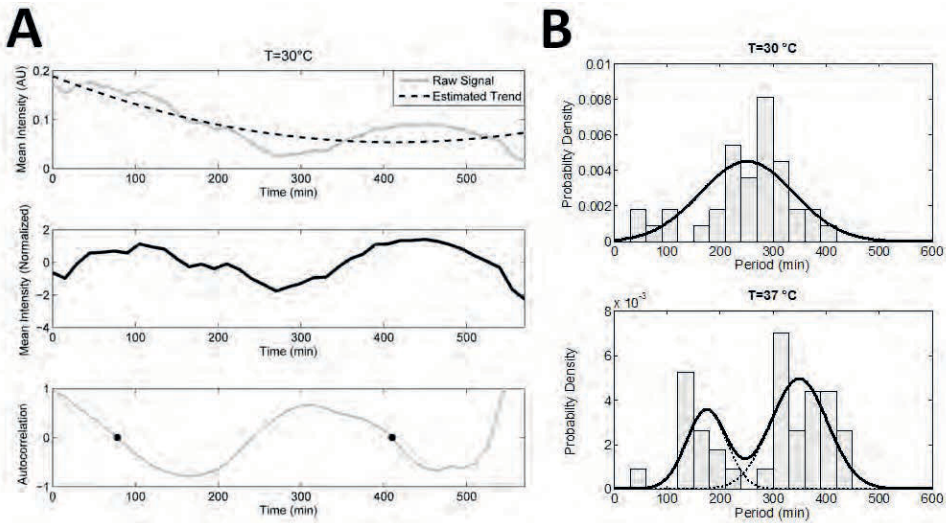


Figure 4.5: (A) Estimating period length of oscillatory signals in individual cells. In the top panel, the raw signal extracted from time-lapse fluorescence images is shown along with the estimated trend. The middle panel shows the new (treated) signal after trend subtraction and residual scaling to unit power. In the bottom panel, the computed autocorrelation function of the treated signal. The first and third zeros are represented by black circles, and the distance between them corresponds to the period of oscillation in each cell. Image extracted and adapted from **Publication I**. (B) Distribution of period lengths of the LCR system, with magnitude scaled to represent probability density, at 30 °C and 37 °C temperatures. Solid lines represent the probability density functions of the fitted model with one (top) or two Gaussians (bottom). Dashed lines represent the probability density functions of the fitted model with two Gaussians (bottom). In the bottom panel, each bimodal distribution presents two bulks of period lengths, with ‘failed’ oscillations having double the length of ‘true’ periods. Image extracted and adapted from the Supplementary Material of **Publication II**.

The results of the LCR oscillation dynamics, in varying temperatures, are presented in Table 1 and Figures 4&5 of the main manuscript of **Publication I**. Meanwhile, the results of the SCR oscillation dynamics, in varying temperatures and induction schemes, are presented in Tables 1-3 of the main manuscript, and Figures S1&S2 of **Publication II**.

4.3.2 Dynamics of Transcription Initiation

In **Publications III, IV and V**, phase contrast images are automatically segmented and manually corrected to determine the location, orientation, and size of cells, and cell lineages determined from overlapping areas between consecutive images, using *CellAging* (Häkkinen, Muthukrishnan, et al. 2013). Next, the information obtained from cell segmentation of each cell is used to align with fluorescence images using cross-correlation maximization, as in (Lloyd-price et al. 2016). In particular, in **Publications III and IV**, to accurately estimate the number of RNAs in individual cells containing both target and reporter systems, the total fluorescent MS2-GFP-RNA spot intensity in each cell was automatically detected inside each cell by kernel density estimation (KDE) method using a Gaussian kernel (Chen et al. 2008) and Otsu's Thresholding (Litcofsky et al. 2012). The changes in RNA numbers in each cell, over time, can be estimated from the changes in the corrected total spot intensity in individual cells, which is obtained by subtracting the mean cell background intensity (multiplied by the area of the spot) from the total fluorescence intensity of the spot (Häkkinen 2016). As such, since the lifetime of an MS2-GFP-RNA molecule is much longer than cell division times (Tran et al. 2015; Peabody 1993; Golding & Cox 2004), the corrected total spot intensity is expected to always increase by 'jumps' of intensities corresponding to the appearance of new tagged RNAs.

The phenotypic heterogeneity of transcription in a cell population can be characterized by measurements of mRNA molecules in individual cells. Because mRNA numbers are also affected by degradation and dilution (due to cell division), this heterogeneity is best characterized by the distribution of time intervals between produced mRNAs in individual cells. To assign integer-valued RNA numbers to the corrected spot intensities for each cell in a population over time, the position and time between the 'jumps' in spots intensity are estimated by least square fit of a monotone piecewise-constant curve (**Figure 4.6-A**). The time interval between these jumps is thus extracted from the measurements (**Figure 4.6-B**). Importantly, because the observed time intervals in microscope measurements are biased, i.e. longer intervals are less likely to be detected due to cell division (~60 minutes on average in these studies) and the time duration of time-lapse measurements (~2 hours duration), to avoid underestimating these time interval durations, we consider right censoring (Häkkinen & Ribeiro 2015). The results of transcription kinetics of the promoters $P_{lac-ara-1}$ and P_{tetA} , as measured by time intervals between jumps, for various temperature and induction schemes, are presented in Figures 2&3&5 of the main manuscript of **Publication III**. Similarly, the results of transcription kinetics of the promoter P_{RM} (one of the promoters of the CI-Cro genetic switch), in varying temperatures, are presented in Table 2 of the main manuscript of **Publication I**.

Meanwhile, in **Study IV**, integer values corresponding to RNA numbers in each cell were quantified from population images using the methods described in (Golding et al. 2005) from the distribution of total spot intensity in each cell in the image, where the first peak of intensity, which corresponds to a single mRNA molecule, is used to normalize all values of the histogram assigning to each spot intensity an integer-valued RNA number. In this study, in particular, this method was used to quantify the number of RNAs produced by the same $P_{lacO3O1}$ promoter when located in a plasmid and in *E. coli*'s chromosome. The results of $P_{lacO3O1}$

transcription kinetics for various temperature conditions are presented in Tables 1&2 and Figures 1-6 of the manuscript and in Tables S1&S3&S9 and Figures S6-S10&S15 of the supplementary results of **Study IV**.

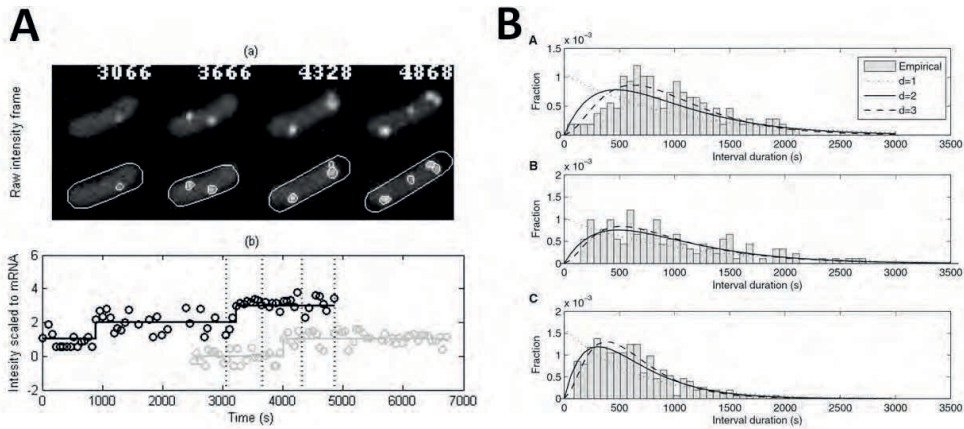


Figure 4.6: (A) In the top panel (a), example fluorescence microscopy images of *E. coli* cells carrying the MS2-GFP tagging system used in the studies of the kinetics of transcription initiation. Unprocessed and segmented cells, the spatial distribution of unbound MS2-GFP tagging proteins (gray background in the cell), RNA fluorescence spots detected (white circles). The time moments of the frame, when images were taken, are indicated on the top (in seconds). In the bottom panel (b), the example of a time series of a cell with scaled intensity levels (circles) and the line representing the monotone piecewise constant fit (solid lines) used to estimated RNA numbers. Image extracted and adapted from **Publication I**. (B) Distributions of time intervals between subsequent productions of mRNA molecules, in individual cells, under the control of the promoter P_{tetA} , in varying temperature and induction schemes. Cells were grown: (A) at 0 ng/ml aTc and 37 °C; (B) at 15 ng/ml aTc and 24 °C; and (C) at 15 ng/ml aTc and 37 °C. The probability density functions of inferred models of transcription initiation with a differing number (N) of rate-limiting steps (denoted as ' $d = N$ ') are presented in each plot. The time intervals were extracted from ~ 110 cells in all conditions. Adapted and reprinted with permission from (Muthukrishnan 2014).

Finally, for fluorescent RNAP studies in varying media richness, the abundance of RNAP in individual cells was quantified from the mean fluorescence intensity in each cell, which was extracted from fluorescence microscopy images using the methods and software previously mentioned, i.e. *CellAging* and alignment of phase-contrast and fluorescence images. The results related to the measured RNAP abundance per cell are presented in Table S7&S8 and Figure S7 of the supplementary results of **Study IV**.

4.3.3 Diffusion Coefficient of Fluorescent Aggregates and Cytoplasmic Viscosity

An osmotic stress adaptation process can result in two distinct populations of cells, adapted and non-adapted, which differ in cell and nucleoid morphology, and in division rate (Jin et al. 2013). In particular, plasmolyzed (non-adapted) cells can exhibit longer length and elliptic shape (visible in phase contrast images), and contain a condensed nucleoid (visibly in fluorescence images) (Konopka et al. 2009; Mika et al. 2010; Jin et al. 2013). Relevantly, in plasmolyzed cells, the high osmolality leads to a heavily reduced diffusion coefficient of GFP (Van Den Bogaart et al. 2007; Konopka et al. 2009). For the studies of osmotic stress in **Publication V**, after selecting cells exhibiting filamentous and elliptical morphology (Konopka et al. 2009) by cell segmentation

using phase-contrast images, tailored image and data analysis tools were developed in the search for cells that had diffusion coefficient of fluorescence spots reduced, that were not dividing during the measurement period and, finally, that contained a single condensed nucleoid (Mika et al. 2010; Jin et al. 2013).

First, spot tracking was performed using a semi-automatic method. Namely, we collected the information on the location of detected RNA molecules (fluorescent spots) in each frame using the method above. Next, the mobility of each fluorescence spot, referred to as synthetic protein aggregates, was measured in each cell using a semi-automatic method. Namely, in each frame, each spot in each cell was assigned an identification number using an automatic method. Next, we manually corrected errors in spot detection and location. In cells containing more than one spot, the spot location in each cell, per frame, is determined beforehand to avoid misidentification of spots between consecutive frames.

From the spatial information of each spot in a time-series, we used the mean squared displacement (MSD), based on the shortest distance between the spot in consecutive frames, to measure the diffusion coefficient (D) of aggregates after a time lag τ , as follows:

$$MSD(\tau) = E\|\mathbf{p}(t) - \mathbf{p}(t - \tau)\|^2 + \xi^2 = 2D\tau + \xi^2 \quad 5.1$$

where $\mathbf{p}(t)$ is the position of a spot at time t , E is the expectation over all spots and over all t , and ξ^2 is the measurement noise. To extract D , discounting ξ^2 , we used the slope of the line taken from the first two points, i.e. $D = (MSD(2) - MSD(1))/2$. From D , and assuming that the aggregates are spherical, we calculated the dynamic viscosity η of the medium in which the diffusive particle is moving (Stokes-Einstein equation):

$$\eta = \frac{k_B T}{6\pi r D} \quad 5.2$$

where k_B is the Boltzmann's constant, temperature is T , and r is the radius of the spherical particle. In equation (5.2), if changes in the relative diffusion coefficient D_1/D_2 are solely explained by changes in between two temperatures T_1/T_2 , the relative dynamic viscosity between conditions η_2/η_1 should be approximately 1. This relative diffusion coefficient D_1/D_2 can be quantified as the following:

$$\frac{D_1}{D_2} = \frac{\eta_2 T_1}{\eta_1 T_2} \quad 5.3$$

The results related to the measured diffusion coefficient and the calculated relative dynamic viscosity of the cellular cytoplasm (η_2/η_1), in varying temperatures and when exposed to osmotic stress, are presented in Table 3 of the main manuscript, and in Table S5 and Figure S8 of the supplementary results of **Publication V**.

4.3.4 Nucleoid and Protein Aggregates Spatial-Distribution

In **Publication V**, the detection and segmentation of nucleoids in each cell, in each frame, were performed using the gradient path labeling algorithm (Mora et al. 2011) and the Levenberg-Marquardt least-squares optimization algorithm (More 1978). After that, PCA was used to obtain the position, dimension, and orientation of the nucleoid in each cell, which were used to define the polar regions of each cell relative to cell extremities at the cell major axis. In that, the normalized fraction of cell poles was calculated from the measured relative values of cell width (along the cell minor axis) and the relative values of cell and nucleoid lengths (along the cell major axis). The results related to the measured nucleoid spatial-distribution in varying temperatures are presented in Table 1 and Figure 1 of the main manuscript, and in Tables S1&S3 and Figure S4 of the supplementary results of **Publication V**.

Therefore, the mean 3-D concentration of aggregate numbers at the cell poles was estimated from the ratio between the mean fraction of aggregate numbers at the poles and the volume of the cell poles, assuming uniform intracellular distribution and accounting for the capped cylindrical shape of the cells. Importantly, the above definitions and equations (5.1-5.3) were applied to cells with two nucleoids. In these cells, the midcell region was defined as the region between the outer borders of the two nucleoids. Further, these equations can be applied to all temperature conditions tested, even for possibly elongated cells when temperature increases. The results related to the measured synthetic aggregates spatial-distribution, and to the calculated aggregates concentrations inside cells, in varying temperatures, are presented in Tables 1&4&5 of the main manuscript, and in Tables S2&S3 and Figure S4 of the supplementary results of **Publication V**.

5 CONCLUSIONS AND DISCUSSION

In this thesis, we studied the effects of temperature fluctuations on the dynamics of genes and small genetic circuits (switches and clocks) in *E. coli* cells, using single-cell, single-RNA, time-lapse fluorescence microscopy and delayed stochastic models. In **Publication I**, we focused on genetic switches and repressilators and how the temperature dependence of one of the proteins present in both circuits affects their robustness to temperature shifts. In **Publication II**, we engineered a single-copy ‘repressilator’ circuit, inserted in a single-copy F-plasmid, to study the effect of gene dosage on the repressilator dynamics, as well as its robustness to temperature shifts. In **Publication III**, we focused on the effects of temperature fluctuations on transcription initiation kinetics, using the MS2-GFP tagging system to detect single RNA molecules as soon as they are produced in live cells. In **Study IV**, we implemented a newly designed MS2-GFP tagging system into *E. coli*’s chromosome, to study the temperature-dependence of the kinetics of transcription of a promoter when chromosomally-integrated and when plasmid-borne. Finally, in **Publication V**, we studied the effects of critically high and low temperatures on the cytoplasm biophysical properties. All studies were supported by microfluidic platforms to perform perturbation assays and by tailored image and signal processing tools to analyze the microscopy images, obtained using various techniques. These studies have contributed to a better understanding of how temperature fluctuations can affect the dynamics of genes and gene networks, as well as of their environment, namely the cytoplasm of *E. coli* cells.

In **Publication I**, we studied the dynamics of the ‘Elowitz & Leibler’ repressilator circuit when subject to temperature fluctuations in live *E. coli* cells. For that, we extracted the raw oscillatory fluorescent signal generated by the repressilator using live cell microscopy and used signal processing techniques to characterize not only the observed period of oscillations but also the failures in oscillation. That allowed us to also characterize the functionality (fraction of cells exhibiting oscillations) and the robustness (fraction of expected oscillations that occur) of this circuit. From this detailed analysis, we found that the repressilator is sensitive to temperature fluctuations, as

the mean and variability of oscillation period lengths differ as temperature changes. Next, we showed that the robustness of the period lengths gradually decreases with increasing temperature, in that most cells exhibit ‘failed’ oscillations when temperature increases beyond 30 °C, while the ‘apparent’ (measured) period was minimized at 30 °C. Meanwhile, the ‘true’ period, extracted from the fitting of a 2-Gaussians model to the measured period length distributions, was found to decrease with increasing temperature. This result is in accordance with the behavior of other synthetic genetic clocks (Stricker et al. 2008) and with the expected increase in speed of thermodynamic processes with temperature, e.g. protein decay rates (Loinger et al. 2007), and increased protein ‘dilution’ rate due to increased cell doubling rate. Meanwhile, at lower-than-optimal temperatures, the ‘true’ periods expectedly increased due to the reduction in rate of most of the chemical processes involved in the functioning of this circuit.

Possible causes that could explain the shift in behavior as temperature increased were further investigated. In particular, we focused on the ‘failed’ oscillations and on why were these temperature-dependent. From previous evidence that the functionality of the CI repressor (one of the circuit’s components) is highly affected at temperatures above 30 °C (Jana et al. 1999; Frank et al. 1997; Hillen et al. 1982), we hypothesized that this was the main cause for the failures in oscillating. Namely, we hypothesized that the increasing temperature compromised the wild-type CI’s binding stability to the promoter region (likely caused by failures in CI’s dimerization (K. S. Koblan & Ackers 1991)).

To test this, we analyzed the dynamics of the CI-Cro genetic switch in the same temperature range. Namely, the transcription dynamics of the P_{RM} promoter, regulated by CI protein produced by the other gene of the switch, was followed at various temperatures, using MS2-GFP RNA tagging. The results showed that, at the same temperature as in the repressilator, there is a significant behavioral change as well. Namely, the dynamics of RNA production by P_{RM} shifts from sub-Poissonian to super-Poissonian (consistent with long periods of repression by the other protein of the switch, Cro). This implies that, with increasing mal-functioning of CI protein (as in the case of the repressilator), the ability to repress the second gene of the switch (P_R) decreases, making P_R to become more active thus increasing the repression on P_{RM} . We find these results to be consistent with the general concept that noise in gene expression and phenotypic cell-to-cell variability ought to be analyzed accounting for context-dependent regulatory interactions (Dunlop et al. 2008).

Supported by the behavioral shifts at the same temperature in the two circuits studied and the fact that they have only one component in common, the CI protein, we concluded that it was the temperature-dependency of CI’s functionality (K. S. Koblan & Ackers 1991; K. S. Koblan & Ackers 1991) the cause for the observed behavioral changes. These results demonstrate that future genetic circuits should consider the temperature sensitivity of the components and make use of

them or, instead, introduce temperature-compensation mechanisms to minimize the effects (Hussain et al. 2014), or the replacement of such components by others shown to be less-sensitive.

In **Publication II**, we compared the dynamics of the original Elowitz & Leibler repressilator circuit (LCR) and a single-copy repressilator (SCR), engineered from its original sequence (Elowitz et al. 2000) into a single-copy F-plasmid. The main questions were whether it is possible to engineer a single-copy genetic clock and whether multiple copies support or decrease the stability of the periodicity. Further, we addressed these questions in multiple temperature conditions, to investigate whether the answers are temperature-dependent. We expected that the period length and the robustness to temperature fluctuations of the two systems would differ, due to one of them having network coupling. In agreement, we found that the variability in period length (measured by the variability of the time length of the oscillations, from the squared coefficient variation CV^2) of the SCR is higher. This agrees with the concept that lowering the copy-number can decrease the rhythmicity of a coupled system (Garcia-Ojalvo et al. 2004). Also, the mean period of oscillations of the SCR was found to be higher than that of the LCR, in all temperatures tested, except 28 °C. Further, we found that the functionality of the SCR, when compared to the LCR, was higher at lower temperatures, but lower at higher temperatures. This suggests that the lowering functionality at non-optimal temperatures is caused by different factors at higher-than-optimal and lower-than-optimal temperatures.

Relevantly, the functionality and the period length of both systems are, respectively, maximized and minimized at 30 °C. Overall, the fact that the dynamics of the two systems differ when at the same conditions demonstrates that the functioning of the original LCR is affected, to some extent, by the dynamical coupling of the multiple functional copies of the circuit in a cell.

Next, we studied the functionality and robustness of the repressilator to external perturbations as a function of the copy numbers and of the perturbation strength, i.e. low and high IPTG concentrations. We found that the functionality of SCR was sensitive to the strength of the perturbation, and more robust than of the LCR at low IPTG concentrations (50 μ M). This was expected, since having multiple copies of the circuit ought to make the coupled networks more sensitive to external perturbations than when only one network is present in the cell. This result was in agreement with the behavior of a delayed stochastic model that assumed perfect coupling between the identical copies of the circuit in the model cell. It also shows that, when perturbing one circuit, all other coupled circuits are perturbed as well.

In summary, our results suggest that the SCR is a more robust clock than the LCR under certain conditions. Therefore, we expect that the implementation of synthetic genetic circuits in a single-copy DNA location (F-plasmids or chromosome), which avoid the phenomenon of network coupling, can be a successful strategy towards designing genetic clocks whose functioning requires robust time tracking under external perturbations.

In **Publication III**, we analyzed the temperature-dependence of the dynamics of the underlying steps in transcription initiation of the $P_{\text{lac-ara-1}}$ promoter for varying induction schemes of IPTG and arabinose, and for a wide range of temperatures above and below the optimal growth temperature. For this, we used time-lapse single-RNA microscopy measurements in live *E. coli* cells and tailored image and signal processing tools. From the images, using those tools, we extracted time intervals between consecutive transcription events. Next, we used statistical methods to find which stochastic model best fitted the empirical data in maximum likelihood sense (for models varying in number and duration of rate-limiting steps in transcription initiation). The selected models informed which rate-limiting steps of transcription initiation are more/less responsible for the changes observed in RNA production kinetics with changing temperature and induction schemes. The results showed that not all steps in transcription initiation are affected equally by temperature changes. In particular, for the $P_{\text{lac-ara-1}}$ promoter, regardless of the induction scheme, varying only some of the steps suffices to explain the changes observed in the measured time intervals between transcription events. Interestingly, we found that changes in RNAP concentrations with changing temperatures can only explain partially the observed changes in transcription dynamics. Further, the analysis indicates that the temperature changes mainly affect the dynamics of the OC formation, rather than the steps prior to this.

Next, we quantified the temperature-dependence of the transcription kinetics of the P_{tetA} promoter using the same methods. We found that the distribution of time intervals between consecutive RNA production events to be less noisy than of the $P_{\text{lac-ara-1}}$, due to having a different rate-limiting steps kinetics. In particular, as temperature is changed, for $P_{\text{lac-ara-1}}$ promoter the changes occur in the stage of CC formation. Meanwhile, for the promoter P_{tetA} , this response arises from changes in the OC formation and/or promoter escape dynamics.

Subsequently, by tuning the dynamics of transcription initiation, by changing the kinetics of the rate-limiting steps, the selected inferred models were further studied to explore why the identified steps were preferred as the main ‘sources’ for tuning both mean and variance of time intervals, with temperature fluctuations. We found that the dynamics of transcription initiation is either insensitive or responds reciprocally to changes in the other steps, thus providing promoters greater adaptability to temperature changes than if tuned by other means. The results may be an explanation for why different natural promoters differ in their response to temperature shifts.

In **Study IV**, we studied the temperature-dependence of the dynamics of transcription initiation of the promoter P_{lacO3O1} , under full induction, when integrated into *E. coli*’s chromosome and when on a single-copy plasmid. Temperature-dependent changes in DNA supercoiling have been proposed to be one of the sensing mechanisms used by *E. coli*, to rapidly identify and transduce external temperature fluctuations (Los 2004; Eriksson et al. 2002). E.g., the buildup of DNA supercoiling on a chromosomal DNA segment, if occurring close to a promoter region, can stop

transcription initiation. When occurring, the promoter activity will only be resumed upon the release of supercoiling topological constraints by gyrase activity (Chong et al. 2014). Interestingly, the same does not appear to occur in plasmids (Chong et al. 2014). We hypothesized that temperature affects the propensity for supercoiling buildup and/or escape from these states. If so, genes would differ in how their activity is affected by temperature shifts, depending on whether they are located on a plasmid or on the chromosome.

To investigate the hypothesis that genes located in plasmids and chromosome differ in sensitivity to temperature fluctuations, we engineered two MS2-GFP tagging systems for single RNA detection under the control of the same promoter, $P_{lacO3O1}$. We then implemented a copy of this system into a single-copy F-plasmid and another into *E. coli*'s chromosome. Importantly, this *lac* derivative promoter lacks the O2 operator site, known to form topological constraints when repressed by its corresponding repressor (Fulcrand et al. 2016).

While both constructs were found to be functional and responsive to temperature changes, at critically low temperatures they differ more in behavior from one another. Namely, unlike the plasmid-borne promoter, the chromosome-integrated promoter presents a weaker and noisier relative RNA production activity. Also, we showed that, in the chromosome-integrated promoter, the process of resuming transcription events due to the release of DNA supercoiling buildup is hampered at lower-than-optimal temperatures, thus drastically reducing the activity of the promoter, when compared to the plasmid-borne promoter. This result was validated by the dissection of the kinetics of transcription initiation of the two promoters, following the methods described in (Lloyd-price et al. 2016). In that, when compared to the plasmid-borne, the chromosomally-integrated $P_{lacO3O1}$ exhibited a widely reduced transcription rate in stages prior to the OC formation, i.e. indicating that it remains longer in a locked state, before committing to transcription.

Overall, the results showed that the chromosomally-integrated promoter has longer-lasting supercoiled states at critically low temperatures, which provides the promoter with increased sensitivity to temperature fluctuations. The results thus suggest that the sensitivity of transcription activity to temperature fluctuations depends on whether the promoter is on the chromosome or on a plasmid.

Next, we studied why this reduction was observed only at low temperatures. If it is true that a locked promoter due to DNA supercoiling buildup can resume transcription activity by gyrase or topoisomerase-I activities, then the promoter activity should be reduced in the presence of the inhibitors of these proteins' activity. To test for this, first, we conducted temporal measurements of RNA numbers in individual cells at 10 °C and 30 °C in the presence and absence of inhibitors. We found that the activity of the chromosomally-integrated $P_{lacO3O1}$ promoter (compared to plasmid-borne) was reduced in the presence of inhibitors at both temperatures, but further reduced at

critically-low temperatures. In agreement with the hypothesis that the proteins gyrase or topoisomerase-I may have evolved mechanisms to function at critically-low temperatures (Yamanaka 1999; Jones et al. 1987), these results suggested that the blocking of the release of $P_{lacO3O1}$ from supercoiled buildup states, at low temperatures, is not due to reduced functionality of either gyrases or topoisomerases-I, but, instead, to an energy-associated increasing difficulty in releasing the promoters from the ‘locked’ state.

To test this hypothesis, we first conducted an additional experiment where cells were subject to energy depletion stress, via DNP treatment, with cells grown at 30 °C. The results showed that promoter release from DNA supercoiling buildup is hampered when cellular energy is depleted, similarly to when at critically-low temperatures, confirming that the activity of the chromosomally-integrated promoter at 10 °C is being hampered by an energy-related increased difficulty in unblocking the DNA from supercoiled states. Relevantly, we also studied the long-term consequences of temperature on the transcription activity of blocked promoters. For that, we conducted an experiment where cells were subject to consecutive shifts between high and low temperatures. The results showed that low temperatures have no long-term consequences on the ability of the promoters to transcribe. In the end, simulations of a stochastic model with realistic parameter values for *lac* derivative promoters showed that the differences observed empirically are consistent with models of transcription that model the effects of supercoiling buildups using a single parameter, namely, the escape time from supercoiled states.

In **Publication V**, we studied the temperature-dependence of the phenomenon of segregation of large protein aggregates to the cell poles by a nucleoid exclusion phenomenon (Gupta et al. 2014), which, following cell division, biasedly partitions the aggregates between daughter cells, thus resulting in the renewal of some cell lineages, at the expense of accelerating cell aging in others (Lindner et al. 2008). For that, we measured the motion of protein aggregates in cells at a wide range of temperatures by assessing the spatial-distribution and mobility of MS2-GFP tagged RNAs, i.e. synthetic fluorescent aggregates, and fluorescently-tagged natural aggregates (IbpA) using confocal microscopy. During imaging, temperatures were changed from optimal (37°C) to substantially low (10°C) in short periods of time. We also studied the long-term consequences of sub-optimal temperatures on the spatial distribution of the aggregates in mother cells and their partitioning between daughter cells. We found that, as temperature decreases, protein aggregates became more homogeneously distributed and less correlated with nucleoid size and location. This was found to be correlated with an increased cytoplasm viscosity in the cells that, not only weakens the anisotropy in aggregate displacements at the nucleoid borders but also increases the concentration of aggregates in between two nucleoids in cells close to division. Consequently, we found reduced asymmetries in aggregates partitioning between daughter cells at the lower temperatures.

Overall, at low temperatures, the nucleoid exclusion phenomenon becomes nonfunctional in that, on average, aggregates no longer preferentially locate at the cell poles, due to enhanced cytoplasm viscosity that renders ‘ineffective’ the interactions between nucleoid and aggregates. We validated these findings by subjecting cells to osmotic stress. Similarly to lowering temperature, we observed that plasmolyzed cells (i.e. with enhanced relative dynamic cytoplasmic viscosity) were also unable to segregate the aggregates to the poles. The results suggest that, in any stress condition where cytoplasm viscosity is increased, such as carbon starvation, energy depletion, and during the stationary phase (Parry et al. 2014), one can expect a loss of aggregate’s preference for polar localization. Further, we expect that the spatial localization of other large cellular components such as plasmids, enzyme complexes, micro-compartments (Kerfeld et al. 2010) and other macromolecules, to be likewise affected by enhanced cytoplasm viscosity.

Interestingly, the increased relative cytoplasmic viscosity with decreasing temperature creates the same glassy-like behavior of the cytoplasm that occurs when the metabolic activity is reduced (Parry et al. 2014). In that study, the cytoplasm viscosity was found to be ‘metabolism dependent’. Namely, enhanced metabolism activity causes the cytoplasm to be fluidized, which allows larger components to move more freely and reach larger regions of the cytoplasm. In the case of the plasmolyzed cells that observed, the increase in viscosity is likely due to increased macromolecular crowding, which reduces the diffusion rate of large components.

Interestingly, due to the above phenomena being biophysical in nature, we expect their effects on aggregate preference of polar localization to occur during the exposure to these stressful conditions but have short-term consequences once normal temperatures are resumed, particularly in the case of *E. coli* cells able to have fast division rates.

In conclusion, from our single-cell based studies on *E. coli*, we found core cellular processes to be widely affected by temperature fluctuations in a variety of aspects. We observed temperature-dependent transcription initiation dynamics configurations, temperature-dependent robustness and sensitivity of synthetic genetic circuits, temperature-dependent effects of DNA topological constraints on transcription activity, and, temperature-dependent diffusion and segregation kinetics of protein aggregates. We expect this knowledge, along with the methodologies presented in this thesis, to contribute to a better understanding of how, in general, temperature affects single-cell organisms. Regarding genetic circuits, we expect that this added knowledge on their temperature-dependent to be of assistance to future efforts in synthetic biology aiming to improve the robustness and sensitivity of genetic circuits to environmental changes.

References

- Alberts, B. et al., 2002. *Molecular biology of the cell* 4th ed. J. H. Wilson & T. Hunt, eds., New York, New York, USA: Garland Science.
- Amit, R., Oppenheim, A.B. & Stavans, J., 2003. Increased bending rigidity of single DNA molecules by H-NS, a temperature and osmolarity sensor. *Biophysical Journal*, 84(4), pp.2467–2473.
- Anderson, J.C., Voigt, C. a & Arkin, A.P., 2007. Environmental signal integration by a modular AND gate. *Molecular Systems Biology*, 3(133), pp.1–8.
- Arkin, A., Ross, J. & Mcadams, H.H., 1998. Stochastic Kinetic Analysis of Developmental Pathway Bifurcation in. *Genetics*, 149(8), pp.1633–1648.
- Arsène, F., Tomoyasu, T. & Bukau, B., 2000. The heat shock response of *Escherichia coli*. *International Journal of Food Microbiology*, 55(1–3), pp.3–9.
- Atkinson, M.R. et al., 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*, 113(5), pp.597–607.
- Axelrod, D., 1981. Cell-substrate Contacts Illuminated by Total-Internal Reflection Fluorescence. *Journal of Cell Biology*, 89(9), pp.141–145.
- Baba, T. et al., 2006. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, 2, p.2006.0008.
- Bailey, J.E. et al., 1986. Studies of host-plasmid interactions in recombinant microorganisms. *Annals of the New York Academy of Sciences*, 469, pp.194–211.
- Bakshi, S. et al., 2012. Superresolution imaging of ribosomes and RNA polymerase in live *Escherichia coli* cells. *Molecular Microbiology*, 85(1), pp.21–38.
- Balaban, N.Q. et al., 2004. Bacterial persistence as a phenotypic switch. *Science*, 305(5690), pp.1622–5.

- Barkley, M.D. et al., 1975. Interaction of Effecting Ligands with Lac Repressor and Repressor-Operator Complex. *Biochemistry*, 14(8), pp.1700–1712.
- Baudin, A. et al., 1993. A simple and efficient method for direct gene deletion in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 21(14), pp.3329–3330.
- Beal, J. et al., 2012. An End-to-End Workflow for Engineering of Biological Networks from High-Level Specifications. *ACS Synth Biology*, 1, pp.317–331.
- Becker, N.A., Peters, J.P. & Maher, L.J., 2013. Mechanism of promoter repression by Lac repressor-DNA loops. *Nucleic Acids Research*, 41(1), pp.156–166.
- Beckwith, J., Grodzicker, T. & Arditti, R., 1972. Evidence for Two Sites in the *lac* Promoter Region. *Journal of Molecular Biology*, 69, pp.155–160.
- Becskei, A. et al., 2005. Contributions of low molecule number and chromosomal positioning to stochastic gene expression. *Nature Genetics*, 37(9), pp.937–944.
- Becskei, A. et al., 2000. Engineering stability in gene networks by autoregulation. *Nature*, 405(6786), pp.590–593.
- Bednarska, N.G. et al., 2013. Protein aggregation in bacteria: The thin boundary between functionality and toxicity. *Microbiology*, 159, pp.1795–1806.
- Bergthorsson, U. & Ochman, H., 1998. Distribution of chromosome length variation in natural isolates of *Escherichia coli*. *Molecular Biology and Evolution*, 15(1), pp.6–16.
- Bernstein, J.A. et al., 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(15), pp.9697–9702.
- Beyreuther, K. et al., 1973. The amino-acid sequence of lac repressor. *Proceedings of the National Academy of Sciences of the United States of America*, 70(12), pp.3576–3580.
- Blattner, F.R. et al., 1997. The Complete Genome Sequence of *Escherichia coli* K-12. *Science*, 277(5331), pp.1453–1462.
- Boeke, J.D. & Model, P., 1982. A prokaryotic membrane anchor sequence: carboxyl terminus of bacteriophage f1 gene III protein retains it in the membrane. *Proceedings of the National Academy of Sciences of the United States of America*, 79(17), pp.5200–5204.
- Van Den Bogaart, G. et al., 2007. Protein mobility and diffusive barriers in *Escherichia coli*: Consequences of osmotic stress. *Molecular Microbiology*, 64(3), pp.858–871.
- Bondeson, K. et al., 1993. Lactose repressor-operator DNA interactions: kinetic analysis by a surface plasmon resonance biosensor. *Analytical Biochemistry*, 214, pp.245–251.
- Bonnet, J. et al., 2013. Amplifying genetic logic gates. *Science*, 340(6132), pp.599–603.
- Born, M. & Wolf, E., 1999. *Principles of Optics* 7th ed., Cambridge Univ. Press, U.K.

- Bratsun, D. et al., 2005. Delay-induced stochastic oscillations in gene regulation. *Proceedings of the National Academy of Sciences of the United States of America*, 102(41), pp.14593–14598.
- Bremer, H. & Dennis, P.P., 1996. *Modulation of Chemical Composition and Other Parameters of the Cell by Growth Rate* F. C. Neidhardt, ed., Washington, DC: American Society for Microbiology Press.
- Brophy, J.A.N. & Voigt, C.A., 2014. Principles of genetic circuit design. *Nature Methods*, 11(5), pp.508–520.
- Browning, D. et al., 2009. Assays for Transcription Factor Activity Douglas. In T. Moss & B. Leblanc, eds. *Methods in Molecular Biology, DNA-Protein Interactions*. Humana Press, pp. 369–387.
- Browning, D.F. & Busby, S.J., 2004. The regulation of bacterial transcription initiation. *Nature Reviews Microbiology*, 2(1), pp.57–65.
- Browning, D.F. & Busby, S.J.W., 2016. Local and global regulation of transcription initiation in bacteria. *Nature Reviews Microbiology*, 14(10), pp.638–650.
- Buc, H. & McClure, W.R., 1985. Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry*, 24(11), pp.2712–2723.
- Bulter, T. et al., 2004. Design of artificial cell–cell communication using gene and metabolic networks. *Proceedings of the National Academy of Sciences of the United States of America*, 101(8), pp.2299–2304.
- Busby, S. & Ebright, R.H., 1999. Transcription activation by catabolite activator protein (CAP). *Journal of Molecular Biology*, 293(2), pp.199–213.
- Çağatay, T. et al., 2009. Architecture-dependent noise discriminates functionally analogous differentiation circuits. *Cell*, 139(3), pp.512–22.
- Cameron, D.E. & Collins, J.J., 2014. Tunable protein degradation in bacteria. *Nature Biotechnology*, 32(November), pp.1–8.
- Cardinale, S. & Arkin, A.P., 2012. Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems. *Biotechnology Journal*, 7(7), pp.856–866.
- Cardinale, S., Joachimiak, M.P. & Arkin, A.P., 2013. Effects of genetic variation on the *E. coli* host-circuit interface. *Cell Reports*, 4, pp.231–237.
- Cech, C.L. & McClure, W.R., 1980. Characterization of ribonucleic acid polymerase-T7 promoter binary complexes. *Biochemistry*, 19(11), pp.2440–2447.
- Chamberlin, M.J., 1974. The selectivity of transcription. *Annual Review of Biochemistry*, 43(1), pp.721–775.
- Chen, H. et al., 2015. Genome-wide study of mRNA degradation and transcript elongation in

- Escherichia coli*. *Molecular Systems Biology*, 11(1), pp.781–781.
- Chen, T.-B.B. et al., 2008. Segmentation of cDNA microarray images by kernel density estimation. *Journal of Biomedical Informatics*, 41(6), pp.1021–1027.
- Chen, Y.Y., Galloway, K.E. & Smolke, C.D., 2012. Synthetic biology: advancing biological frontiers by building synthetic systems. *Genome biology*, 13(240), pp.1–10.
- Chew, L.C. & Tacon, W.C., 1990. Simultaneous regulation of plasmid replication and heterologous gene expression in *Escherichia coli*. *Journal of Biotechnology*, 13(1), pp.47–60.
- Chong, S. et al., 2014. Mechanism of transcriptional bursting in bacteria. *Cell*, 158(2), pp.314–326.
- Chubukov, V. et al., 2014. Coordination of microbial metabolism. *Nature Reviews Microbiology*, 12(5), pp.327–40.
- Clancy, K. & Voigt, C. a, 2010. Programming cells: towards an automated “Genetic Compiler.” *Current Opinion in Biotechnology*, 21, pp.572–81.
- Coquel, A.S. et al., 2013. Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLoS Computational Biology*, 9(4), p.e1003038.
- Cormack, B.P., Valdivia, R.H. & Falkow, S., 1996. FACS-optimized mutants of the green fluorescent protein (GFP). *Gene*, 173(1 Spec No), pp.33–8.
- Coulon, A. et al., 2013. Eukaryotic transcriptional dynamics: from single molecules to cell populations. *Nature Reviews Genetics*, 14(8), pp.572–84.
- Crick, F.H.C., 1970. Central Dogma of Molecular Biology. *Nature*, 227(5258), pp.561–563.
- Danino, T. et al., 2010. A synchronized quorum of genetic clocks. *Nature*, 463(7279), pp.326–30.
- Day, R.N. & Davidson, M.W., 2009. The fluorescent protein palette: tools for cellular imaging. *Chemical Society Reviews*, 38, pp.2887–2921.
- Deuerling, E. et al., 1999. Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature*, 400(8), pp.693–696.
- Dillon, S.C. & Dorman, C.J., 2010. Bacterial nucleoid-associated proteins, nucleoid structure and gene expression. *Nature Reviews Microbiology*, 8(3), pp.185–195.
- Dong, H. et al., 1995. Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction. *Journal of Bacteriology*, 177(6), pp.1497–504.
- Dong, T. & Schellhorn, H.E., 2009. Global effect of RpoS on gene expression in pathogenic *Escherichia coli* O157:H7 strain EDL933. *BMC Genomics*, 10(349), pp.1–17.
- Dorman, C.J., 2006. DNA supercoiling and bacterial gene expression. *Science Progress*, 89,

pp.151–166.

- Dorman, C.J. & Dorman, M.J., 2016. DNA supercoiling is a fundamental regulatory principle in the control of bacterial gene expression. *Biophysical Reviews*, 8, pp.89–100.
- Dunlap, J.C., 1999. Molecular bases for circadian clocks. *Cell*, 96(2), pp.271–290.
- Dunlop, M.J. et al., 2008. Regulatory activity revealed by dynamic correlations in gene expression noise. *Nature Genetics*, 40(12), pp.1493–8.
- Dymond, J.S. et al., 2011. Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, 477(7365), pp.471–476.
- Ebersbach, G. & Gerdes, K., 2005. Plasmid Segregation Mechanisms. *Annual Review of Genetics*, 39(1), pp.453–479.
- Eliasson, A. et al., 1992. Direct visualization of plasmid DNA in bacterial cells. *Molecular Microbiology*, 6(2), pp.165–70.
- Ellefson, J.W. et al., 2014. Directed evolution of genetic parts and circuits by compartmentalized partnered replication. *Nature Biotechnology*, 32(1), pp.97–101.
- Ellis, T., Adie, T. & Baldwin, G.S., 2011. DNA assembly for synthetic biology: from parts to pathways and beyond. *Integrative Biology*, 3(2), p.109. Available at: <http://xlink.rsc.org/?DOI=c0ib00070a>.
- Elowitz, M.B. et al., 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*, 403(6767), pp.335–8.
- Elowitz, M.B. et al., 2002. Stochastic gene expression in a single cell. *Science*, 297(5584), pp.1183–6.
- Emmer, M. et al., 1970. Cyclic AMP receptor protein of *E. coli*: its role in the synthesis of inducible enzymes. *Proceedings of the National Academy of Sciences of the United States of America*, 66(2), pp.480–487.
- Endy, D., 2005. Foundations for engineering biology. *Nature*, 438(7067), pp.449–453.
- Engler, C., Kandzia, R. & Marillonnet, S., 2008. A one pot, one step, precision cloning method with high throughput capability. *PLoS One*, 3(11), p.e3647.
- Englesberg, E. et al., 1965. Positive control of enzyme synthesis by gene C in the L-arabinose system. *Journal of Bacteriology*, 90(4), pp.946–957.
- Eriksson, S., Hurme, R. & Rhen, M., 2002. Low-temperature sensors in bacteria. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 357(1423), pp.887–893.
- Farewell, A., Kvint, K. & Nystro, T., 1998. Negative regulation by RpoS : a case of sigma factor competition. *Molecular Microbiology*, 29(4), pp.1039–1051.
- Fleischmann, R. et al., 1995. Whole-genome random sequencing and assembly of *Haemophilus*

- influenzae* Rd. *Science*, 269(5223), pp.496–512.
- Frank, D.E. et al., 1997. Thermodynamics of the interactions of lac repressor with variants of the symmetric lac operator: effects of converting a consensus site to a non-specific site. *Journal of Molecular Biology*, 267(5), pp.1186–206.
- Friedman, D.I. & Gottesman, M., 1983. Lytic Mode of Lambda Development. In *LambdaII*. Cold Spring Harbor Laboratory, pp. 21–51.
- Fulcrand, G. et al., 2016. DNA supercoiling, a critical signal regulating the basal expression of the lac operon in *Escherichia coli*. *Scientific Reports*, 6(19243), pp.1–12.
- Fung, E. et al., 2005. A synthetic gene – metabolic oscillator. *Nature*, 435, pp.118–122.
- Fusco, D. et al., 2003. Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Current Biology*, 13(2), pp.161–167.
- Fussenegger, M., Tigges, M. & Marquez-lago, T.T., 2009. A tunable synthetic mammalian oscillator. *Nature*, 457(January), pp.309–312.
- Gally, D.L. et al., 1993. Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in *Escherichia coli* K-12: Effects of temperature and media. *Journal of Bacteriology*, 175(19), pp.6186–6193.
- Garcia-Ojalvo, J., Elowitz, M.B. & Strogatz, S.H., 2004. Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing. *Proceedings of the National Academy of Sciences of the United States of America*, 101(30), pp.10955–60.
- Garcia, H.G. et al., 2010. Transcription by the numbers redux: experiments and calculations that surprise. *Trends in cell biology*, 20(12), pp.723–33.
- Gardner, T.S., Cantor, C.R. & Collins, J.J., 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403(6767), pp.339–42.
- Gardner, T.S. & Collins, J.J., 2000. Neutralizing noise in gene networks. *Nature Reviews Genetics*, 405(06), pp.520–521.
- Gasnier, M. et al., 2013. Fluorescent mRNA labeling through cytoplasmic FISH. *Nature Protocols*, 8(12), pp.2538–2547.
- Gibson, D.G. et al., 2010. Chemical synthesis of the mouse mitochondrial genome. *Nature Methods*, 7(11), pp.901–3.
- Gibson, D.G. et al., 2008. Complete Chemical Synthesis, Assembly, and Cloning of a *Mycoplasma genitalium* Genome. *Science*, 319(5867), pp.1215–1220.
- Gibson, D.G. et al., 2010. Creation of a Bacterial Cell Controlled by a Chemically Synthesized Genome. *Science*, 329(5987), pp.52–56.
- Gibson, D.G. et al., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, 6(5), pp.343–347.

- Gibson, M. a. & Bruck, J., 2000. Efficient Exact Stochastic Simulation of Chemical Systems with Many Species and Many Channels. *The Journal of Physical Chemistry A*, 104(9), pp.1876–1889.
- Gilbert, W. & Maxam, A., 1973. The Nucleotide Sequence of the *lac* Operator. *Proceedings of the National Academy of Sciences*, 70(12), pp.3581–3584.
- Gillespie, D.T., 1976. A general method for numerically simulating coupled chemical reactions. *Journal of Computational Physics*, 22(4), pp.403–434.
- Gillespie, D.T., 1977. Concerning the validity of the stochastic approach to chemical kinetics. *Journal of Statistical Physics*, 16(3), pp.311–318.
- Gillespie, D.T., 1977. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(4), pp.2340–2361.
- Gillespie, D.T., 2007. Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58, pp.35–55.
- Glossop, N.R.J., Lyons, L.C. & Hardin, P.E., 1999. Interlocked Feedback Loops Within the *Drosophila* Circadian Oscillator. *Science*, 286(5440), pp.766–768.
- Goeddel, D. V et al., 1979. Expression in *Escherichia coli* of chemically synthesized genes for human insulin. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1), pp.106–110.
- Golding, I. et al., 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*, 123(6), pp.1025–1036.
- Golding, I. & Cox, E.C., 2006. Eukaryotic Transcription: What Does It Mean for a Gene to Be ‘on’? *Current Biology*, 16(10), pp.371–373.
- Golding, I. & Cox, E.C., 2004. RNA dynamics in live *Escherichia coli* cells. *Proceedings of the National Academy of Sciences of the United States of America*, 101(31), pp.11310–5.
- Goncalves, N.S.M. et al., 2018. Temperature-dependence of the single-cell variability in the kinetics of transcription activation in *Escherichia coli*. *Physical Biology*, 15(2), p.026007.
- Gordon, G.S. et al., 1997. Chromosome and low copy plasmid segregation in *E. coli*: visual evidence for distinct mechanisms. *Cell*, 90(6), pp.1113–21.
- Govers, S.K., Dutré, P. & Aertsen, A., 2014. *In vivo* disassembly and reassembly of protein aggregates in *Escherichia coli*. *Journal of Bacteriology*, 196(13), pp.2325–2332.
- Gray, M. & Honigberg, S.M., 2001. Effect of chromosomal locus, GC content and length of homology on PCR-mediated targeted gene replacement in *Saccharomyces cerevisiae*. *Nucleic Acids Research*, 29(24), pp.5156–62.
- Greive, S.J. & von Hippel, P.H., 2005. Thinking quantitatively about transcriptional regulation. *Nature Reviews Molecular Cell Biology*, 6(3), pp.221–32.

- Gupta, A. et al., 2014. *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*. *Biophysical Journal*, 106(9), pp.1928–37.
- Guye, P. et al., 2013. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Research*, 41(16), p.e156.
- Häkkinen, A., Muthukrishnan, A.-B., et al., 2013. CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli*. *Bioinformatics*, 29, pp.1708–9.
- Häkkinen, A., Tran, H., et al., 2013. Effects of rate-limiting steps in transcription initiation on genetic filter motifs. *PLoS One*, 8(8), p.e70439.
- Häkkinen, A., 2016. *Quantifying Transcriptional Dynamics and Their Effects on Genetic Motifs from Live Cell Fluorescence Microscopy*. PhD thesis, Tampere University of Technology.
- Häkkinen, A. & Ribeiro, A.S., 2015. Characterizing rate limiting steps in transcription from RNA production times in live cells. *Bioinformatics*, 32(9), pp.1346–1352.
- Ham, T.S. et al., 2006. A Tightly Regulated Inducible Expression System Utilizing the fim Inversion Recombination Switch. *Biotechnology and Bioengineering*, 94(1), pp.1–4.
- Hardin, P.E., Hall, J.C. & Rosbash, M., 1990. Feedback of the *Drosophila* period gene product on circadian cycling of its messenger RNA levels. *Nature*, 343(6258), pp.536–540.
- Hardy, C.D. & Cozzarelli, N.R., 2005. A genetic selection for supercoiling mutants of *Escherichia coli* reveals proteins implicated in chromosome structure. *Molecular Microbiology*, 57(6), pp.1636–1652.
- Harley, C.B. & Reynolds, R.P., 1987. Analysis of *E. coli* promoter sequences. *Nucleic Acids Research*, 15(5), pp.2343–2361.
- Haseltine, E.L. & Arnold, F.H., 2007. Synthetic Gene Circuits: Design with Directed Evolution. *Annual Review of Biophysics and Biomolecular Structure*, 36(1), pp.1–19.
- Hastings, J.W. & Sweeney, B.M., 1957. On the Mechanism of Temperature Independence in a Biological Clock. *Proceedings of the National Academy of Sciences of the United States of America*, 43(9), pp.804–811.
- Hasty, J. et al., 2001. Computational studies of gene regulatory networks: in numero molecular biology. *Nature Reviews Genetics*, 2(4), pp.268–79.
- Hasty, J. et al., 2002. Synthetic Gene Network for Entraining and Amplifying Cellular Oscillations. *Physical Review Letters*, 88(14), p.148101.
- Helling, R.B. & Weinberg, R., 1963. Complementation Studies of Arabinose Genes in *Escherichia coli*. *Genetics*, 48(October), pp.1397–1410.
- Henderson, K.L. et al., 2017. Mechanism of transcription initiation and promoter escape by *E. coli* RNA polymerase. *Proceedings of the National Academy of Sciences*, 114(15), pp.E3032–E3040.

- Hensel, Z. et al., 2012. Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis. *Nature Structural & Molecular Biology*, 19(8), pp.797–802.
- Hilborn, R.C. & Erwin, J.D., 2008. Stochastic coherence in an oscillatory gene circuit model. *Journal of Theoretical Biology*, 253(2), pp.349–354.
- Hilfinger, A. & Paulsson, J., 2011. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proceedings of the National Academy of Sciences of the United States of America*, 108(29), pp.12167–72.
- Hill, A.D. et al., 2008. SynBioSS: The synthetic biology modeling suite. *Bioinformatics*, 24(21), pp.2551–2553.
- Hillen, W. et al., 1982. Purification of the TET repressor and TET operator from the transposon Tn10 and characterization of their interaction. *The Journal of Biological Chemistry*, 257(11), pp.6605–6613.
- von Hippel, P.H. et al., 1984. Protein-Nucleic Acid Interactions in Transcription: A Molecular Analysis. *Annual Review of Biochemistry*, 53, pp.389–446.
- Hoch, J.A., 2000. Two-component and phosphorelay signal transduction. *Current Opinion in Microbiology*, 3(2), pp.165–170.
- Hocine, S. et al., 2013. Single-molecule analysis of gene expression using two-color RNA labeling in live yeast. *Nature Methods*, 10(2), pp.119–21.
- Hogg, R.W. & Englesberg, E., 1969. L-arabinose binding protein from *Escherichia coli* B/r. *Journal of Bacteriology*, 100(1), pp.423–32.
- Holmes, V.F. & Cozzarelli, N.R., 2000. Closing the ring: Links between SMC proteins and chromosome partitioning, condensation, and supercoiling. *Proceedings of the National Academy of Sciences*, 97(4), pp.1322–1324.
- Hsu, L.M., 2002. Promoter clearance and escape in prokaryotes. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1577(2), pp.191–207.
- Huang, B., Bates, M. & Zhuang, X., 2009. Super-Resolution Fluorescence Microscopy. *Annual Review of Biochemistry*, 78(1), pp.993–1016.
- Huh, D. & Paulsson, J., 2011. Non-genetic heterogeneity from stochastic partitioning at cell division. *Nature Genetics*, 43(2), pp.95–100.
- Hussain, F. et al., 2014. Engineered temperature compensation in a synthetic genetic clock. *Proceedings of the National Academy of Sciences of the United States of America*, 111(3), pp.972–7.
- Isaacs, F.J. et al., 2004. Engineered riboregulators enable post-transcriptional control of gene expression. *Nature Biotechnology*, 22(7), pp.841–547.
- Isaacs, F.J. et al., 2003. Prediction and measurement of an autoregulatory genetic module. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13),

pp.7714–9.

Isaacs, F.J., 2012. Synthetic biology: Automated design of RNA devices. *Nature Chemical Biology*, 8(5), pp.413–415.

Ishihama, A., 2000. Functional modulation of *Escherichia coli* RNA polymerase. *Annual Review of Microbiology*, 54, pp.499–518.

Jacob, F. & Monod, J., 1961. Genetic regulatory mechanism in the synthesis of proteins. *Journal of Molecular Biology*, 3(3), pp.318–356.

Jana, N.K. et al., 1999. Amino acid changes in the repressor of bacteriophage lambda due to temperature-sensitive mutations in its cI gene and the structure of a highly temperature-sensitive mutant repressor. *Protein Engineering*, 12(3), pp.225–33.

Jin, D.J., Cagliero, C. & Zhou, Y.N., 2013. Role of RNA polymerase and transcription in the organization of the bacterial nucleoid. *Chemical Reviews*, 113(11), pp.8662–8682.

Johnson, C.M. & Schleif, R.F., 1995. *In vivo* induction kinetics of the arabinose promoters in *Escherichia coli*. *Journal of Bacteriology*, 177(12), p.3438.

Jones, D.L., Brewster, R.C. & Phillips, R., 2014. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*, 346(6216), pp.1533–1536.

Jones, K.L., Kim, S.W. & Keasling, J.D., 2000. Low-copy plasmids can perform as well as or better than high-copy plasmids for metabolic engineering of bacteria. *Metabolic Engineering*, 2(4), pp.328–38.

Jones, P.G., VanBogelen, R.A. & Neidhardt, F.C., 1987. Induction of proteins in response to low temperature in *Escherichia coli*. *Journal of Bacteriology*, 169(5), pp.2092–2095.

Kaern, M. et al., 2005. Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, 6, pp.451–64.

Kahl, L.J. & Endy, D., 2013. A survey of enabling technologies in synthetic biology. *Journal of Biological Engineering*, 7(13), pp.1–18.

Kandavalli, V.K., Tran, H. & Ribeiro, A.S., 2016. Effects of σ factor competition are promoter initiation kinetics dependent. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1859(10), pp.1281–1288.

Kannan, G. et al., 2008. Rapid acid treatment of *Escherichia coli*: transcriptomic response and recovery. *BMC Microbiology*, 8(37), pp.1–13.

Kapanidis, A.N. et al., 2006. Initial transcription by RNA polymerase proceeds through a DNA-scrunching mechanism. *Science*, 314(5802), pp.1144–1147.

Kauffman, S., 1969. Metabolic stability and epigenesis in randomly constructed genetic nets. *Journal of Theoretical Biology*, 22(3), pp.437–67.

Kennel, D. & Riezman, H., 1977. Transcription and translation initiation frequencies of the *E.*

- coli lac* operon. *Journal of Molecular Biology*, 114(1), pp.1–21.
- Kerfeld, C.A., Heinhorst, S. & Cannon, G.C., 2010. Bacterial microcompartments. *Annual Review of Microbiology*, 64, pp.391–408.
- Keryer-Bibens, C., Barreau, C. & Osborne, H.B., 2008. Tethering of proteins to RNAs by bacteriophage proteins. *Biology of the Cell*, 100(2), pp.125–138.
- Khalil, A.S. et al., 2012. A synthetic biology framework for programming eukaryotic transcription functions. *Cell*, 150(3), pp.647–58.
- Kittleson, J.T., Cheung, S. & Anderson, J.C., 2011. Rapid optimization of gene dosage in *E. coli* using DIAL strains. *Journal of Biological Engineering*, 5(10), pp.1–7.
- Klappenbach, J.A., Dunbar, J.M. & Schmidt, T.M., 2000. rRNA operon copy number reflects ecological strategies of bacteria. *Applied and Environmental Microbiology*, 66(4), pp.1328–1333.
- Klump, S., Zhang, Z. & Hwa, T., 2009. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell*, 139(7), pp.1366–75.
- Ko, C.H. & Takahashi, J.S., 2006. Molecular components of the mammalian circadian clock. *Human Molecular Genetics*, 15(SUPPL. 2), pp.271–277.
- Koblan, K.S. & Ackers, G.K., 1991. Cooperative protein-DNA interactions: effects of KCl on lambda cI binding to OR. *Biochemistry*, 30(31), pp.7822–7.
- Koblan, K.S. & Ackers, G.K., 1991. Energetics of Subunit Dimerization in Bacteriophage λ cI Repressor: Linkage to Protons, Temperature, and KCl. *Biochemistry*, 30(31), pp.7817–7821.
- Konopka, C.A. & Bednarek, S.Y., 2008. Variable-angle epifluorescence microscopy: A new way to look at protein dynamics in the plant cell cortex. *Plant Journal*, 53(1), pp.186–196.
- Konopka, M.C. et al., 2009. Cytoplasmic protein mobility in osmotically stressed *Escherichia coli*. *Journal of Bacteriology*, 91(1), pp.231–237.
- Kontur, W.S. et al., 2008. Late Steps in the Formation of *E. coli* RNA Polymerase-lambdaPR Promoter Open Complexes: Characterization of Conformational Changes by Rapid [Perturbant] Upshift Experiments. *Journal of Molecular Biology*, 376(4), pp.1034–1047.
- Kramer, B.P. et al., 2004. An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology*, 22, pp.867–870.
- Lajoie, M.J. et al., 2013. Genomically Recoded Organisms Expand Biological Functions. *Science*, 342(6156), pp.357–360.
- Larson, M.H., Landick, R. & Block, S.M., 2011. Single-molecule studies of RNA polymerase: one singular sensation, every little step it takes. *Molecular Cell*, 41(3), pp.249–62.
- Lebar, T. et al., 2014. A bistable genetic switch based on designable DNA-binding domains. *Nature Communications*, 5, pp.1–13.

- Lederberg, E.M. & Lederberg, J., 1953. Genetic Studies of Lysogenicity in *Escherichia coli*. *Genetics*, 38(1), pp.51–64.
- Lee, J.H., Al-Zarban, S. & Wilcox, G., 1981. Genetic characterization of the *araE* gene in *Salmonella typhimurium* lt2. *Journal of Bacteriology*, 146(1), pp.298–304.
- Leibler, S. & Kussell, E., 2010. Individual histories and selection in heterogeneous populations. *Proceedings of the National Academy of Sciences of the United States of America*, 107(29), pp.13183–8.
- Leng, F., Chen, B. & Dunlap, D.D., 2011. Dividing a supercoiled DNA molecule into two independent topological domains. *Proceedings of the National Academy of Sciences*, 108(50), pp.19973–19978.
- Lentini, R. et al., 2013. Fluorescent proteins and in vitro genetic organization for cell-free synthetic biology. *ACS Synthetic Biology*, 2(9), pp.482–489.
- Levine, E. et al., 2007. Quantitative characteristics of gene regulation by small RNA. *PLoS Biology*, 5(9), p.e229.
- Levine, J.H., Lin, Y. & Elowitz, M.B., 2013. Functional roles of pulsing in genetic circuits. *Science*, 342(6163), pp.1193–200.
- Levsky, J.M., 2003. Fluorescence in situ hybridization: past, present and future. *Journal of Cell Science*, 116(14), pp.2833–2838.
- Lewin, B., 2008. *Genes IX* 9th ed., Sudbury, MA, USA: Jones and Bartlett.
- Lewis, M. et al., 1996. Crystal Structure of the Lactose Operon Repressor and Its Complexes with DNA and Inducer. *Science*, 271(5253), pp.1247–1254.
- Lewis, M., 2005. The lac repressor. *Comptes Rendus Biologies*, 328(6), pp.521–48.
- Lindner, A.B. et al., 2008. Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proceedings of the National Academy of Sciences*, 105(8), pp.3076–3081.
- Lineweaver, H. & Burk, D., 1934. The Determination of Enzyme Dissociation Constants. *Journal of the American Chemical Society*, 56(3), pp.658–666.
- Lipshtat, A. et al., 2006. Genetic Toggle Switch without Cooperative Binding. *Physical Review Letters*, 96(188101), pp.1–4.
- Litcofsky, K., Afeyan, R. & Krom, R., 2012. Iterative plug-and-play methodology for constructing and midifying synthetic gene networks. *Nature*, 9(11), pp.1077–1080.
- Livak, K.J. & Schmittgen, T.D., 2001. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods*, 25(4), pp.402–408.
- Lloyd-price, J. et al., 2016. Dissecting the stochastic transcription initiation process in live *Escherichia coli*. *DNA Research*, 23(3), pp.203–214.

- Lobell, R. & Schleif, R., 1990. DNA looping and unlooping by AraC protein. *Science*, 250(4980), pp.528–532.
- Loinger, A. et al., 2007. Stochastic simulations of genetic switch systems. *Physical Review E*, 75(051917), pp.1–14.
- López-García, P. & Forterre, P., 2000. DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *BioEssays*, 22(8), pp.738–746.
- Los, D.A., 2004. The effect of low-temperature-induced DNA supercoiling on the expression of the desaturase genes in *synechocystis*. *Cellular and Molecular Biology*, 50(5), pp.605–612.
- Lucks, J.B. et al., 2011. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences*, 108(21), pp.8617–8622.
- Lutz, R. et al., 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), pp.3873–81.
- Lutz, R. & Bujard, H., 1997. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*, 25(6), pp.1203–10.
- Mäkelä, J., 2016. *On the Variability of the Transcription Process in Escherichia coli*. PhD thesis, Tampere University of Technology.
- Mäkelä, J., Kandavalli, V. & Ribeiro, A.S., 2017. Rate-limiting steps in transcription dictate sensitivity to variability in cellular components. *Scientific Reports*, 7(1), pp.1–10.
- Marchisio, M.A. & Stelling, J., 2009. Computational design tools for synthetic biology. *Current Opinion in Biotechnology*, 20(4), pp.479–485.
- Martin, V.J.J. et al., 2003. Engineering a mevalonate pathway in *Escherichia coli* for production of terpenoids. *Nature Biotechnology*, 21, pp.796–802.
- Martínez-Antonio, A. et al., 2012. Hierarchical dynamics of a transcription factors network in *E. coli*. *Molecular BioSystems*, 8, pp.2932–2936.
- Martinez, N.J. & Walhout, A.J.M., 2009. The interplay between transcription factors and microRNAs in genome-scale regulatory networks. *BioEssays*, 31(4), pp.435–45.
- Mashaghi, A. & Dekker, C., 2014. Systems and synthetic biology approaches to cell division. *Systems and Synthetic Biology*, 8(3), pp.173–178.
- McClure, W.R., 1985. Mechanism and control of transcription initiation in prokaryotes. *Annual Review of Biochemistry*, 54(1), pp.171–204.
- McClure, W.R., 1980. Rate-limiting steps in RNA chain initiation. *Proceedings of the National Academy of Sciences*, 77(10), pp.5634–8.
- McClure, W.R., Cech, C.L. & Johnston, D.E., 1978. A steady state assay for the RNA polymerase

- initiation reaction. *The Journal of biological chemistry*, 253(24), pp.8941–8948.
- McMillen, D. et al., 2002. Synchronizing genetic relaxation oscillators by intercell signaling. *Proceedings of the National Academy of Sciences of the United States of America*, 99(2), pp.679–84.
- Megerle, J.A. et al., 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. *Biophysical Journal*, 95(4), pp.2103–15.
- Mihalcescu, I., Hsing, W. & Leibler, S., 2004. Resilient circadian oscillator revealed in individual cyanobacteria. *Nature*, 430, pp.81–85.
- Mika, J.T. et al., 2010. Molecular sieving properties of the cytoplasm of *Escherichia coli* and consequences of osmotic stress. *Molecular Microbiology*, 77(1), pp.200–207.
- Mileyko, Y., Joh, R.I. & Weitz, J.S., 2008. Small-scale copy number variation and large-scale changes in gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 105(43), pp.16659–64.
- Mileyko, Y. & Weitz, J.S., 2010. Bifurcation Analysis of Gene Regulatory Circuits Subject to Copy Number Variation. *SIAM Journal on Applied Dynamical Systems*, 9(3), pp.799–826.
- Miller, J.H., 1980. The lacI Gene: Its Role in lac Operon Control and Its Use as a Genetic System. In *The operon*. Cold Spring Harbor Monograph Archive, pp. 31–86.
- Miller, O.L., Hamkalo, B.A. & Thomas, C.A., 1970. Visualization of Bacterial Genes in Action. *Science*, 169(3943), pp.392–395.
- Million-Weaver, S. & Camps, M., 2014. Mechanisms of plasmid segregation: have multicopy plasmids been overlooked? *Plasmid*, 0(9), pp.27–36.
- Mitarai, N., Sneppen, K. & Pedersen, S., 2008. Ribosome collisions and translation efficiency: optimization by codon usage and mRNA destabilization. *Journal of Molecular Biology*, 382(1), pp.236–45.
- Miyamoto, T. et al., 2013. Synthesizing Biomolecule-based Boolean Logic Gates. *ACS Synthetic Biology*, 2(2), pp.72–82.
- Molin, S. et al., 1980. Runaway replication of plasmid R1 is not caused by loss of replication inhibitor activity of gene cop. *Journal of Bacteriology*, 143(2), pp.1046–8.
- Moon, T.S. et al., 2011. Construction of a genetic multiplexer to toggle between chemosensory pathways in *Escherichia coli*. *Journal of Molecular Biology*, 406(2), pp.215–227.
- Moon, T.S. et al., 2012. Genetic programs constructed from layered logic gates in single cells. *Nature*, 491(7423), pp.249–253.
- Mora, A.D. et al., 2011. Automated drusen detection in retinal images using analytical modelling algorithms. *Biomedical Engineering Online*, 10(59), pp.1–15.
- More, J.J., 1978. The Levenberg-Marquardt algorithm: implementation and theory. In G. A.

- Watson, ed. *Numerical Analysis*. Heidelberg: Springer-Verlag, pp. 105–116.
- Moser, F. et al., 2012. Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synthetic Biology*, 1(11), pp.555–64.
- Moulin, L., Rahmouni, A.R. & Boccard, F., 2005. Topological insulators inhibit diffusion of transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Molecular Microbiology*, 55(2), pp.601–610.
- Murakami, K.S. et al., 2002. Structural basis of transcription initiation: An RNA polymerase holoenzyme-DNA complex. *Science*, 296(5571), pp.1285–1290.
- Mutalik, V.K. et al., 2013. Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, 10, pp.1–15.
- Mutalik, V.K. et al., 2012. Rationally designed families of orthogonal RNA regulators of translation. *Nature Chemical Biology*, 8(5), pp.447–54.
- Muthukrishnan, A., 2014. *Studies of the Plasticity of Transcription in Escherichia coli using Single-Molecule , in vivo Detection Techniques*. PhD thesis, Tampere University of Technology.
- 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), p.e109005.
- Nandagopal, N. & Elowitz, M.B., 2011. Synthetic biology: integrated gene circuits. *Science*, 333(6047), pp.1244–8.
- Neice, A., 2010. Methods and Limitations of Subwavelength Imaging. In *Advances in Imaging and Electron Physics*. pp. 117–140.
- Newman, J.R.S. et al., 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, 441(7095), pp.840–6.
- Nielsen, A.K. et al., 2016. Genetic circuit design automation. *Science*, 352(6281), pp.53–64.
- Norman, T.M. et al., 2015. Stochastic Switching of Cell Fate in Microbes. *Annual Review of Microbiology*, 69(1), pp.381–403.
- Nudler, E. & Gottesman, M.E., 2002. Transcription termination and anti-termination in *Escherichia coli*. *Genes to cells*, 7(8), pp.755–68.
- Oehler, S., 2009. Feedback regulation of Lac repressor expression in *Escherichia coli*. *Journal of Bacteriology*, 191(16), pp.5301–3.
- Oehler, S. et al., 1994. Quality and position of the three lac operators of *E. coli* define efficiency of repression. *The EMBO Journal*, 13(14), pp.3348–3355.
- Oehler, S. et al., 1990. The three operators of the lac operon cooperate in repression. *The EMBO Journal*, 9(4), pp.973–979.

- Ogura, T. & Hiraga, S., 1983. Mini-F plasmid genes that couple host cell division to plasmid proliferation. *Proceedings of the National Academy of Sciences*, 80(15), pp.4784–4788.
- Ohshima, Y., Mizokoshi, T. & Horiuchi, T., 1974. Binding of an inducer to the lac repressor. *Journal of Molecular Biology*, 89(1), pp.127–136.
- Oleksiuk, O. et al., 2011. Thermal robustness of signaling in bacterial chemotaxis. *Cell*, 145(2), pp.312–321.
- Olson, E.J. & Tabor, J.J., 2012. Post-translational tools expand the scope of synthetic biology. *Current Opinion in Chemical Biology*, 16(3–4), pp.300–306.
- Ortín, J. & Parra, F., 2006. Structure and Function of RNA Replication. *Annual Review of Microbiology*, 60(1), pp.305–326.
- Ozbudak, E.M. et al., 2004. Multistability in the lactose utilisation network of *Escherichia coli*. *Nature*, 427(6976), pp.737–740.
- Paddon, C.J. et al., 2013. High-level semi-synthetic production of the potent antimalarial artemisinin. *Nature*, 496(7446), pp.528–532.
- Parry, B.R. et al., 2014. The Bacterial Cytoplasm Has Glass-like Properties and Is Fluidized by Metabolic Activity. *Cell*, 156, pp.183–194.
- Patrick, M. et al., 2015. Free RNA polymerase in *Escherichia coli*. *Biochimie*, 119, pp.80–91.
- Paulsson, J., 2005a. Models of stochastic gene expression. *Physics of Life Reviews*, 2(5), pp.157–175.
- Paulsson, J., 2005b. Prime movers of noisy gene expression. *Nature Genetics*, 37(9), pp.925–926.
- Paulsson, J., 2004. Summing up the noise in gene networks. *Nature*, 427(29), pp.415–418.
- Paulsson, J. & Ehrenberg, M., 2001. Noise in a minimal regulatory network: plasmid copy number control. *Quarterly Reviews of Biophysics*, 34(1), pp.1–59.
- Pawley, J.E., 2006. *Handbook of Biological Confocal Microscopy* 3rd ed. Springer Science + Business Media, ed., New York.
- Peabody, D.S., 1993. The RNA binding site of bacteriophage MS2 coat protein. *The EMBO Journal*, 12(2), pp.595–600.
- Peabody, D.S. & Lim, F., 1996. Complementation of RNA binding site mutations in MS2 coat protein heterodimers. *Nucleic Acids Research*, 24(12), pp.2352–2359.
- Peccoud, J., 1995. Markovian Modelling of Gene Product Synthesis. *Theoretical Population Biology*, 48, pp.222–234.
- Phadtare, S., 2004. Recent developments in bacterial cold-shock response. *Current Issues in Molecular Biology*, 6(2), pp.125–136.

- Pingoud, A. & Jeltsch, A., 2001. Structure and function of type II restriction endonucleases. *Nucleic Acids Research*, 29(18), pp.3705–27.
- Pitchiaya, S. et al., 2014. Single Molecule Fluorescence Approaches Shed Light on Intracellular RNAs. *Chemical Reviews*, 114(6), pp.3224–3265.
- Pomerening, J.R., Kim, S.Y. & Jr., J.E.F., 2005. Systems-level dissection of the cell-cycle oscillator: bypassing positive feedback produces damped oscillations. *Cell*, 122(4), pp.565–78.
- Pomerening, J.R., Ubersax, J.A. & Jr., J.E.F., 2008. Rapid Cycling and Precocious Termination of G1 Phase in Cells Expressing CDK1AF. *Molecular Biology of the Cell*, 19(08), pp.33426–3441.
- Postow, L. et al., 2004. Topological domain structure of the *Escherichia coli* chromosome. *Genes & Development*, 18, pp.1766–1779.
- Potvin-Trottier, L. et al., 2016. Synchronous long-term oscillations in a synthetic gene circuit. *Nature*, 538(7626), pp.1–4.
- Prasher, D.C. et al., 1992. Aequorea victoria. *Gene*, 111, pp.229–233.
- Prindle, A. et al., 2014. Rapid and tunable post-translational coupling of genetic circuits. *Nature*, 508(7496), pp.387–391.
- 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 of the United States of America*, 83(23), pp.8952–6.
- Qi, L.S. & Arkin, A.P., 2014. A versatile framework for microbial engineering using synthetic non-coding RNAs. *Nature Reviews Microbiology*, 12, pp.341–354.
- Rahmouni, A.R. & Wells, R.D., 1992. Direct evidence for the effect of transcription on local DNA supercoiling *in vivo*. *Journal of Molecular Biology*, 223(1), pp.131–44.
- Raj, A. et al., 2008. Imaging individual mRNA molecules using multiple singly labeled probes. *Nature Methods*, 5(10), pp.877–879.
- Raj, A. & Oudenaarden, A. Van, 2008. Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell*, 135(10), pp.216–226.
- Rajala, T. et al., 2010. Effects of Transcriptional Pausing on Gene Expression Dynamics. *PLoS Biology*, 6(3), pp.29–30.
- Ramakrishnan, V., 2002. Ribosome Structure and the Mechanism of Translation. *Cell*, 108, pp.557–572.
- Rao, C. V., 2012. Expanding the synthetic biology toolbox: Engineering orthogonal regulators of gene expression. *Current Opinion in Biotechnology*, 23(5), pp.689–694.
- Revyakin, A. et al., 2006. Abortive initiation and productive initiation by RNA polymerase

- involve DNA scrunching. *Science*, 314(5802), pp.1139–1143.
- Reznikoff, W.S., Winter, R.B. & Hurley, C.K., 1974. The location of the repressor binding sites in the lac operon. *Proceedings of the National Academy of Sciences of the United States of America*, 71(6), pp.2314–2318.
- Ribeiro, A.S., 2007a. A Model of Genetic Networks with Delayed Stochastic Dynamics. In F. Emmert-Streib & M. Dehmer, eds. *Analysis of Microarray Data: A Network-Based Approach*. pp. 1–30.
- Ribeiro, A.S., 2007b. Dynamics of a two-dimensional model of cell tissues with coupled stochastic gene networks. *Physical Review E*, 76(051915), pp.1–14.
- Ribeiro, A.S., 2007c. Effects of coupling strength and space on the dynamics of coupled toggle switches in stochastic gene networks with multiple-delayed reactions. *Physical Review E*, 75(061903), pp.1–12.
- Ribeiro, A.S., 2010. Stochastic and delayed stochastic models of gene expression and regulation. *Mathematical Biosciences*, 223(1), pp.1–11.
- Ribeiro, A.S. & Kauffman, S.A., 2007. Noisy attractors and ergodic sets in models of gene regulatory networks. *Journal of Theoretical Biology*, 247, pp.743–755.
- Ribeiro, A.S., Zhu, R. & Kauffman, S.A., 2006. A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics. *Journal of Computational Biology*, 13(9), pp.1630–1639.
- Richter, K., Haslbeck, M. & Buchner, J., 2010. The Heat Shock Response: Life on the Verge of Death. *Molecular Cell*, 40(2), pp.253–266.
- Riggs, A.D., Suzuki, H. & Bourgeois, S., 1970. Lac repressor-operator interaction: I. Equilibrium studies. *Journal of Molecular Biology*, 48(1), pp.67–83.
- Roberts, E. et al., 2011. Noise Contributions in an Inducible Genetic Switch: A Whole-Cell Simulation Study. *PLoS Computational Biology*, 7(3), p.e1002010.
- Roberts, R.J. & Murray, K., 1976. Restriction Endonuclease. *CRC Critical Reviews in Biochemistry*, 4(2), pp.123–164.
- Rodrigo, G. et al., 2012. Perspectives on the automatic design of regulatory systems for synthetic biology. *FEBS Letters*, 586(15), pp.2037–2042.
- Roe, J.-H. et al., 1985. Temperature Dependence of the Rate Constants of the *Escherichia coli* RNA Polymerase-Lambda Pr Promoter Interaction Assignment of the Kinetic Steps Corresponding to Protein Conformational Change and DNA Opening. *Journal of Molecular Biology*, 184, pp.441–453.
- Roussel, M.R. & Zhu, R., 2006. Stochastic kinetics description of a simple transcription model. *Bulletin of Mathematical Biology*, 68(7), pp.1681–1713.
- Rouvière, P.E. et al., 1995. rpoE, the gene encoding the second heat-shock sigma factor, sigma E,

- in *Escherichia coli*. *The EMBO Journal*, 14(5), pp.1032–1042.
- Rovinskiy, N. et al., 2012. Rates of Gyrase Supercoiling and Transcription Elongation Control Supercoil Density in a Bacterial Chromosome. *PLoS Genetics*, 8(8), p.e1002845.
- Rutkauskas, D. et al., 2009. Tetramer opening in LacI-mediated DNA looping. *Proceedings of the National Academy of Sciences of the United States of America*, 106(39), pp.16627–32.
- Sabate, R., De Groot, N.S. & Ventura, S., 2010. Protein folding and aggregation in bacteria. *Cellular and Molecular Life Sciences*, 67(16), pp.2695–2715.
- Sadowski, P.D., Zhu, X.-D. & Sadowski, P.D., 1995. Cleavage-dependent Ligation by the FLP Recombinase. *Journal of Biological Chemistry*, 270(39), pp.23044–23054.
- 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), pp.754–771.
- Samad, H. El et al., 2005. Repressilators and promotilators: loop dynamics in synthetic gene networks. *Proceedings of the 2005, American Control Conference, 2005.*, pp.4405–4410.
- Samoilov, M., Arkin, A. & Ross, J., 2002. Signal Processing by Simple Chemical Systems. *Journal of Physical Chemistry A*, 106(43), pp.10205–10221.
- 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), pp.925–935.
- Sanchez, A. et al., 2011. Mechanism of transcriptional repression at a bacterial promoter by analysis of single molecules. *The EMBO Journal*, (07), pp.1–7.
- Sanchez, A. & Golding, I., 2013. Genetic determinants and cellular constraints in noisy gene expression. *Science*, 342(6163), pp.1188–1193.
- Sander, J.D. & Joung, J.K., 2014. CRISPR-Cas systems for editing, regulating and targeting genomes. *Nature Biotechnology*, 32(4), pp.347–350.
- Santillán, M. & Mackey, M.C., 2004. Why the lysogenic state of phage lambda is so stable: a mathematical modeling approach. *Biophysical Journal*, 86(1), pp.75–84.
- Schlx, P.J., Capp, M.W. & Jr, M.T.R., 1995. Inhibition of Transcription Initiation by lac Repressor. *Journal of Molecular Biology*, 245, pp.331–350.
- Schleif, R., 2010. AraC protein, regulation of the l-arabinose operon in *Escherichia coli*, and the light switch mechanism of AraC action. *FEMS Microbiology Reviews*, 34(5), pp.779–796.
- Schleif, R., 2000a. DNA looping and regulation of the arabinose operon. *Harvey Lectures*, 84, pp.27–39.
- Schleif, R., 2002. Regulation of the L-arabinose operon in *Escherichia coli*. In D. A. Hodgson &

- C. M. Thomas, eds. *SGM Symposium 61: Signals, switches, regulons and cascades: control of bacterial gene expression*. Cambridge University Press, pp. 156–168.
- Schleif, R., 2000b. Regulation of the L-arabinose operon of *Escherichia coli*. *Trends in Genetics*, 16(12), pp.559–565.
- Segel, I.H., 1975. *Enzyme Kinetics, Behavior and Analysis of Rapid Equilibrium and Steady-State Enzyme Systems*, New York, New York, USA: John Wiley & Sons.
- Shaner, N.C. et al., 2004. Improved monomeric red, orange and yellow fluorescent proteins derived from *Discosoma* sp. red fluorescent protein. *Nature Biotechnology*, 22(12), pp.1567–1572.
- Shih, Y.-L., Le, T. & Rothfield, L., 2003. Division site selection in *Escherichia coli* involves dynamic redistribution of Min proteins within coiled structures that extend between the two cell poles. *Proceedings of the National Academy of Sciences of the United States of America*, 100(13), pp.7865–70.
- Shuman, H.A. & Silhavy, T.J., 2003. Microbial genetics: The art and design of genetic screens: *Escherichia coli*. *Nature Reviews Genetics*, 4(6), pp.419–431.
- Siuti, P., Yazbek, J. & Lu, T.K., 2013. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology*, 31(5), pp.448–452.
- Snijder, B. et al., 2009. Population context determines cell-to-cell variability in endocytosis and virus infection. *Nature*, 461(7263), pp.520–523.
- So, L. et al., 2011. General properties of transcriptional time series in *Escherichia coli*. *Nature Genetics*, (4), pp.1–9.
- Speer, M.A. & Richard, T.L., 2011. Amplified insert assembly: An optimized approach to standard assembly of BioBrick TM genetic circuits. *Journal of Biological Engineering*, 5(17), pp.1–10.
- Sprinzak, D. & Elowitz, M.B., 2005. Reconstruction of genetic circuits. *Nature*, 438(11), pp.443–448.
- Stevenson, B.S. & Schmidt, T.M., 2004. Life History Implications of rRNA Gene Copy Number in *Escherichia coli*. *Society*, 70(11), pp.6670–6677.
- Stewart-ornstein, J., Weissman, J.S. & El-samad, H., 2012. Article Cellular Noise Regulons Underlie Fluctuations in *Saccharomyces cerevisiae*. *Molecular Cell*, 45(4), pp.483–493.
- Straight, P.D. et al., 2007. A singular enzymatic megacomplex from *Bacillus subtilis*. *Proceedings of the National Academy of Sciences of the United States of America*, 104(1), pp.305–310.
- Straney, S.B. & Crothers, D.M., 1987. Kinetics of the Stages of Transcription Initiation at the *Escherichia coli* lac UV5 Promoter. *Biochemistry*, 26(16), pp.5063–5070.
- Stricker, J. et al., 2008. A fast, robust and tunable synthetic gene oscillator. *Nature*, 456(7221),

pp.516–9.

- Studier, F. & Moffatt, B., 1986. Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *Journal of Molecular Biology*, 189(1), pp.113–130.
- Stuger, R. et al., 2002. DNA supercoiling by gyrase is linked to nucleoid compaction. *Molecular Biology Reports*, 29, pp.79–82.
- Stylianidou, S., Kuwada, N.J. & Wiggins, P.A., 2015. Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility. *Biophysical Journal*, 107(11), pp.2664–2692.
- Süel, G.M. et al., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature*, 440(23), pp.545–550.
- Susa, M., Kubori, T. & Shimamoto, N., 2006. A pathway branching in transcription initiation in *Escherichia coli*. *Molecular Microbiology*, 59(6), pp.1807–1817.
- Tamsir, A., Tabor, J.J. & Voigt, C.A., 2011. Robust multicellular computing using genetically encoded NOR gates and chemical ‘wires.’ *Nature*, 469(7329), pp.212–215.
- Taniguchi, Y. et al., 2010. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*, 329(5991), pp.533–538.
- Thomas, C.M. & Summers, D., 2008. *Bacterial Plasmids* 2nd ed., John Wiley & Sons.
- Togna, A.P., Shuler, M.L. & Wilson, D.B., 1993. Effects of Plasmid Copy Number and Runaway Plasmid Replication on Overproduction and Excretion of β -Lactamase from *Escherichia coli*. *Biotechnology Progress*, 9(1), pp.31–39.
- Touchon, M. et al., 2009. Organised genome dynamics in the *Escherichia coli* species results in highly diverse adaptive paths. *PLoS Genetics*, 5(1), p.e1000344.
- Tran, H. et al., 2015. Kinetics of the cellular intake of a gene expression inducer at high concentrations. *Molecular BioSystems*, 11(9), pp.2579–2587.
- Tsien, R.Y., 1998. The green fluorescent protein. *Annual Review of Biochemistry*, 67, pp.509–44.
- Tuttle, L.M. et al., 2005. Model-driven designs of an oscillating gene network. *Biophysical Journal*, 89(6), pp.3873–83.
- Tyedmers, J., Mogk, A. & Bukau, B., 2010. Cellular strategies for controlling protein aggregation. *Nature Publishing Group*, 11(11), pp.777–788.
- Tyson, J.J., Chen, K. & Novak, B., 2001. Network dynamics and cell physiology. *Nature Reviews Molecular Cell Biology*, 2(12), pp.908–916.
- Ullmann, A., Jacob, F. & Monod, J., 1967. Characterization by in vitro complementation of a peptide corresponding to an operator-proximal segment of the β -galactosidase structural gene of *Escherichia coli*. *Journal of Molecular Biology*, 24(2), pp.339–343.
- Vecchiarelli, A.G., Mizuuchi, K. & Funnell, B.E., 2012. Surfing Biological surfaces: Exploiting

- the nucleoid for partition and transport in bacteria. *Molecular Microbiology*, 86(3), pp.513–523.
- Veliz-Cuba, A. et al., 2015. Sources of Variability in a Synthetic Gene Oscillator. *PLoS Computational Biology*, 11(12), pp.1–23.
- Venter, J.C. et al., 2001. The sequence of the human genome. *Science*, 291(5507), pp.1304–1351.
- Verghese, J. et al., 2012. Biology of the Heat Shock Response and Protein Chaperones: Budding Yeast (*Saccharomyces cerevisiae*) as a Model System. *Microbiology and Molecular Biology Reviews*, 76(2), pp.115–158.
- Viaplana, E. et al., 1997. Reversible activation of a cryptic cleavage site within *E. coli* beta-galactosidase in beta-galactosidase fusion proteins. *Biochimica et Biophysica Acta*, 1343(2), pp.221–6.
- Voigt, C.A., 2006. Genetic parts to program bacteria. *Current Opinion in Biotechnology*, 17, pp.548–557.
- Walter, G. et al., 1967. Initiation of DNA-Dependent RNA Synthesis and the Effect of Heparin on RNA Polymerase. *European Journal of Biochemistry*, 3(7), pp.194–201.
- Walter, N.G. et al., 2008. Do-it-yourself guide: How to use the modern single-molecule toolkit. *Nature Methods*, 5(6), pp.475–489.
- Wang, B. et al., 2011. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nature Communications*, 2(508), pp.1–9.
- Weber, E. et al., 2011. A Modular Cloning System for Standardized Assembly of Multigene Constructs. *PLoS ONE*, 6(2), p.e16765.
- Weiss, R.O.N. et al., 2003. Genetic circuit building blocks for cellular computation, communications, and signal processing. *Natural Computing*, 2(1), pp.47–84.
- Wickner, S., Maurizi, M.R. & Gottesman, S., 1999. Posttranslational quality control: folding, refolding, and degrading proteins. *Science*, 286(5446), pp.1888–1893.
- Winkler, J. et al., 2010. Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular ageing. *The EMBO Journal*, 29(5), pp.910–923.
- Wolf, D.M. & Arkin, A.P., 2003. Motifs, modules and games in bacteria. *Current Opinion in Microbiology*, 6(6), pp.125–134.
- Wu, B. et al., 2011. Modern fluorescent proteins and imaging technologies to study gene expression, nuclear localization, and dynamics. *Current Opinion in Cell Biology*, 23(3), pp.310–317.
- Wu, B., Chao, J.A. & Singer, R.H., 2012. Fluorescence Fluctuation Spectroscopy Enables Quantitative Imaging of Single mRNAs in Living Cells. *Biophysical Journal*, 102(12), pp.2936–2944.

- Xia, B. et al., 2011. Chapter five - Developer's and User's Guide to Clotho v2.0: A Software Platform for the Creation of Synthetic Biological Systems. *Methods in Enzymology*, 498, pp.97–135.
- Xie, X.S. et al., 2008. Single-Molecule Approach to Molecular Biology in Living Bacterial Cells. *Annual Review of Biophysics*, 37, pp.417–44.
- Yamanaka, K., 1999. Cold shock response in *Escherichia coli*. *Journal of Molecular Microbiology and Biotechnology*, 1(2), pp.193–202.
- Yang, S. et al., 2014. Contribution of RNA polymerase concentration variation to protein expression noise. *Nature Communications*, 5(476), pp.1–9.
- Young, B.A., Gruber, T.M. & Gross, C.A., 2002. Views of Transcription Initiation. *Cell*, 109, pp.417–420.
- Yu, J. et al., 2006. Probing gene expression in live cells, one protein molecule at a time. *Science*, 311(5767), pp.1600–1603.
- Zhang, F., Wen, Y. & Guo, X., 2014. CRISPR/Cas9 for genome editing: Progress, implications and challenges. *Human Molecular Genetics*, 23, pp.40–46.
- Zhang, Y. et al., 1998. A new logic for DNA engineering using recombination in *Escherichia coli*. *Nature Genetics*, 20, pp.123–128.
- Zhu et al., 2007. Efficiency, robustness, and stochasticity of gene regulatory networks in systems biology: λ switch as a working example. *Introduction to Systems Biology*, pp.336–371.
- Zhu, R. et al., 2007. Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models. *Journal of Theoretical Biology*, 246(2), pp.725–745.
- Zong, C. et al., 2010. Lysogen stability is determined by the frequency of activity bursts from the fate-determining gene. *Molecular Systems Biology*, (440), pp.1–12.
- Zopf, C.J. et al., 2013. Cell-Cycle Dependence of Transcription Dominates Noise in Gene Expression. *PLoS Computational Biology*, 9(7), pp.1–12.
- Zubay, G., Schwartz, D. & Beckwith, J., 1970. Mechanism of Activation of Catabolite-Sensitive Genes: A Positive Control System. *Proceedings of the National Academy of Sciences of the United States of America*, 66(1), pp.104–110.

PUBLICATION I

Effects of temperature on the dynamics of the LacI-TetR-CI Repressilator

J.G. Chandraseelan, S.M.D. Oliveira, A. Häkkinen, H. Tran, A. Sala, I. Potapov, M. Kandhavelu, and
A.S. Ribeiro

Mol. Biosyst., 9:3117-23, 2013.
doi: 10.1039/c3mb70203k

Publication reprinted with the permission of the copyright holders.

PAPER

Effects of temperature on the dynamics of the
LacI-TetR-CI repressilator

Cite this: *Mol. BioSyst.*, 2013,
9, 3117

Jerome G. Chandraseelan, Samuel M. D. Oliveira, Antti Häkkinen, Huy Tran,
Ilya Potapov, Adrien Sala, Meenakshisundaram Kandhavelu and Andre S. Ribeiro*

We studied the behaviour of the repressilator at 28 °C, 30 °C, 33 °C, and 37 °C. From the fluorescence in each cell over time, we determined the period of oscillations, the functionality (fraction of cells exhibiting oscillations) and the robustness (fraction of expected oscillations that occur) of this circuit. We show that the oscillatory dynamics differs with temperature. Functionality is maximized at 30 °C. Robustness decreases beyond 30 °C, as most cells exhibit 'failed' oscillations. These failures cause the distribution of periods to become bimodal, with an 'apparent period' that is minimal at 30 °C, while the true period decreases with increasing temperature. Based on previous studies, we hypothesized that the failures are due to a loss of functionality of one protein of the repressilator, CI. To test this, we studied the kinetics of a genetic switch, formed by the proteins CI and Cro, whose expression is controlled by P_{RM} and P_R , respectively. By probing the activity of P_{RM} by *in vivo* detection of MS2-GFP tagged RNA, we find that, beyond 30 °C, the production of the CI-coding RNA changes from sub-Poissonian to super-Poissonian. Given this, we suggest that the decrease in efficiency of CI as a repressor with temperature hinders the robustness of the repressilator beyond 30 °C. We conclude that the repressilator is sensitive but not robust to temperature. Replacing CI for a less temperature-dependent protein should enhance robustness.

Received 31st May 2013,
Accepted 27th September 2013

DOI: 10.1039/c3mb70203k

www.rsc.org/molecularbiosystems

Introduction

Natural genetic circuits can efficiently perform various tasks, such as time counting,¹ state holding,² and signal filtering,³ while maintaining robustness to environmental changes. This is necessary for them to be able to regulate complex cellular processes under various conditions^{4–6} or to efficiently determine cells' response to environmental shifts and signals. Much effort has been made to reproduce their behaviour in synthetic circuits.^{6–8} Once proven reliable, these synthetic circuits should have a wide range of applications.^{9–12} For example, synthetic genetic clocks promise to be of use as regulators of intracellular processes. For that, they will need to be robust to environmental changes, similarly to natural circuits.

One of the most famous synthetic circuits is the 'repressilator', engineered by Elowitz *et al.*⁷ This circuit has three genes, whose interactions form a negative feedback loop. Namely, the three genes form a cycle, with each gene expressing a protein that represses the next gene in the cycle. At 30 °C, the repressilator exhibits periodic oscillations,⁷ visible in time-lapse measurements

of a green fluorescent protein (GFP) reporter that is under the control of a promoter that is also present in the 3-gene circuit.

Temperature affects the dynamics of most cellular processes, including gene expression.¹³ Evidence suggests that natural, time-keeping circuits, such as circadian oscillators, evolved robustness to temperature fluctuations.^{14–16} Similar robustness is desired in synthetic circuits designed for time keeping.

The degree of robustness of the repressilator to temperature is unknown, but studies on some of its components suggest that its behaviour is bound to be strongly affected by small changes in temperature. For example, one of its proteins, the wild-type CI,⁷ has temperature-dependent DNA-binding stability.¹⁷ Namely, it is maximized at ~30 °C and is gradually lost as temperature increases, becoming ~50% weaker at 42 °C.¹⁷ This decrease may arise from the fact that the ability of CI to discriminate between operator sites depends on ion binding/release reactions¹⁸ and/or from the temperature-dependence of the CI's dimerization process.¹⁹

Here, we investigate how temperature affects the dynamics of the repressilator. Afterwards, we search for causes. Motivated by previous evidence that CI's functionality is temperature-dependent, we also study the temperature-dependence of another circuit, the CI–Cro switch. After comparing the effects of temperature on the kinetics of the two circuits, we propose

Computational Systems Biology Research Group, Tampere University of Technology,
P.O. Box 553, 33101 Tampere, Finland. E-mail: andre.ribeiro@tut.fi

modifications to the repressilator that may enhance its robustness to temperature fluctuations.

Methods

Repressilator: strain, plasmid, and microscopy

Cells of *E. coli* lac[−] strain MC 4100 with the repressilator (pZS1-ITrLLtCL) and the reporter plasmid (pZE21-GFPaav) were generously provided by M. Elowitz, Princeton University, NJ, USA. Minimal media were prepared with 2 mM MgSO₄·7H₂O (Sigma-Aldrich, USA), 7.6 mM (NH₄)₂SO₄ (Sigma Life Science, USA), 30 μM FeSO₄·7H₂O (Sigma Life Science, USA), 1 mM EDTA (Sigma Life Science, USA), 60 mM KH₂PO₄ (Sigma Life Science, USA) pH 6.8 with Glycerol 0.5% (Sigma Life Science, USA) and Casaminoacids 0.1% (Fluka Analytical, USA).

E. coli cells with the repressilator and reporter plasmids were grown in minimal media overnight at 28 °C, 30 °C, 33 °C or 37 °C with shaking at 300 rpm, to an optical density (OD) of 0.1 at 600 nm. Next, cells were diluted into fresh media and a few μl of the culture was placed between a cover-slip and a slab of 2% low melting agarose in minimal media, 0.75 mm thick. During time lapse microscopy, the temperature of the samples was kept stable by a control chamber (Bioptechs, FCS2, Pennsylvania, USA). Images were obtained every 15 minutes for 10 hours by a Nikon Eclipse (TE2000-U, Nikon, Tokyo, Japan) inverted C1 confocal laser-scanning system with a 100× Apo TIRF (1.49 NA, oil) objective. GFP fluorescence was measured using a 488 nm laser (Melles-Griot) and a 515/30 nm detection filter. For image acquisition, we used Nikon software EZ-C1.

Switch: strain, plasmid, and microscopy

E. coli CZ071 with a reporter plasmid PLtetO-1-MS2d-GFP and a target plasmid pIG-BAC (P_{RM}-limm(rexAB::bs48)) were generously provided by I. Golding (University of Illinois, USA). The target plasmid is a single-copy F-plasmid with a genetic switch coding for CI, under the control of P_{RM}, and Cro,²⁰ under the control of P_R. Further, the plasmid contains the immunity region of wild-type λ,²¹ where the rexA and rexB genes were replaced by a 48 binding site array for MS2d proteins, so as to detect individual RNAs whose production is controlled by P_{RM}. Depending on the occupation of the sites OR1, OR2 and OR3, one of the two promoters will be in a repressed state.^{32,33} Note that OR3 is absent in the repressilator. Nevertheless, the existence of oscillations⁷ shows that CI still achieves repression of P_R.

Cells were grown in Luria–Bertani (LB) medium with the following components: 10 g L^{−1} of Tryptone (Sigma Aldrich, USA), 5 g L^{−1} of yeast extract (LabM, UK) and 10 g L^{−1} of NaCl (LabM, UK), with addition of 34 μg ml^{−1} of Kanamycin and 34 μg ml^{−1} of Chloramphenicol (both antibiotics from Sigma Aldrich, USA). Cells were grown overnight with shaking at 260 rpm, in an orbital shaker (Labnet), at 30 °C for 12–16 h to an optical density (OD) of 0.1 at 600 nm. Thereafter, cells were grown until they reached an OD of ≈0.01 and diluted to 1:10 in LB medium with antibiotics. Then, they were grown at 37 °C with shaking at 260 rpm for a few hours, until they reached the exponential phase and an OD of ≈0.3.

The reporter gene, TetO1-MS2d, was activated using 10 ng ml^{−1} of anhydrotetracycline (aTc) (IBA GmbH, Germany), for at least 45 minutes, to allow the production and maturation of enough reporter MS2-GFP proteins. For acclimatization, cells were grown at room temperature for 1 hour. Afterwards, they were transferred to a microscope chamber, for image acquisition.

Cells were kept at 24 °C, 27 °C, 30 °C, 34 °C, or 37 °C during microscopy in a thermal chamber (Bioptechs, FCS2, Pennsylvania, USA). We poured 100 μl of melted agarose-medium with 1% agarose (Sigma life science, USA), LB medium, and aTc (10 ng ml^{−1}), into a microscope slide with a glass coverslip on top. After waiting for the gel-pad to solidify, prior to adding cells, we removed the coverslip and left the gel-pad to dry for 2–5 minutes at room temperature. Finally, we added 5–8 μl of cell suspension into the gel and placed this sandwich in the thermal chamber for image acquisition.

Cells were visualized in a Nikon Eclipse (TE2000-U, Nikon, Japan) inverted microscope with C1 confocal laser-scanning and a 100× Apo TIRF objective. Images were taken every minute for 2 hours. GFP fluorescence was measured using 488 nm argon ion laser (Melles-Griot) and a 515/30 nm emission filter. Images were acquired with Nikon EZ-C1 software and were analysed by custom software written in MATLAB 2011b (MathWorks).

Image analysis

Images of cells with the repressilator and with the switch were analysed differently. To detect cells with the repressilator from images (Fig. 1), we segment them by manually masking the area each occupies in each frame. Next, the total fluorescence intensity in each mask is extracted and the mean pixel intensity of each cell is calculated for each time moment.

For cells containing the switch, thus expressing MS2-GFP and its target RNA, the region occupied by each cell over time was manually masked. In each mask, principal component analysis (PCA) was used to obtain dimensions and orientation of the cell at each moment. By kernel density estimation using a Gaussian kernel²² and Otsu's thresholding,²³ fluorescent spots were automatically segmented. To obtain the intensity of each spot, the cell background was subtracted. Finally, RNA numbers in each cell were obtained from the time series of the corrected total spot intensity by a least squares fit of a monotone piecewise-constant curve (Fig. 2b).²⁴ The number of terms in the curve was selected by an *F*-test with a *p*-value of 0.01. Each jump corresponds to the production of one RNA²⁴ (Fig. 2, for details see ref. 25).

Assessing functionality and apparent period of oscillations

To determine if a repressilator is 'functional' during a time series, we use the criterion used in Elowitz *et al.*⁷ A fast Fourier transform is applied to the temporal fluorescence signal from



Fig. 1 Cell exhibiting oscillatory fluorescence. 5 frames are shown, along with time stamps in minutes. In this case, the images were taken at 30 °C.

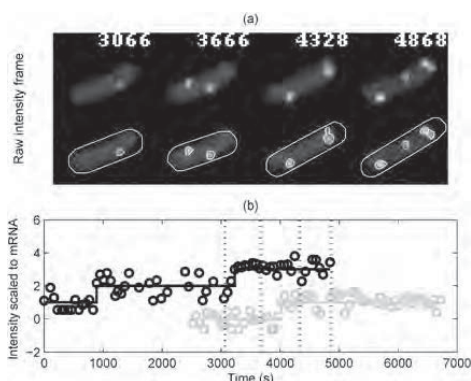


Fig. 2 MS2-GFP tagged RNAs in *E. coli* cells. Unprocessed and segmented cells and RNA spots (a). Moments when images were taken are indicated in each frame. Examples of time series in cells with scaled intensity levels (circles) and estimated RNA numbers (solid lines) (b).

each cell and divided by the transform of a decaying exponential with a time constant of 90 min, the measured lifetime of the fluorescent protein used (GFPaav).⁷ Power spectra with peaks 4.5 times higher than the background at frequencies of 0.15–0.5 per hour were classified as oscillatory. The bandwidth used here is larger than in ref. 7 so as to include failed oscillations that should cause apparent oscillations with close to double period.

For cells considered functional, we estimated the ‘apparent period’ as follows. First, we fit a quadratic curve in the least-squares sense to the intensity time series, to estimate the general trend (Fig. 3, top panel) since the measured intensity is affected by, *e.g.*, photo-bleaching. After subtracting the estimated trend, the residual is scaled to unit power (Fig. 3, middle panel), and then the autocorrelation function is computed (Fig. 3, bottom panel). From this function, we estimate the period by locating the first and the third zeros of the autocorrelation function and computing their distance (Fig. 3, bottom panel, black circles).

Detecting failed oscillations and estimating the true period

The above method of period estimation relies on robust periodic behaviours. If a repressilator halts its activity for a while and then resumes it, the above method cannot detect it. Instead, it assumes an oscillation length that includes the halting and the ‘true’ oscillation. We observed by inspection that, in some cells, the GFP reporter failed to report an oscillation, either because the oscillation itself failed or because the reporters’ expression failed. In general, the reporter signal ‘recovered’ in the next cycle. In these cases, the measured time was double that between other consecutive oscillations.

To extract the ‘true period’, we employed a method that relies on the fact that the distributions of period lengths, when failures occur, resemble bimodal distributions. Namely, we estimate the mean and standard deviation of the true period in the population and the fraction of failed oscillations from

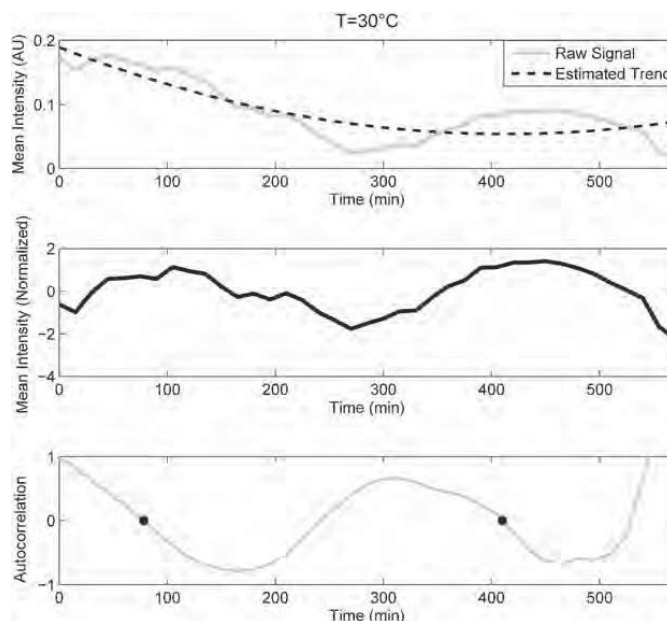


Fig. 3 Period estimation from the fluorescence intensity signal. In the top panel, the raw signal extracted from images is shown along with the estimated trend. In the middle panel, the trend was subtracted from the raw signal and the residual was scaled to unit power. The bottom panel shows the autocorrelation function of the treated signal. The distance between first and third zeros (black circles) corresponds to the period of oscillation in the cell.

the measured periods from each cell. For that, we find the maximum likelihood estimates for a single Gaussian (given by the mean and standard deviation of the measured periods) and for a mixture of two Gaussians, such that the mean and variance of the second are double that of the first (found using an iterative expectation maximization algorithm).²⁶ The appropriate model is selected by a likelihood ratio test with significance level of 0.01 between the two models. That is, we only select the 2-Gaussian model if the p -value of this test is smaller than 0.01. Finally, we performed the fitting with each subset of data lacking one of the measured periods (leave-one-out technique). This procedure results in N estimates each using $N - 1$ measured periods from which the variance of the estimates is estimated.

Results

We measured the behaviour of the repressilator at 28 °C, 30 °C, 33 °C, and 37 °C. We also conducted measurements for lower and higher temperatures than these, but the number of functional repressilators was negligible or non-existing. We limited the measurements' length to 10 h, as cells tend to enter the stationary phase beyond this point, which halts the repressilator.⁷ Cells with a non-functional repressilator, for this or other reasons, were discarded by the method used to determine if the GFP levels oscillate throughout the measurement period (see Methods).

We first tested if the distributions of lengths of the oscillations (Fig. 4) here referred to loosely as 'periods', differ with temperature. For that, we compared all pairs of distributions using the Kolmogorov–Smirnov (KS) test. All, except 33 °C *vs.* 37 °C, differ in a statistical sense (p -values smaller than 0.03), which implies that the circuit is sensitive to temperature.

Under all conditions, as visible from the distributions in Fig. 4, the period lengths vary widely. Given their mean and variability, a number of short-lasting periods (<100 min) are expected (visible in Fig. 4). To verify that these did not occur in a higher than expected frequency, for the condition '30 °C' (the one with most samples), we computed the probability of having such or a more extreme number of periods smaller than 100 min (*i.e.* a p -value) assuming the fitted model (see below and the Methods section). From the model, 2.93 'short periods' are expected while 3 were detected, which results in a p -value of 0.56 *i.e.*, the number of events observed is not unlikely.

The effects of temperature on the distribution of periods' length are visible in Fig. 4. The distribution appears to become bimodal for $T > 30$ °C. This bimodality, not possible if the oscillations in protein numbers were robust, appears to arise from 'failed oscillations' that occur with non-negligible probability. Namely, in some of the cells at $T > 30$ °C, the GFP levels appear to remain low for approximately one cycle and only increase again in the following cycle.

To test for bimodality, for each of the four distributions, we determined the maximum likelihood estimates for a single Gaussian and for a mixture of two Gaussians with the mean and variance of the second Gaussian being double those of the first. The preferred model (see Methods) in each condition is shown in Fig. 4 as well. For 33 °C and 37 °C, the model of two Gaussians was preferred.

Using the fitting, we estimated the number of failed oscillations in each cell, under each condition (see Methods). The fraction of successful oscillations (R) is shown in Table 1, for each condition. Beyond 30 °C, the repressilator loses much of its robustness, as several expected oscillations were not detected. This agrees with the observed decrease in functionality (F) for temperatures above 30 °C (Table 1).

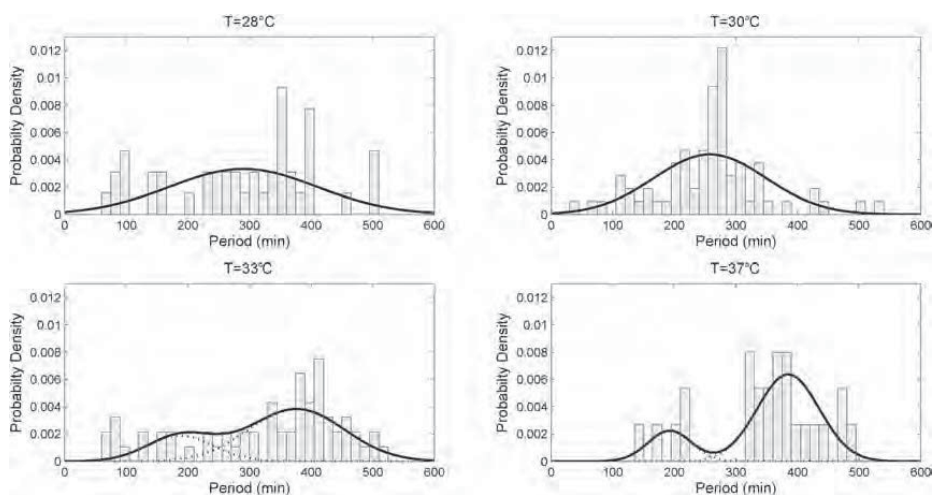


Fig. 4 Distribution of periods (magnitude scaled to represent probability density) for each temperature. Solid lines represent the probability densities of the fitted model with one or two Gaussians. Dashed lines represent the densities of individual components in the case of two Gaussians. For 28 °C and 30 °C, the p -values of the likelihood ratio tests are 0.08 and 1, respectively, indicating a lack of evidence for the two-Gaussian model, whereas for 33 °C and 37 °C, the p -values are 0.0065 and 0.0015, respectively, indicating that the two-Gaussian model should be favored over the one-Gaussian model.

Table 1 Kinetics of the repressilator at different temperatures. Temperature (T), fraction of functional cells (F), total number of cells exhibiting oscillations, fraction of robust oscillations (R), mean (m) and standard deviation (s) of the apparent period, and mean (μ) and standard deviation (σ) of the estimated true period are shown

T (°C)	F (%)	No. of cells		R (%)	m (min)	s (min)	μ (min)	σ (min)
		oscillating						
28	20	43		100	290	120	290	120
30	30	71		100	258	91	258	91
33	15	62		26	328	126	188	59
37	5	25		20	347	92	192	36

Also in Table 1, we show the mean and standard deviation of both the apparent period and the true, estimated period. The mean true period, μ , always decreases with increasing temperature. On the other hand, the mean apparent period, m , is minimal at 30 °C.

Given this, we investigated whether the distributions of durations of true oscillations alone also differ with temperature, as the distributions of apparent oscillations do. Namely, we estimated the mean true period (Fig. 5) and then the one standard deviation of this estimate (error bars in Fig. 5). From Fig. 5, this mean always decreases significantly as temperature increases, except beyond 33 °C.

Next, we investigated the causes for the decrease in robustness with temperature. In particular, we investigated how temperature affects the functionality of the three component proteins of the repressilator, namely, CI, LacI, and TetR. First, studies suggest that as temperature increases from 30 °C to 42 °C, CI loses approximately half of its DNA-binding stability.¹⁷ On the other hand, the DNA-binding affinity of LacI does not vary significantly between 28 °C and 37 °C.²⁷ Similarly, TetR's functionality is unaltered from 20 °C to 40 °C.²⁸ We thus hypothesized that a possible cause for the loss of robustness of the repressilator with increasing T was the weakening effectiveness of CI as a repressor.

There is another circuit, the CI-Cro genetic switch, of which CI is a component. If CI loses functionality with increasing temperature (partially or completely) the behaviour of this switch should change with temperature. To determine whether this is

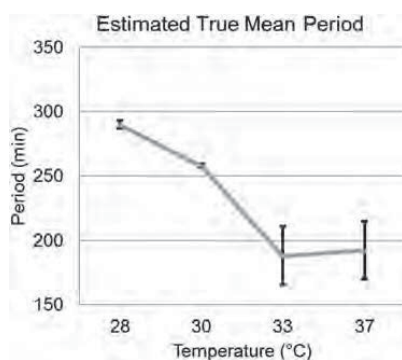


Fig. 5 Estimated mean values of the true period. Error bars indicate one standard deviations of the mean period estimated by the leave-one-out technique.

Table 2 Intervals between the appearances of novel, consecutive RNA molecules in individual cells. For condition, the table shows the number of intervals, mean (μ), standard deviation (σ) and, square of the coefficient of variation (CV^2) of the interval duration

T (°C)	No. intervals	μ (s)	σ (s)	CV^2
24	157	1242	1166	0.88
27	229	1452	1191	0.67
30	88	1130	1040	0.85
33	539	788	807	1.05
37	324	714	785	1.21

Table 3 P -values of the Kolmogorov–Smirnov test between distributions of intervals between consecutive RNA production events, under the control of P_{RM} , obtained at different temperatures. For p -values < 0.01 , the hypothesis that the two distributions are the same is rejected

T (°C)	24	27	30	33
27	0.149			
30		0.006		
33			0.009	
37				0.478

the case, we conducted *in vivo* measurements of RNA production, one event at a time, by one of the two genes of this switch. This particular gene is controlled by the promoter P_{RM} , and codes for CI as well as for a 48 MS2d binding array. The second gene of the switch, whose activity is not followed, is controlled by the promoter P_R and codes for Cro. Relevantly, Cro-DNA interactions do not vary significantly from 24–37 °C,²⁹ thus, behaviour changes in this switch with increasing temperature should mostly arise from the changes in CI-DNA interactions.

We measured intervals between consecutive productions of the RNA target for MS2-GFP in individual cells, from *in vivo* measurements 2 h long, with images taken every minute, at 24 °C, 27 °C, 30 °C, 33 °C and 37 °C. In Table 2, we show for each condition the number of samples (*i.e.* intervals) and the mean and standard deviation of the intervals' duration. As temperature increases, the kinetics of production of the target RNA changes. Specifically, aside from a decrease in the mean interval between consecutive transcription events, one observes that the production kinetics changes from sub-Poissonian ($CV^2 < 1$) for $T \leq 30$ °C, to super-Poissonian ($CV^2 > 1$) for $T > 30$ °C.

To verify if the change is significant, we compared the distributions of intervals in consecutive temperatures with the K–S test. The results in Table 3 indicate that the distributions at 24 °C and 27 °C cannot be statistically distinguished from one another. Similarly, the distributions at 33 °C and 37 °C cannot be distinguished. Meanwhile, the distributions from 27 °C and 30 °C, as well from 30 °C and 33 °C, differ from one another. Thus, there is a change in the dynamics of transcript production, and it occurs around 30 °C, which is similar to the point where changes in behaviour of the repressilator are observed.

Conclusions and discussion

We studied the behaviour of the repressilator at different temperatures. We observed that the fraction of functional cells

(i.e. exhibiting oscillations), the robustness of the oscillations in functional cells, and also the apparent and the real period all differ with temperature.

Because the robustness decreases at higher-than-optimal temperatures, the extraction of the period in this regime requires the identification of failed oscillations. Otherwise, the period will likely be overestimated. The extraction method here proposed should be applicable to other genetic clocks as well.

The apparent period was minimized at 30 °C. However, the results of employing the novel method of period extraction suggest that the increase in apparent period when increasing temperature beyond 30 °C is due to an increasing rate of failed oscillations. Meanwhile, the true period decreased significantly with increasing temperature (until 33 °C), in accordance with the response of other synthetic genetic clocks to increasing temperature.³⁰ This decrease is likely caused by the increased rate of the underlying thermodynamic processes (see ref. 30). In particular, we expect the decay rates of the proteins to increase, which decreases the period length.³⁵ The increased protein decay rates are expected from both increased rates of degradation and increased doubling rate of the cells. This allows the repressilator to be sensitive to temperature changes in the range tested.

We hypothesize that the design of genetic clocks that are insensitive to temperature will have to be able to compensate for increased speed of processes such as cell division, open complex formation,¹³ among others.

Subsequently, based on previous studies on the functionality of the proteins of the repressilator,^{17,27,28} we hypothesized that the loss of robustness with increasing temperature was associated with the temperature-dependent functionality of one component protein, CI. We tested this indirectly, by studying how temperature affects the CI-Cro switch. In particular, we conducted *in vivo* measurements, one event at a time, of the kinetics of production of an MS2-GFP tagged RNA that probed the transcription kinetics of the RNA coding for CI. From these, we observed that, when increasing temperature beyond 30 °C, the dynamics of production of the tagged RNA changed from sub-Poissonian to super-Poissonian, which suggests that the production of the tagged RNA became subject to repression.

Recent studies in *E. coli* suggest that, provided no repression, RNA production is a sub-Poissonian process, within the range of temperature tested here.^{13,24,25,31} To be super-Poissonian, the promoter ought to have intervals of inactivity^{21,34} (e.g. due to repressors) or due to another, similar mechanism. In the case of the CI-Cro switch, the occurrence of periods of inactivity of P_{RM} is expected if CI loses functionality, allowing Cro to be expressed.^{32,33} Thus, these results suggest that CI loses functionality with increasing temperature.

The repressilator and the CI-Cro switch only share CI in common, while the other component proteins differ. Relevantly, the interactions between all these other proteins and their respective DNA binding sites are not temperature-dependent in the range studied.^{27–29} Given this and all of the above, it is therefore reasonable to conclude that, in both circuits, the

behavioural changes with temperature observed are primarily due to the temperature-dependence of CI's activity.^{18,19}

Further, we hypothesize that it is possible to modify the repressilator so as to make it more robust to a wider range of temperatures. For that, the CI-DNA interaction should be replaced by a less temperature-dependent repression mechanism. This modification is not expected to compromise the sensitivity (which likely depends more heavily on the temperature-dependent cell division rate, among others).

It is worthwhile discussing the different effects of temperature on robustness and functionality. Namely, while functionality is maximized at 30 °C, robustness was only compromised at higher-than-optimal temperatures. In the latter regime, the two decreases are likely related. As robustness decreases, we expect a higher chance for repressilators to not function during the measurements. However, at lower-than-optimal temperatures, the loss in functionality is likely caused by other reasons, as the robustness was not compromised. Future research is needed to identify such causes.

Finally, the results presented here demonstrate that the behavioural changes in genetic circuits upon changing conditions depend not only on the topology of the circuit, but also on how each of its components responds to the environmental changes.

Acknowledgements

This work was supported by Academy of Finland (ASR), Finnish Funding Agency for Technology and Innovation (ASR), and Tampere City Science Foundation (AH). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank M. Elowitz and I. Golding for generously providing the genetic circuits.

Notes and references

- 1 C. H. Ko and J. S. Takahashi, *Hum. Mol. Genet.*, 2006, **15**, R271–R277.
- 2 Z. Neubauer and E. Calef, *J. Mol. Biol.*, 1970, **51**, 1–13.
- 3 D. L. Gally, J. A. Bogan, B. I. Eisenstein and I. C. Blomfield, *J. Mol. Biol.*, 1970, **51**, 1–13.
- 4 A. Becskei and L. Serrano, *Nature*, 2000, **405**, 590–593.
- 5 N. Nandagopal and M. B. Elowitz, *Science*, 2011, **333**, 1244–1248.
- 6 D. M. Wolf and A. P. Arkin, *Curr. Opin. Microbiol.*, 2003, **6**, 125–134.
- 7 M. B. Elowitz and S. Leibler, *Nature*, 2000, **403**, 335–338.
- 8 T. S. Gardner, C. Cantor and J. J. Collins, *Nature*, 2000, **403**, 339–342.
- 9 A. S. Khalil, T. K. Lu, C. J. Bashor, C. L. Ramirez, N. C. Pyenson, J. K. Joung and J. J. Collins, *Cell*, 2012, **150**, 647–658.
- 10 J. M. Callura, C. R. Cantor and J. J. Collins, *Proc. Natl. Acad. Sci. U. S. A.*, 2012, **109**, 5850–5855.
- 11 M. B. Elowitz and W. A. Lim, *Nature*, 2010, **468**, 889–890.
- 12 K. D. Litcofsky, R. B. Afeyan, R. J. Krom, A. S. Khalil and J. J. Collins, *Nat. Methods*, 2012, **9**, 1077–1080.

- 13 A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2012, **40**, 8472–8483.
- 14 D. M. Virshup and D. B. Forger, *Cell*, 2009, **137**, 602–604.
- 15 I. Mihalcescu, W. Hsing and S. Leibler, *Nature*, 2004, **430**, 81–85.
- 16 O. Oleksiuk, V. Jakovljevic, N. Vladimirov, R. Carvalho, E. Paster, W. S. Ryu, Y. Meir, N. S. Wingreen, M. Kollmann and V. Sourjik, *Cell*, 2011, **145**(2), 312–321.
- 17 N. Jana, S. Roy, B. Bhattacharyya and N. C. Mandal, *Protein Eng.*, 1999, **12**(3), 225–233.
- 18 K. Koblan and G. Ackers, *Biochemistry*, 1991, **30**, 7822–7827.
- 19 K. Koblan and G. Ackers, *Biochemistry*, 1991, **30**, 7817–7821.
- 20 *Lambda II*, ed. R. W. Hendrix, J. W. Roberts, F. W. Stahl and R. A. Weisberg, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1983.
- 21 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–1036.
- 22 T. B. Chen, H. H. Lu, Y. S. Lee and H. J. Lan, *J. Biomed. Inf.*, 2008, **41**, 1021–1027.
- 23 N. Otsu, *IEEE Trans. Syst. Man Cybern.*, 1979, **9**, 62–66.
- 24 M. Kandhavelu, H. Mannerstrom, A. Gupta, A. Hakkinen, J. Lloyd-Price, O. Yli-Harja and A. S. Ribeiro, *BMC Syst. Biol.*, 2011, **5**, 149.
- 25 M. Kandhavelu, J. Lloyd-Price, A. Gupta, A.-B. Muthukrishnan, O. Yli-Harja and A. S. Ribeiro, *FEBS Lett.*, 2012, **586**, 3870–3875.
- 26 A. P. Dempster, N. M. Laird and D. B. Rubin, *J. R. Stat. Soc. Ser. B (Methodological)*, 1977, **39**(1), 1–38.
- 27 D. E. Frank, R. M. Saecker, J. P. Bond, M. W. Capp, O. V. Tsodikov, S. E. Melcher, M. M. Levandoski and M. T. Record Jr, *J. Mol. Biol.*, 1997, **267**, 1186–1206.
- 28 W. Hillen, G. Klock and I. Kaffenberger, *J. Biol. Chem.*, 1982, **257**, 6605–6613.
- 29 Y. Takeda, P. Ross and C. Mudd, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 8180–8184.
- 30 J. Stricker, S. Cookson, M. R. Bennet, W. H. Mather, L. S. Tsimring and J. Hasty, *Nature*, 2008, **456**, 516–520.
- 31 J. Mäkelä, M. Kandhavelu, S. Oliveira, J. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja and A. Ribeiro, *Nucleic Acids Res.*, 2013, **41**, 6544–6552.
- 32 L. Anderson and H. Yang, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 5827–5832.
- 33 S. Svenningsen, N. Costantino, D. Court and S. Adhya, *Proc. Natl. Acad. Sci. U. S. A.*, 2005, **102**, 4465–4469.
- 34 C. W. Gardiner, *Handbook of Stochastic Methods*, Springer, NY, 3rd edn, 2004.
- 35 A. Loinger and O. Biham, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **76**, 051917.

PUBLICATION II

Single-cell kinetics of the Repressilator when inserted into a single-copy plasmid

S.M.D. Oliveira*, J.G. Chandraseelan*, A. Häkkinen, N.S.M. Goncalves, O. Yli-Harja, S. Startceva and
A.S. Ribeiro

Mol. Biosyst., 11:1939-1945, 2015. *Equal contributions.
doi: 10.1039/c5mb00012b

Publication reprinted with the permission of the copyright holders.

PAPER



Cite this: *Mol. BioSyst.*, 2015,
11, 1939

Single-cell kinetics of a repressilator when
implemented in a single-copy plasmid†

Samuel M. D. Oliveira,^{‡a} Jerome G. Chandraseelan,^{‡a} Antti Häkkinen,^a
Nadia S. M. Goncalves,^a Olli Yli-Harja,^b Sofia Startceva^a and Andre S. Ribeiro*^a

Synthetic genetic clocks, such as the Elowitz–Leibler repressilator, will be key regulatory components of future synthetic circuits. We constructed a single-copy repressilator (SCR) by implementing the original repressilator circuit on a single-copy F-plasmid. After verifying its functionality, we studied its behaviour as a function of temperature and compared it with that of the original low-copy-number repressilator (LCR). Namely, we compared the period of oscillations, functionality (the fraction of cells exhibiting oscillations) and robustness to internal fluctuations (the fraction of expected oscillations that would occur). We found that, under optimal temperature conditions, the dynamics of the two systems differs significantly, although qualitatively they respond similarly to temperature changes. Exception to this is in the functionality, in which the SCR is higher at lower temperatures but lower at higher temperatures. Next, by adding IPTG to the medium at low and high concentrations during microscopy sessions, we showed that the functionality of the SCR is more robust to external perturbations, which indicates that the oscillatory behaviour of the LCR can be disrupted by affecting only a few of the copies in a cell. We conclude that the SCR, the first functional, synthetic, single-copy, ring-type genetic clock, is more robust to lower temperatures and to external perturbations than the original LCR. The SCR will be of use in future synthetic circuits, since it complements the array of tasks that the LCR can perform.

Received 8th January 2015,
Accepted 17th April 2015

DOI: 10.1039/c5mb00012b

www.rsc.org/molecularbiosystems

Introduction

Efforts in synthetic biology dedicated to the engineering of artificial genetic circuits have focused on constructing functional switches (for decision making), clocks (for time keeping), and noise and frequency filters,¹ as these modules are likely candidate regulatory components of the activity of future, more complex synthetic circuits.

One of the first reported functional synthetic circuits is the ‘repressilator’,² a ring-oscillator with three genes, each expressing a protein that represses the next gene in the loop. From the study of the signal from GFP reporters in cells at 30 °C, it was shown that it oscillates (stochastically) at a slower rhythm than the cell cycle. Interestingly, for unknown reasons, only approximately 40% of the cells exhibit oscillations, *i.e.* are ‘functional’. Further, even though the circuits’ behaviour is uncoupled from the cell cycle in the previous phase,² these functional cells

become non-functional in the stationary phase, suggesting that this synthetic network is not fully uncoupled from the regulatory mechanisms of cell growth. Finally, the oscillatory behaviour can be halted by external signals. *E.g.* most cells lose the functionality following the addition of 50 μM of isopropyl β-D-1-thiogalactopyranoside (IPTG) to the medium.²

A subsequent study³ analysed the behaviour of the repressilator at temperatures below and above the optimal (from 28 °C to 37 °C, with 30 °C being considered as optimal), focusing on the period of oscillations, the functionality (the fraction of cells exhibiting oscillations) and the robustness (the fraction of expected oscillations that would occur) of the signal from the cells. Both the functionality (maximum at 30 °C) and the period length were found to be temperature-dependent. The minimum period length was observed at 30 °C. While the reason for longer periods at lower-than-optimal temperatures is likely the slower rate of most chemical processes, at temperatures beyond optimal, longer periods emerge due to the loss of functionality of one of the component proteins of the repressilator, CI.³

Originally, the repressilator was implemented on a low-copy plasmid (pZS1-ITrLLtCL). Because of this (as the origin of replication is pSC101), each cell has, on average, 3–4 copies of the repressilator,⁴ which are functionally coupled in that the proteins coded by a gene in one of the copies can act as repressors of the next gene in the loop in all other copies of

^a Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland.
E-mail: andre.ribeiro@tut.fi

^b Computational Systems Biology Research Group, Department of Signal Processing, Tampere University of Technology, P.O. Box 553, 33101 Tampere, Finland

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5mb00012b

‡ Equal contribution.

the plasmid in the cell. This coupling, according to simulations of stochastic models, is expected to reduce the fluctuations in period lengths that arise from the stochasticity in gene expression and in RNA and protein degradation.^{5,6} So far, it is unknown whether the repressilator would function if implemented on a single-copy plasmid.

If functional, a single-copy repressilator (SCR) ought to be of use to ongoing efforts in synthetic biology. For example, by comparing its behaviour with that of the original LCR, we may obtain a better understanding of how the copy number variation in bacteria can lead to changes in bacterial growth rates^{7,8} and phenotypic innovation,⁹ among others. In the case of the repressilator, it is expected that the copy-number will affect the dynamics severely enough to allow the system to change from a single steady state to sustained oscillations.¹⁰

As these and other expectations are, so far, solely based on theoretical models,^{7,10–13} we have implemented the Elowitz–Leibler repressilator² on a single-copy F-plasmid (pBAC2) in order to conduct an empirical analysis of the behavioural changes due to copy-number differences. This plasmid is well-known for its high hereditary, *i.e.* copy-number and stability.^{14,15} For most of the cell cycle there is only one copy of the plasmid in the cell, which is replicated once, prior to cell division.

After verifying the functionality of our SCR, we compared its dynamics with the original LCR at optimal temperatures. Next, we compared their responses to changing temperatures. Finally, we studied their robustness to external perturbations.

The results provide new insights into the effects of coupling on genetic circuits in general, and clocks in particular. Understanding the functioning of natural, as well as synthetic clocks, such as the repressilator, will assist in the understanding of how cells regulate the timing of several processes¹⁶ and contribute to ongoing efforts in synthetic biology to produce circuits useful in assisting medicine and biotechnology, particularly given the important role that synthetic clocks are expected to play as sensors and regulators in future synthetic circuits.

Methods

Design, validation, and functioning of single-copy repressilators

The repressilator consists of a three-gene network in a loop formation, with each gene repressing the next gene in the loop² (Fig. 1, top). Such a network is expected to exhibit periodic oscillations in the protein levels of the component genes (Fig. 1, bottom). To observe them, a GFP reporter is used, which is regulated by one of the proteins of the circuit.²

To build the SCR, we transferred the sequence coding for the repressilator from the original low-copy plasmid into a single-copy F-plasmid (pBAC2-ITrLLtCL). The original GFP reporter system² was left unchanged. The construction history of the SCR is shown in Fig. S1 (ESI†). Images of the gels of the SCR plasmid and PCR are shown in Fig. S4 and S5 (ESI†), respectively. To further confirm the proper construction of the SCR

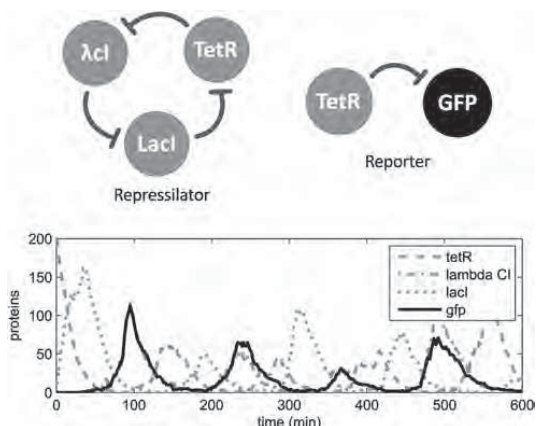


Fig. 1 Top: graphical representation of the 3-gene network (repressilator) along with the GFP reporter. Bottom: an example output from a stochastic model of the repressilator.¹¹ The black curve represents the output of the reporter while the grey curves represent the actual protein numbers of the three genes of the repressilator. The signals of the GFP reporter and λ CI are nearly superimposed, as expected.

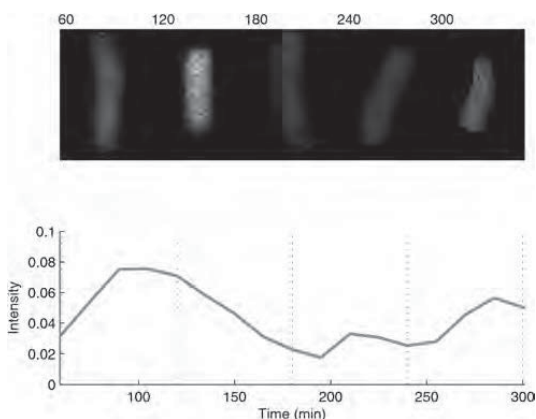


Fig. 2 Top: example images of a cell exhibiting oscillatory fluorescence levels. 5 frames are shown along with the time stamps in minutes. Images taken at 37 °C. Bottom: the mean fluorescence intensity level (in arbitrary units) over time of the cell shown above. The dashed lines indicate points at which the above frames were captured.

plasmid we performed sequencing and qPCR (Fig. S6 and ESI†, respectively).

Finally, we conducted live cell microscopy to determine whether cells with the SCR exhibited a fluorescent signal whose intensity oscillates (for example Fig. 2 and Fig. S7, ESI†), similar to the original LCR. The observations confirmed the existence of oscillations.

Strains, genetic circuit assembly, and growth conditions

Cells of *E. coli* host strain *lac*[−] MC 4100, containing the low-copy repressilator (pZS1-ITrLLtCL) and the reporter (pZE21-GFPaav)

plasmids, were generously provided by M. Elowitz (Princeton University, NJ, USA).² In cells of the same strain, MC4100, containing only the reporter system (also generously provided by M. Elowitz), controlled by the tetracycline repressor (*TetR*) and the promoter P_L *tetO1*,² we inserted the engineered single-copy F-plasmid containing the repressilator system (pBAC2-ITrLLtCL) from pZS1-ITrLLtCL (Fig. S1, ESI†).

The low-copy (LCR) and single-copy repressilator (SCR) strains were grown in agar lysogeny broth (LB) medium from glycerol stock (kept at $-80\text{ }^{\circ}\text{C}$) for 12 hours until single colonies could be detected. Single colonies selected from the LB plates were transferred to LB-agar plates for 8 hours of fast growth. A single colony was then inoculated into a minimal medium for 10 hours at $28\text{ }^{\circ}\text{C}$, $30\text{ }^{\circ}\text{C}$, $33\text{ }^{\circ}\text{C}$ or $37\text{ }^{\circ}\text{C}$ with shaking at 250 rpm (6 rcf), to an optical density (OD) ≈ 0.1 at 600 nm. Next, cells were centrifuged at 8000 rpm (6093 rcf) for 1 minute and diluted into fresh minimal medium. In all steps, besides image acquisition, LCR cells' preparation contained $35\text{ }\mu\text{g ml}^{-1}$ of kanamycin and $20\text{ }\mu\text{g ml}^{-1}$ of ampicillin, while SCR cells' preparation contained $35\text{ }\mu\text{g ml}^{-1}$ of kanamycin and $35\text{ }\mu\text{g ml}^{-1}$ of chloramphenicol (all antibiotics from Sigma Aldrich, USA). For imaging, a few μl of the culture were placed between a cover-slip and a 2.5% low melting agarose gel pad of minimal medium with 1 mm thickness.

As mentioned above, the LCR and SCR differ in their antibiotic markers, as the SCR uses chloramphenicol instead of ampicillin. However, as we do not use either of antibiotics during the microscopy measurements and given the identical growth rates of the two strains during those measurements (see below), this difference is not expected to affect their dynamics.

Cell culturing optimization

To avoid plasmid instability^{17–19} and to optimize culturing protocols^{2,3} we proceeded as follows: (i) as in the original protocol,^{2,3} cells were taken from a stock (at $-80\text{ }^{\circ}\text{C}$) and streaked onto an LB agar medium with appropriate antibiotics for 14–16 hours; (ii) at this stage, we added to the original protocol an extra step of 8 hours of cell growth from single colonies in LB agar medium;^{2,3} (iii) next, as in the original protocol, a few colonies from the second plate were inoculated in M63 liquid minimal medium with antibiotics for 10 hours. Finally, we placed cells at $28\text{ }^{\circ}\text{C}$, $30\text{ }^{\circ}\text{C}$, $33\text{ }^{\circ}\text{C}$, or $37\text{ }^{\circ}\text{C}$ for 8 hours, after which we measured the optical density (OD). After these 8 hours of culturing $\text{OD}_{600} \approx 0.1\text{--}0.2$ was reached under all conditions, as reported in ref. 3 No differences in behaviour were found between cells with the LCR and with the SCR during this procedure.

Microscopy

During time lapse microscopy, cells were kept at a stable temperature in a thermal chamber (Biopetechs, FCS2, PA, USA). Images of both LCR and SCR cells were obtained every 15 minutes for 10 hours using a Nikon Eclipse (Ti-E, Nikon, Japan) inverted microscope equipped with a C2+ confocal laser-scanning system and a $100\times$ Apo TIRF (1.49 NA, oil) objective. Images were taken from multiple locations at each moment. GFP fluorescence was

excited using a 488 nm argon ion laser (Melles-Griot) and measured using a 515/30 nm emission filter. The pixel dwell time was set to $2.4\text{ }\mu\text{s}$, so that the total image acquisition time per location was $\approx 2.5\text{ s}$. The laser shutter was open only during exposure to minimize photo-bleaching. We used NIS-Elements software (Nikon) for image acquisition.

Data and image analysis

For image and data analysis, we used custom software written in MATLAB 2011b (MathWorks). Cells with either the LCR or the SCR were manually segmented in the images.³ Next, the segments were automatically tracked based on the overlapping areas of the segments in consecutive frames, and the total fluorescence intensity was extracted and used to calculate the mean pixel intensity of the cell at each moment.³

We used the following criterion to determine the functionality of the repressilator:^{2,3} for cells presenting the fluorescence signal from start to end, a discrete Fourier transform was applied and divided by the transform of a decaying exponential of the measured lifetime of the fluorescent protein used (*GFPaav*), with a time constant of 90 min.² From these, cells with power spectra with peaks 3 times higher than the background, at frequencies of 0.2–0.5 per hour, were classified as oscillatory. As discussed in ref. 3, the bandwidth was larger than that reported in ref. 2 to detect failed oscillations, which create apparent periods close to the double mean and standard deviation.³

The same method as reported in ref. 3 was used to estimate the period of oscillations for each cell. It consists of subtracting the detected trend from a raw signal, followed by scaling the residual to unit power and computing an autocorrelation function. Afterwards, periods were estimated by locating the first and third zeros of the autocorrelation function, and computing their distance.

Robustness of the oscillations to internal fluctuations

It is known that some cells fail to report an oscillation at times, particularly at temperatures higher than $30\text{ }^{\circ}\text{C}$.³ This occurs either because no oscillation occurred or because the reporter failed to report it. Typically, such 'halted' signals resume in the next cycle. Regardless of the cause, these events are evidence for the lack of robustness to internal fluctuations of the repressilator-reporter system. To quantify this phenomenon and assess its temperature dependence in each system, we defined robustness to internal fluctuations as the fraction of expected oscillations that do occur according to the reporter (*i.e.* the ratio between true and apparent cycles). To find the fraction of 'apparent' and 'true' cycles in functional cells under each condition, distributions of period lengths were analysed (Fig. S2 and S3, ESI†).

As reported in ref. 3, these distributions were fitted to either a single Gaussian or to a mixture of two Gaussians. In the distributions where bimodality was observed, the 'apparent' and 'true' periods were extracted from the fitting of a 2-Gaussian model such that the mean and the variance of the second period were twice that of the first (found using an iterative expectation

maximization algorithm).²⁰ The appropriate model was selected by a likelihood ratio test with a significance level of 0.01 between the two models. That is, we only select the 2-Gaussian model if the *p*-value of this test is smaller than 0.01. This methodology was used here solely to quantify the robustness of the repressor–reporter system to internal fluctuations.

External perturbation of the activity of the repressor

In one experiment, we assessed the robustness of the SCR and LCR to external perturbations by introducing a certain concentration of IPTG into the medium, 180 minutes after starting the microscopy measurements (to allow at least one cycle of oscillation). IPTG induces the P_L lacO1 promoter and, as such, it should disrupt the functioning of the repressor. We performed three such experiments. First, we perturbed cells with the SCR and cells with the LCR by adding 50 μ M of IPTG to the medium at 30 °C, to compare the robustness of these two circuits to external perturbations. Next, we perturbed cells with the SCR by adding 1 mM of IPTG to the medium, so as to compare the effects of ‘weak’ (50 μ M) and ‘strong’ (1 mM) perturbations on the SCR dynamics. In all tests, pre-warmed fresh M63 medium containing IPTG at the desired concentration was added to the medium using a peristaltic pump at the rate of 0.3 ml min^{−1}. Images were taken every 15 minutes for 6 hours (3 hours prior to perturbation and 3 hours after the perturbation).

Results

Effects of the plasmid copy number under optimal conditions

We imaged cells with the SCR and LCR at 30 °C, the temperature at which the LCR exhibits shorter period and higher functionality and robustness,³ due to, among others, the temperature dependence of $CI^{3,21-23}$ (see results in Table 1). It is noted that in both cases, cells grow at a relatively slow rate under the microscope (division time of ~ 60 min). Thus, it is reasonable to assume that, in the case of SCR cells, most of the time only one copy of the repressor is present in the cells. Nevertheless, in all cases, the SCR cells contain significantly fewer copies of repressors than the LCR cells (see below).

From these data, we assessed if the dynamics of the LCR and SCR differed significantly by performing a Kolmogorov–Smirnov (K–S) test of statistical significance to determine whether the two sets of time lengths of oscillations could be obtained from equal distributions. We obtained a *p*-value of 0.006, from

Table 1 Kinetics of the LCR and the SCR at 30 °C. The table shows the fraction of functional cells (*F*), the number of cells exhibiting oscillations, the fraction of robust oscillations in functional cells (*R*), the mean (*m*) and standard deviation (*s*) of the period, and the squared coefficient of variation (CV^2) of the period in functional cells

Copy no.	<i>F</i> (%)	No. of oscillating cells	<i>R</i> (%)	<i>m</i> (min)	<i>s</i> (min)	CV^2
LCR	42	37	100	251	89	0.126
SCR	48	59	100	313	122	0.152

Table 2 Kinetics of the LCR and the SCR at various temperatures. The temperature (*T*), the fraction of functional cells (*F*), the number of cells exhibiting oscillations, the fraction of robust oscillations (*R*), the mean (*m*) and standard deviation (*s*) of the period are shown

Copy no.	<i>T</i> (°C)	<i>F</i> (%)	No. of oscillating cells	<i>R</i> (%)	<i>m</i> (min)	<i>s</i> (min)
LCR	28	30	41	100	393	40
LCR	30	42	37	100	251	89
LCR	33	35	38	43	275	100
LCR	37	30	38	34	291	111
SCR	28	32	46	100	342	124
SCR	30	48	59	100	313	122
SCR	33	24	84	39	364	161
SCR	37	21	49	31	404	145

which we concluded that the dynamics of the circuits differs statistically (typically, the null hypothesis is rejected at a significance level of 0.01). From this difference, and since the repressor circuits in the SCR and LCR implementation do not differ, it is possible to conclude that more than one copy of the 3–4 copies⁴ of the LCR present in each cell is active.

Interestingly, this difference in the period distributions (particularly the higher variance) is not reflected in the robustness of the oscillations of the SCR, which does not differ from the robustness of the LCR. Also, the SCR exhibits higher functionality at 30 °C than the LCR (similar values to those previously reported in ref. 2). Finally, the period of the SCR is longer and noisier (higher CV^2).

Dynamics at different temperatures

Next, we measured the behaviour of the LCR and of the SCR at 28 °C, 30 °C, 33 °C, and 37 °C (Table 2). We also conducted measurements at lower and higher temperatures than these, but the number of functional repressors was negligible. We limited the measurement period to 10 h, as cells tend to enter the stationary phase at this stage, halting the repressor.^{2,3}

From the images, for each condition, we extracted the fraction of functional cells (*F*), the number of cells exhibiting the oscillatory fluorescent signal, the robustness (*R*) of the oscillations in ‘functional’ cells, and the mean and standard deviation of the period (*m* and *s*). The results are shown in Table 2.

In Table 3, we show the results from K–S tests of statistical significance to determine whether the distributions of periods from the LCR and the SCR could be obtained from equal distributions, at each temperature. This table indicates that the two circuits exhibit different dynamics at all temperatures. Nevertheless, Table 2 indicates that both circuits respond similarly (but not identically) to temperature changes, in the range tested. Specifically, one similarity is that in both circuits

Table 3 *p*-values of the K–S test between distributions of periods from different copy number plasmids. For *p*-values < 0.01, the null hypothesis that the two distributions are equal is rejected

<i>T</i> (°C)	28	30	33	37
<i>p</i> -value	5.02×10^{-4}	0.006	0.36×10^{-4}	7.04×10^{-7}

Table 4 p -values of the Kolmogorov–Smirnov test between distributions of the LCR and SCR periods from different temperatures. For p -values < 0.01, the null hypothesis that the two distributions are equal is rejected

Copy no.	28 °C vs. 30 °C	30 °C vs. 33 °C	33 °C vs. 37 °C
LCR	3.09×10^{-14}	0.13	0.33
SCR	0.44	6.70×10^{-5}	0.30

the mean period is minimized at 30 °C. Another is that the robustness of both circuits is hampered at temperatures beyond 30 °C, due to the loss of effectiveness of CI as a repressor.^{21–23} Finally, in both systems, the functionality is maximized at 30 °C.

On the other hand, as indicated in Table 2, one main difference in how the two circuits respond to temperature changes is that the functionality of the SCR has a steeper decrease with increasing temperature. This causes the SCR's functionality to be lowered at higher temperatures (33 °C and 37 °C). The other significant difference is in how the oscillations change with temperature. While in the LCR the steepest change in the mean period length occurs when raising the temperature from 28 °C to 30 °C (decrease of 142 min), in the SCR it occurs when raising the temperature from 30 °C to 33 °C (increase by 51 min). We verified this by K–S tests of statistical significance to determine whether the sets of oscillation lengths at different temperatures could be obtained from equal distributions, for both the LCR and the SCR (Table 4). From these, one observes a p -value smaller than 0.01 in the LCR only when comparing data from 28 °C and 30 °C, while in the SCR such observation only occurs when comparing data from 30 °C and 33 °C, in agreement with the observed changes in the oscillations' mean time length with temperature.

Perturbing the functioning by IPTG induction

One important property of genetic clocks is their robustness and/or responsiveness to external perturbations. In natural systems, depending on the tasks that they are involved in, it is expected that the genetic clocks have evolved specific robustness and/or responsiveness to perturbations. *E.g.*, some clocks likely evolved robustness to weak, spurious perturbations but responsiveness to strong perturbations (such as due to an environmental shift). For similar reasons, these properties are also important in the case of synthetic circuits, as they will define their applicability.

We compared the robustness of the LCR and the SCR to a 'weak' perturbation, by addition of a small amount of IPTG to the medium. Also, we compared the robustness of the SCR to a 'weak' and to a 'strong' perturbation. For this, after starting measurements as before, we introduced IPTG into the medium at the end of the third hour of the measurements, as this is approximately the time length of one oscillation (see the Methods section). The expected effect of this permanent perturbation is the continuous induction of the P_L lacO1 promoter (*i.e.* up-regulation of *TetR*) in the repressilator. Consequently, P_L tetO1 ought to become permanently repressed. Since this promoter also drives the reporter, the reporter signal should

become negligible, following the perturbation, if the perturbation succeeds in disrupting the oscillations.

First, we compared the effects of perturbation (50 μ M of IPTG) on the dynamics of the LCR with of the SCR. For that, we assessed the functionality (see the Methods section) in the first 3 hours prior to perturbation and in the subsequent 3 hours after the perturbation. We found that the functionality of cells with the LCR equalled 93.3% in the first 3 hours, and 1.64% in the subsequent 3 hours (61 cells imaged) (*i.e.* 98% of the cells were perturbed). Meanwhile, the functionality of cells containing the SCR equalled 100.0% in the first 3 hours, and 8.96% in the subsequent 3 hours (145 cells imaged) (*i.e.* 91% of the cells were perturbed). Thus, surprisingly, we conclude that the LCR is less robust to this perturbation than the SCR. As a side note, the reason why the functionality values are much higher than those shown in Table 2 is the shorter duration of the present measurements and the criteria of the functionality (see the Methods section).

Next, we compared the effects of a 'strong' *versus* a 'weak' perturbation in cells containing the SCR (by adding 1 mM or 50 μ M of IPTG to the medium, respectively), at the end of the third hour of the measurements (112 cells imaged). We measured a functionality of 100.0% in the first 3 hours as before, but only 0.89% in the later 3 hours (99.1% of the cells were perturbed). We conclude that, as predicted in simulations of models of this and similar circuits,^{6,11} the robustness of the SCR's functionality to external perturbations decreases with the strength of the perturbation, in this case defined by the concentration of IPTG in the medium.

Assessing the robustness to perturbations

Given that the LCR and the SCR exhibit different dynamics at any of the temperatures tested, we concluded that more than one repressilator circuit is active in cells with the LCR. Also, it is reasonable to assume that, following the introduction of IPTG into themedia with LCR cells, in each cell, it is always equally or more likely that at least one circuit is affected by IPTG than all of its circuits. Finally, it is reasonable to assume that for the same perturbation, it is more likely that at least one circuit is perturbed in cells with the LCR than in cells with the SCR, due to the larger number of circuits.

Given the above and the observation that cells with the LCR exhibit weaker robustness to the external perturbations than cells with the SCR, it is possible to conclude that, in cells with the LCR, not all copies of the repressilator need to be perturbed in order to disrupt the periodic signal. This is expected, given that all circuits of the LCR are necessarily dynamically coupled in a cell (as demonstrated by the existence of a periodic signal prior to perturbation), since they produce and are affected by identical proteins, which are equally available to interact with any of the circuits.

To exemplify this, we implemented stochastic models of the SCR and the LCR (based on a model proposed by Zhu *et al.*¹¹). The methods are described in the ESI.† We simulated the two models for each of the 1000 instances (cells) and extracted the functionality of each cell prior to and after perturbation. In the model with three repressilator copies the functionality is

reduced from 81.7% to 1.09% as a result of the perturbation, while in the model of the SCR the corresponding numbers are 96.3% and 9.60%. The functionalities do not differ significantly from the measurements in either case (p -values larger than 0.01), as determined by a set of Fisher's exact tests. Meanwhile, the LCR and SCR models differ significantly (p -values are 5.2×10^{-27} and 6.5×10^{-19} for before and after perturbation, respectively).

Conclusions and discussion

We inserted the genetic repressilator of Elowitz and Leibler² into a single-copy F-plasmid to obtain, to our knowledge, the first functional, synthetic, single-copy, ring-type genetic clock.

The SCR was found to exhibit stronger fluctuations (the lower copy number is expected to decrease the rhythmicity of the coupled system²⁴) and longer mean periods and, as such, to differ in dynamics from the LCR. Regardless, the signal of the SCR is stable enough so as to maintain its main feature, periodicity. Interestingly, this difference in dynamics is a demonstration that the stability of the signal of the original LCR relies, to some extent, on the existence of more than one functional copy of the repressilator in each cell. In addition, as these multiple copies exhibit a periodic signal, one can conclude that they are dynamically coupled (as expected, given the indistinguishability between the proteins they produce and are regulated by).

On the other hand, the response of the two systems to temperature changes is similar. In both circuits, the mean period is minimized and the functionality is maximized at 30 °C. Also, both systems lose robustness at temperatures above 30 °C. These behaviours have been explained in a previous study.³

There are only two differences in their response to temperature changes. First, the functionality appears to have a more rapid decrease with increasing temperature in the SCR. Second, the most temperature-sensitive regions of the two systems differ (between 28 °C vs. 30 °C in the LCR and between 30 °C vs. 33 °C in the SCR). At present, we do not have sufficient information to further investigate the causes of these two differences between the SCR and the LCR, particularly since it is presently unknown which underlying parameters regulate the functionality. Our study suggests that the number of functional circuits in a cell is likely one of these parameters.

Finally, we studied the effects of external perturbations on the robustness of the repressilator as a function of the copy numbers and the perturbation strength. First, we observed that the LCR is less robust to a constant perturbation (50 μ M of IPTG in the medium) than the SCR, which shows that not all copies of a repressilator in a cell have to be perturbed in order to disrupt the periodic signal. This result was exemplified using a model, which assumed perfect coupling within a cell and no differences in the promoter strength of the two circuits. Consequently, we find it reasonable to hypothesize that the measured differences between the dynamics of the LCR and the SCR are solely due to the differences in copy-numbers.

From the perturbation studies, we also observed that the SCR is sensitive to the strength of the perturbation, which is particularly relevant in that it increases the number of possible future applications for this circuit.

Overall, we find that the differences in robustness to external perturbations as well as the differences in the dynamics of the two circuits reported here justify the need for a version of the synthetic repressilator implemented on a single-copy plasmid. In particular, its higher robustness to external perturbations and higher functionality at lower temperatures allow the SCR to be more useful than the LCR under certain conditions (*i.e.* by being a more robust clock). This is important for future efforts of synthetic biology aiming to engineer artificial genetic circuits whose proper functioning requires robust time tracking.

Finally, our study also provides much needed empirical data for developing more accurate models of coupled genetic circuits which, so far, have relied on arbitrary parameter values (see *e.g.* ref. 5, 6, 25 and 26). In this regard, the observed fluctuations in the length of the oscillations strongly supports the need to use detailed stochastic modelling strategies^{27,28} to accurately mimic the behaviour of the circuits.

Acknowledgements

This work is supported by Academy of Finland [126803 to A.S.R.], Fundacao para a Ciencia e a Tecnologia [PTDC/BBB-MET/1084/2012 A.S.R.], and Jenny and Antti Wihuri Foundation [A.H.]. We thank the Helsinki Bio-center for sequencing services. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Notes and references

- 1 M. D. Wolf and A. P. Arkin, *Curr. Opin. Microbiol.*, 2003, **6**, 125–134.
- 2 M. Elowitz and S. Leibler, *Nature*, 2000, **403**, 335–338.
- 3 J. G. Chandraseelan, S. M. D. Oliveira, A. Häkkinen, H. Tran, I. Potapov, A. Sala, M. Kandhavelu and A. S. Ribeiro, *Mol. BioSyst.*, 2013, **9**, 3117–3123.
- 4 R. Lutz and H. Bujard, *Nucleic Acids Res.*, 1997, **25**(6), 1203–1210.
- 5 A. Ribeiro, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **75**, 061903.
- 6 A. Ribeiro and S. A. Kauffman, *J. Theor. Biol.*, 2007, **247**, 743–755.
- 7 J. A. Klappenbach, J. M. Dunbar and T. M. Schmidt, *Appl. Environ. Microbiol.*, 2000, **66**, 1328–1333.
- 8 B. S. Stevenson and T. M. Schmidt, *Appl. Environ. Microbiol.*, 2004, **70**, 6670–6677.
- 9 M. Lynch and J. S. Conery, *Science*, 2000, **290**, 1151–1155.
- 10 Y. Mileyko, R. I. Joh and J. S. Weitz, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**(43), 16659–16664.
- 11 R. Zhu, A. S. Ribeiro, D. Salahub and S. A. Kauffman, *J. Theor. Biol.*, 2007, **246**, 725–745.

- 12 A. S. Ribeiro, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **75**(6), 061903.
- 13 A. S. Ribeiro, *Phys. Rev. E: Stat., Nonlinear, Soft Matter Phys.*, 2007, **76**(5), 051915.
- 14 G. Gordon, D. Sitnikov, C. Webb, A. Teleman, A. Straight, R. Losick, A. Murray and A. Wright, *Cell*, 1997, **90**, 1113–1121.
- 15 T. Ogura and S. Hiraga, *Proc. Natl. Acad. Sci. U. S. A.*, 1983, **80**, 4784–4788.
- 16 D. Bray, *Nature*, 1995, **376**, 307–312.
- 17 V. Dinçbaşı, A. Hortaçsu and A. Çamurdan, *Biotechnol. Prog.*, 1993, **9**, 218–220.
- 18 M. Smith and M. Bidochka, *Can. J. Microbiol.*, 1998, **44**, 351–355.
- 19 X. Chen, Z. Xu, P. Cen and W. Wong, *Biochem. Eng. J.*, 2006, **28**, 215–219.
- 20 A. P. Dempster, N. Laird and D. B. Rubin, *J. R. Stat. Soc.*, 1977, **39**, 1–38.
- 21 N. K. Jana, S. Roy, B. Bhattacharyya and N. C. Mandal, *Protein Eng.*, 1999, **12**, 225–233.
- 22 K. S. Koblan and G. K. Ackers, *Biochemistry*, 1991, **30**, 7817–7821.
- 23 K. S. Koblan and G. K. Ackers, *Biochemistry*, 1991, **30**, 7822–7827.
- 24 J. Garcia-Ojalvo, M. B. Elowitz and S. H. Strogatz, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 10955–10960.
- 25 A. Goldbeter and O. Pourquié, *J. Theor. Biol.*, 2008, **252**, 574–585.
- 26 D. Gonze, *BioSystems*, 2010, **99**, 60–69.
- 27 H. H. McAdams and A. Arkin, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 814–819.
- 28 A. S. Ribeiro, R. Zhu and S. A. Kauffman, *J. Comput. Biol.*, 2006, **13**, 1630–1639.

Electronic supplementary information

Single-cell kinetics of the Repressilator when inserted into a single-copy plasmid

Samuel M.D. Oliveira^{a,*}, Jerome G. Chandraseelan^{a,*}, Antti Häkkinen^a, Nadia S.M. Goncalves^a and Andre S. Ribeiro^{a,+}

*^aLaboratory of Biosystem Dynamics, Tampere University of Technology,
P.O. Box 553, 33101 Tampere, Finland*

*** Equal contributions**

+ Corresponding author

Andre S. Ribeiro
Tampere University of Technology,
P.O. Box 553, 33101 Tampere, Finland
E-mail: andre.ribeiro@tut.fi.
Tel: +358408490736

Supplementary Methods

Media and chemicals

Media used were Lysogeny Broth (LB) and minimal nutrient (M63) with the following components, respectively: (i) 10g/L of Tryptone (Sigma Aldrich, USA), 5g/L of yeast extract (LabM, UK) and 10g/L of NaCl (LabM, UK); (ii) 2 mM MgSO₄·7H₂O (Sigma-Aldrich, USA), 7.6 mM (NH₄)₂SO₄ (Sigma Life Science, USA), 30 μM FeSO₄·7H₂O (Sigma Life Science, USA), 1 mM EDTA (Sigma Life Science, USA), 60 mM KH₂PO₄ (Sigma Life Science, USA) pH 6.8 with Glycerol 0.5% (Sigma Life Science, USA) and Casaminoacids 0.1% (Fluka Analytical, USA).

Isopropyl β-D-1-thiogalactopyranoside (IPTG) was used for testing the effects of external perturbations on the Repressilator. All antibiotics used for SCR and LCR strains culturing were purchased from Sigma-Aldrich (USA), respectively: (i) 35 mg/mL kanamycin and 35 mg/mL chloramphenicol; (ii) 35 mg/mL kanamycin and 20 μg/mL Ampicillin. Agarose (Sigma Life Science, USA) was used for the microscopic slide gel preparation.

Bacterial strains and single-copy repressilator plasmid construction and validation

Cells of *E. coli lac*⁻ strain MC 4100 with the repressilator (pZS1-ITlrLLtCL) here named, low-copy repressilator (LCR), and the reporter plasmid (pZE21-GFPaav) were generously provided by M. Elowitz, Princeton University, NJ, USA. Cloning and measurements were performed on this strain.

To construct the single-copy F-plasmid repressilator (SCR) system pBAC2-ITlrLLtCL, we amplified the functional repressilator cassette from the original plasmid (de Novo *Sma*I restriction sites were added to the end of cassette during this procedure). The primers used were:

1-Rep.SmaI-Fw: 5' CCCGGGTCGAGAATTGTGAGCG 3'

2-Rep.SmaI-Rev: 5' CCCGGGTCAAGCTGCTAAAGCGTAG 3'

The vector, pTB-BAC2 F-plasmid, containing the origin of replication and Chloramphenicol resistance gene, was amplified using PCR, also amplified with *Sma*I restriction enzyme sites, using the following primers:

3-Sc.ori.Cam-SmaI-Fw: 5' CCCGGGTTCGAACGCGTATGCATGAG 3'

4-Sc.ori.Cam-SmaI-Rev: 5' CCCGGGTTAGGGCCGTCGACCAA 3'

The amplified sequences of the repressilator and pTB-BAC2 vector were digested using *Sma*I and then ligated. The plasmid was then transferred into *lacI* *E. coli* MC4100 containing the reporter plasmid.

We validated the SCR construction as follows. First, we performed gel electrophoresis (Fig.S4, for the construct; Fig. S5 for the final product) to confirm the presence of the plasmid containing the SCR.

Next, a fraction of the new plasmid (covering the vector and insert) was amplified by PCR and sequenced using appropriate primers. The sequence obtained from sequencing was aligned with the expected sequence using BLAST. The results confirm the SCR construction (Fig. S6).

Sequencing for confirmation of the plasmid

A part of the plasmid covering the vector and the insert was amplified from the Chloramphenicol resistance gene in the vector to the *tetR* region of the Repressilator, and sequenced. The primers used for amplification were:

5. CmR-1-F: CCGCTGGCGATTCAGGTTC

6. *tetR*-3-R: AGCAAAGCCCGCTTATTTTTTACATG

The alignment of sequence obtained from sequencing against the expected sequence using NCBI BLAST²² is shown in Fig S6.

qPCR for SCR validation and LCR copy-number estimation

qPCR was used to validate whether the SCR plasmid is functional in the host cell. For that, we quantified the levels of mRNA of the component genes, *lacI*, *tetR* and *cl*, and compared their expression levels against the No-Template and No-RT controls (see Methods of the main manuscript). We also compared the mRNA numbers of *lacI*, *tetR* and *cl* of the SCR and of LCR plasmids.

lacI *E. coli* MC 4100 cells, containing either LCR or SCR with the reporter systems, were grown following the same culturing protocols described in the methods section of the manuscript. After 10 hours of culturing in 5 mL liquid M63 medium at shaking 250 rpm, one sample of the LCR or SCR strain was taken, and rifampicin was immediately added, to prevent further transcription. RNA protect reagent was used to fix cells before their enzymatic lysis with Tris-EDTA lysozyme buffer (pH 8.3). RNA was isolated from cells using RNeasy mini-kit (Qiagen) following manufacturer's instructions. 1 µg of RNA was used as a starting material. To ensure purity of the

RNA, the RNA samples were treated with DNase free of RNase to remove residual DNA. Next, RNA was reverse transcribed into cDNA using iSCRIPT reverse transcription super mix (Biorad). qPCR was performed using Power SYBR-green master mix (Life Technologies) with primers for the amplification of the target and the reference genes at a concentration of 200 nM. Reactions were carried out in 20 µL reactions triplicates with 500 nM per primer. The following primers were used for quantification:

-For *lacI* gene:

7-lacI.pro-Fw: 5' GTGGTGTCTGATGGTAGAACG 3'

8-lacI.pro-Rev: 5' CTGTTGATGGGTGTCTGGTC 3'

- For *tetR* gene:

9-tetR.pro-Fw: 5' CGCTGTGGGGCATTTCAC 3'

10-tetR.pro-Rev: 5' AAGAAGGCTGGCTCTGCAC 3'

- For *cI* gene:

11-cI.pro-Fw: 5' GATGCGGAGAGATGGGTAAG 3'

12-cI.pro-Rev: 5' ACTCATCACCCCCAAGTCTG 3'

The length of amplicons was kept at 90 bp. The sequences of the primers of the reference gene 16 S rRNA (EcoCyc Accession Number: EG30090) were obtained from Thermo Scientific:

13- Fw-5' CGTCAGCTCGTGTGTGAA 3'

14- Rev 5' GGACCGCTGGCAACAAAG 3'

The level of each target gene was normalized to the level of the 16 S rRNA for all samples. The PCR cycling protocol used was 94 °C for 15 s, 51 °C for 30 s, and 72 °C for 30 s, up to 39 cycles. We used NO-RT enzyme and NO-Template as controls. The C_q values were obtained from the CFX ManagerTM Software and the fold change of the genes from the LCR were analysed using as reference their expression in the SCR according to the Livak method (reference⁴ in the main manuscript). Results are described in the main manuscript.

Supplementary Figures

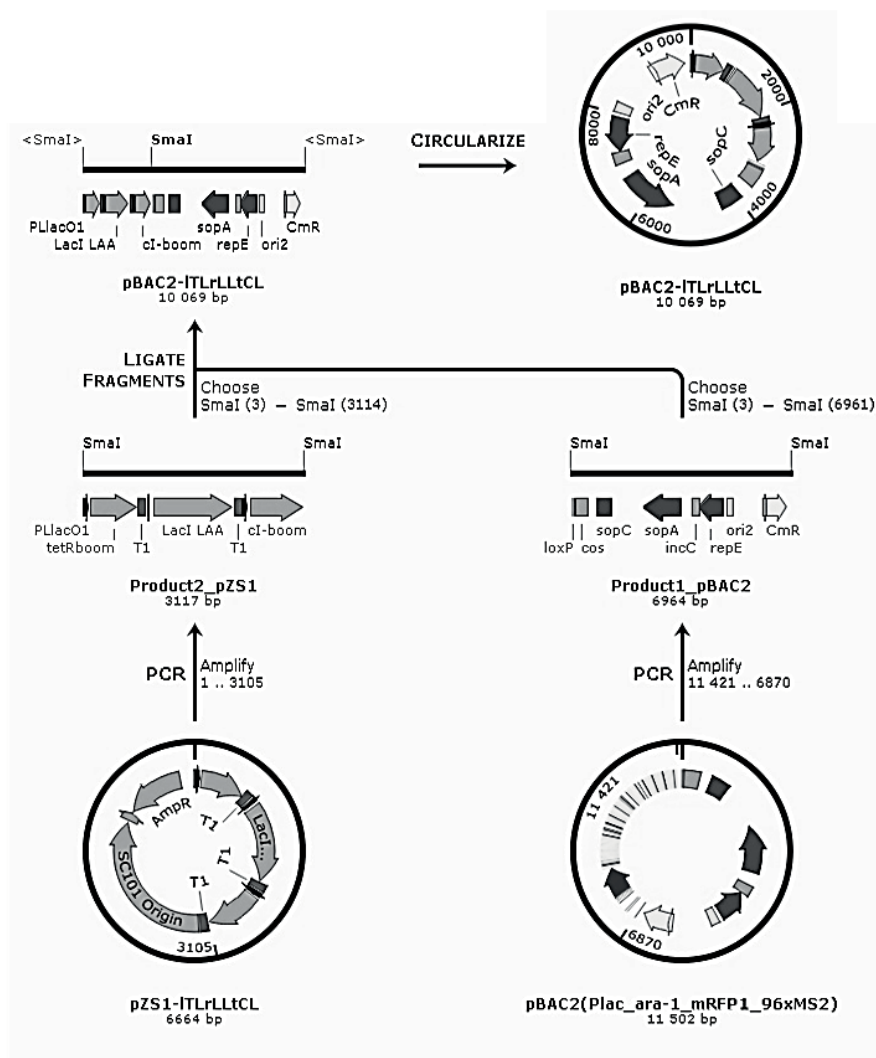


Fig. S1 Plasmids used for the construction of the SCR plasmid. The pBAC2-ITLrLLtCL plasmid was engineered by inserting the repressor cassette into pBAC2 (*P_{lac/ara-1}-mRFP1-MS2-96x*) vector (generously provided by Ido Golding of the University of Illinois, USA), containing the single-copy origin of replication (the construction history was generated and adapted using SnapGene® 1.5.2).

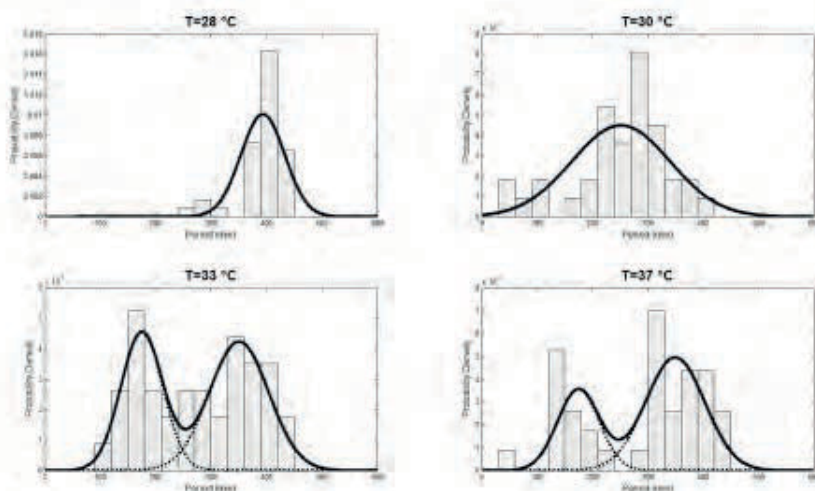


Fig. S2 LCR period distributions at different temperatures. Dashed lines represent the probability densities of the fitted model with two Gaussians. Solid lines represent the densities of individual components in the case of one and two Gaussians. Magnitudes were scaled to represent the probability density.

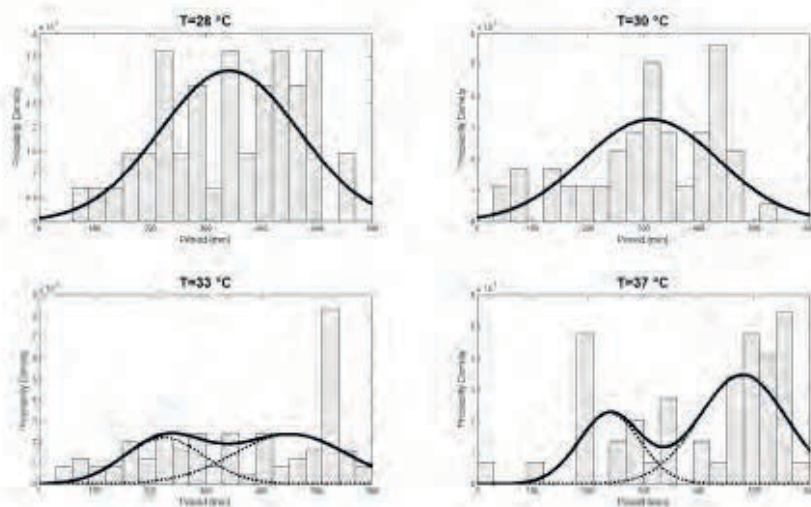


Fig. S3 SCR period distributions at different temperatures. Dashed lines represent the probability densities of the fitted model with two Gaussians. Solid lines represent the densities of individual components in the case of one and two Gaussians. Magnitudes were scaled to represent the probability density.

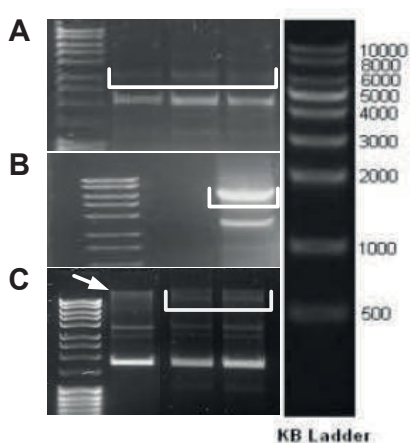


Fig. S4 Split gels for intermediate steps of the SCR plasmid construction. (A) PCR fragment of 3114 bp amplified from original pZS1-ITlrLLtCL with appropriate primers (in triplicate). (B) Lane containing pTB-BAC2 backbone amplified region with the single-copy origin of replication (6961 bp). (C) Lane 1: plasmid profile of the strain containing only the reporter plasmid. Lanes 2 and 3: two replicates of the plasmid profile of the strain with the reporter and the final construct pBAC-ITlrLLtCL (SCR plasmid, 10069 bp). Note the white arrow and the white lines (indicating the lanes). The numbers of the DNA ladders on the left side of Figures A, B, and C, are shown on an identical ladder on the right side of the figure, for easier visualization.

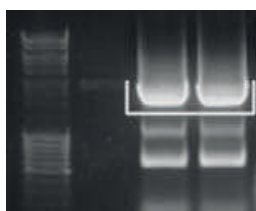


Fig. S5 Split gel of the final product of the SCR plasmid construction. Lane 1: unused. Lanes 2 and 3: PCR amplification of 3114 bp from the SCR plasmid, pBAC-ITlrLLtCL, with appropriate primers. Note the white line indicating the lanes. The ladder is identical to those in Fig. 1.

```

Query 1125 GATGGCTTCCATGTCTGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAG 1184
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 4 GATGGCTT-CATGTCTGGCAGAATGCTTAATGAATTACAACAGTACTGCGATGAGTGGCAG 62

Query 1185 GGCGGGGCGTAAttttttAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCC 1244
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 63 GGCGGGGCGTAATTTTTTAAGGCAGTTATTGGTGCCCTTAAACGCCTGGTTGCTACGCC 122

Query 1245 TGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGATGATAAGCTGTCAAACATGA 1304
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 123 TGAATAAGTGATAATAAGCGGATGAATGGCAGAAATTCGATGATAAGCTGTCAAACATGA 182

Query 1305 GAATTGGTCGACGGCCCTAACCCGGGTCGAGAATTGTGAGCGGATAACAATTGACATTGT 1364
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 183 GAATTGGTCGACGGCCCTAACCCGGGTCGAGAATTGTGAGCGGATAACAATTGACATTGT 242

Query 1365 GAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCGAATTCATTAAAGAG 1424
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 243 GAGCGGATAACAAGATACTGAGCACATCAGCAGGACGCACTGACCGAATTCATTAAAGAG 302

Query 1425 GAGAAAGGTACCATGTCCAGATTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTG 1484
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 303 GAGAAAGGTACCATGTCCAGATTAGATAAAAGTAAAGTGATTAACAGCGCATTAGAGCTG 362

Query 1485 CTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTA 1544
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 363 CTTAATGAGGTCGGAATCGAAGGTTTAACAACCCGTAAACTCGCCCAGAAGCTAGGTGTA 422

Query 1545 GAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCT 1591
          ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Sbjct 423 GAGCAGCCTACATTGTATTGGCATGTAAAAAATAAGCGGGCTTTGCT 469

```

Fig. S6. Alignment of sequence obtained from sequencing, using primer CmR-1-F, against the expected sequence.

PUBLICATION III

Temperature-Dependent Model of Multi-Step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements

S.M.D. Oliveira*, A. Häkkinen*, J. Lloyd-Price, H. Tran, V. Kandavalli, and A.S. Ribeiro

PLoS Comp. Biol., 12(10): e1005174, 2016. *Equal contributions.
doi: 10.1371/journal.pcbi.1005174

Publication reprinted with the permission of the copyright holders.

RESEARCH ARTICLE

Temperature-Dependent Model of Multi-step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements

Samuel M. D. Oliveira[✉], Antti Häkkinen[✉], Jason Lloyd-Price, Huy Tran, Vinodh Kandavalli, Andre S. Ribeiro*

Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, Tampere, Finland

✉ These authors contributed equally to this work.

* andre.ribeiro@tut.fi



OPEN ACCESS

Citation: Oliveira SMD, Häkkinen A, Lloyd-Price J, Tran H, Kandavalli V, Ribeiro AS (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

Editor: Teresa M. Przytycka, National Center for Biotechnology Information (NCBI), UNITED STATES

Received: May 2, 2016

Accepted: September 23, 2016

Published: October 28, 2016

Copyright: © 2016 Oliveira et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: Work supported by Vilho, Yrjö and Kalle Väisälä Foundation (SMDO), Alfred Kordelin Foundation (AH), Tampere University of Technology President's Graduate Programme (JLP), Academy of Finland: General Research Grant (No. 295027; ASR) and Key Project Funding (No. 305342; ASR), and Jane and Aatos Erkkö

Abstract

Transcription kinetics is limited by its initiation steps, which differ between promoters and with intra- and extracellular conditions. Regulation of these steps allows tuning both the rate and stochasticity of RNA production. We used time-lapse, single-RNA microscopy measurements in live *Escherichia coli* to study how the rate-limiting steps in initiation of the $P_{lac/ara-1}$ promoter change with temperature and induction scheme. For this, we compared detailed stochastic models fit to the empirical data in maximum likelihood sense using statistical methods. Using this analysis, we found that temperature affects the rate limiting steps unequally, as nonlinear changes in the closed complex formation suffice to explain the differences in transcription dynamics between conditions. Meanwhile, a similar analysis of the P_{tetA} promoter revealed that it has a different rate limiting step configuration, with temperature regulating different steps. Finally, we used the derived models to explore a possible cause for why the identified steps are preferred as the main cause for behavior modifications with temperature: we find that transcription dynamics is either insensitive or responds reciprocally to changes in the other steps. Our results suggests that different promoters employ different rate limiting step patterns that control not only their rate and variability, but also their sensitivity to environmental changes.

Author Summary

Temperature affects the behavior of cells, such as their growth rate. However, it is not well understood how these changes result from the changes at the single molecule level. We observed the production of individual RNA molecules in live cells under a wide range of temperatures. This allowed us to determine not only how fast they are produced, but also how much variability there is in this process. Next, we fit a stochastic model to the data to identify which rate-limiting steps during RNA production are responsible for the observed differences between conditions. We found that genes differ in how their RNA production

Foundation (No. 610536; ASR). 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.

is limited by different steps and in how these are affected by the temperature, which explains why different genes respond differently to temperature fluctuations.

Introduction

Temperature is known to affect gene expression patterns in cells. This has profound effects, as changes in transcription and translation dynamics propagate to the behavior of genetic networks, which manifests in their sensitivity to temperature changes [1–3].

The expression patterns are not solely characterized by the rates at which the genes are expressed, but also by the associated stochasticity. The latter affects the phenotypic variability of populations of genetically identical cells [4–6] and the temporal variations in the behavior of the individual cells [7]. In unicellular organisms, such variations, even at the level of single molecules [8], can determine a cell fate.

In bacteria, much of the stochasticity in gene expression stems from transcription [9]. Live-cell measurements report that different promoters and intra- and extracellular conditions result in wide differences in transcription dynamics, both in rate and stochasticity [10]. Sub-Poissonian [11], Poissonian [12], and super-Poissonian [10, 13] dynamics (featuring less, equal to, and more variability than a corresponding Poisson process, respectively) have been reported, each resulting from a different combination of mechanistic properties that shape RNA production dynamics.

The way the effects of temperature changes on transcription kinetics propagate to the cellular behavior is still poorly understood. Rates of biochemical reactions are known to be affected by temperature changes, as dictated by the laws of physics, such as the Arrhenius law; however, biochemical laws have been found to not suffice to explain changes in more complex biological processes such as bacterial growth [14].

For example, it is expected that the number of RNA polymerases, the rate at which they work, and RNA lifetimes are affected by temperature. Also, at suboptimal temperatures, *Escherichia coli* shifts to specific expression patterns by changing in transcription factor numbers [15], by regulating of the relative σ -factor concentrations [16], and by affecting DNA conformation: negative supercoiling increases at low temperatures, and relaxes at high ones [17]. Despite these findings, quantitative information of the changes and on their contribution to the changes in transcription dynamics is still lacking.

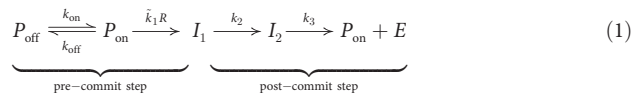
Some progress toward these goals has been made through *in vitro* measurements of the closed and open complex formation dynamics [18–20]. Another study reports that RNA polymerase, rRNA, and tRNA concentrations, and the fraction of stable RNA remained constant, while the elongation rate, ppGpp concentration, and cell growth rate increase to up to 40°C, while after 40°C the changes become complex and, e.g. the growth rate decreases [21]. So far, these studies only identified changes in mean expression rates. Further information on other dynamical properties, such as stochasticity, is required. In addition, it is unclear to what extent these measurements reflect what occurs in live cells.

Here, we study the transcription dynamics of the $P_{\text{lac/ara-1}}$ promoter in live *E. coli* at the single RNA level, under various inducer concentrations, for a wide range of temperatures. Using statistical models, we identify the most likely changes with temperature in the rate limiting steps of transcription initiation. This expands on our previous work [22–24] in that the analysis at different temperatures allows identifying more complex underlying details of the kinetics of transcription. Also, we test if similar changes are observed in the P_{tetA} promoter. Finally, we use the inferred models to study the possible causes of the underlying changes.

Materials and Methods

Model of Transcription Initiation

We use the following model of transcription initiation [25], which combines the active-inactive promoter model [26] with a sequential model of transcription initiation [27, 28]:



where P_{off} and P_{on} represent the promoter in an inactive (i.e. repressed) and active (i.e. free from repressors or bound by an activator) states, respectively, R represents RNA polymerases, and I_k are intermediate complexes of transcription initiation. Finally, the product E represents the elongation complex. As the RNA polymerase numbers are not observed, we let $k_1 \doteq \tilde{k}_1 R$ to represent the effective forward rate for an active gene.

In the model, the promoter switches between being active and inactive (on/off) for transcription, i.e. whether an RNA polymerase can unobstructedly reach the start site and initiate transcription [9], depending on the binding and unbinding of regulatory molecules [9, 29]. In some cases, these molecules can also affect subsequent steps [28, 30].

Note that once the promoter is in a state that allows the RNA polymerase to bind and form the closed complex, not necessarily will this result in the production of an RNA, as the closed complex can reverse to the previous state [20, 31]. As such, the system is not yet fully committed to transcription. This “full commitment” only occurs once reaching the next state (I_1). At this stage, it becomes highly unlikely that any reversion occurs (e.g. in the λ PR promoter, I_2 is *always* unstable compared to the complete open complex state, even at 0°C [31]).

Once this commitment occurs, it is followed by a sequence of steps responsible for the formation of a stable open complex, which includes the isomerization steps [20, 28, 31, 32]. The last step in this model represents the complex escaping the start site, clearing the promoter region. Note that the reversibility of the closed complex formation [20, 24] does not reduce the applicability of the model [24, 25]. Also, regardless of this reversibility, the model can still be separated into an R -dependent (pre-commit) and R -independent (post-commit) parts.

The ability of regulatory molecules to create an on/off promoter dynamics allows bursty RNA production when fast production events are separated by long, random off periods (due to, e.g. slow repressor unbinding) [7]. If this process dominates RNA production, the transcription intervals are highly noisy (coefficient of variation $c_v \geq 1$). Meanwhile, if the subsequent sequential process dominates transcription, the intervals between RNA production events are more regular, resulting in less noisy transcription ($c_v \leq 1$).

Which and how many steps most contribute to the observed transcription dynamics appears to depend on the promoter and intra- and extracellular conditions [10, 11, 13, 22, 23]. In what concerns modeling this process, for example, if the promoter’s visit to the off-state or a sequential step are fast, these can be eliminated from the model, as they do not contribute significantly to the transcription dynamics. The steps with most influence on the transcription rate are called rate limiting [32].

Here we equate the transcription intervals with those of transcription initiation, which implies that transcription elongation is neglected. This is justified by the fact that, on average, elongation is not expected to affect the transcription intervals, as each transcript is expected to be delayed by a similar time. As a result, elongation primarily adds solely extra variance in the inter-transcription intervals. However, given the timescales of elongation and transcription initiation, this additional variance can be ignored: e.g. chain elongation at 50 nt/s [33] for a

5000 nt gene is expected to add a noise term with a standard deviation of 2 s [25]. Even considering elongational pauses under GTP-starved conditions [34], the elongation noise term is therefore likely negligible at the resolution of our measurements (cf. sampling interval of 60 s). Further, we note that no differences have been found between genes with elongation regions of different length [7, 35].

As each of the steps are complex processes rather than elementary chemical reactions, it is not yet well understood how the temperature affects of each of the steps in Eq (1) in live cells. For this reason, we let each of the model parameters change as a function of the temperature according to a polynomial function, rather than according to some biochemical model. As we have relatively few samples on the temperature axis, a quadratic equation was deemed sufficient to capture either temperature-independent or a linear and nonlinear temperature-dependent relationship:

$$k_x(T)^{-1} = a_{k_x^{-1},2} T^2 + a_{k_x^{-1},1} T + a_{k_x^{-1},0} \quad (2)$$

where $a_{k_x^{-1},j}$ are the order- j coefficients of the polynomial of the temperature-dependence of the parameter k_x , and T is the temperature. While this model is not expected to provide particular insight of the mechanisms of temperature-dependence (i.e. $a_{k_x^{-1},j}$ do not encode a particular physical meaning), it allows us to identify how each parameter likely responds to temperature changes, be it independent or dependent, linear or nonlinear, or monotonic or bitonic.

The parameters of the model can be estimated from the transcription intervals in a maximum likelihood sense [25]. Confidence in the model parameters can be estimated using the delta method [36]. Further, to determine if each of the steps in the above model play a significant role in the overall dynamics, we use the Bayesian information criterion (BIC) [37] as the model selection criterion. A difference in BICs (Δ BIC) of 0 to 2, 2 to 6, or ≥ 6 indicates weak, positive, or strong evidence, respectively, against the model with the greater BIC [37]. For the purposes of BIC, a censored sample is assumed to be worth 0.5 exact samples (the exact value varies depending on how badly the sample is censored). To avoid drawing false conclusions due to this approximation, we also compute a lower bound for Δ BICs when determining if the best models fits significantly better than the alternatives (see S1 Appendix).

Measurements and Data Acquisition

Transcription intervals in individual cells were measured in live *E. coli* using the MS2-GFP RNA-tagging system [7]. The cells feature a multi-copy reporter gene expressing MS2-GFP and a single-copy target gene, controlled by the promoter of interest, containing 96 MS2-GFP binding sites. Shortly after a target RNA is produced, the binding sites are quickly occupied by the abundant GFPs, allowing the fully tagged RNAs to be visualized using fluorescence microscopy [38]. Once formed, the RNA-96-MS2-GFP complex remains fluorescent for much longer than the cell lifetime [38]. The constructs used here were engineered previously [7, 11]. An example microscopy image is shown in Fig 1.

The activity of the target genes was also analyzed by quantitative PCR (qPCR) as a function of the media composition. To quantify intracellular RNA polymerase concentrations in each condition, we measured the amount of RpoC subunits by western blot (WB), as these are the limiting factor in the assembly of the RNA polymerase holoenzyme [39, 40]. Measuring gene activity as a function of the RNA polymerase abundance allows extrapolating the relative duration of the pre- and post-commit steps [24, 32].

More details of the constructs, measurement procedures, and data acquisition methods is given in S1 Appendix.

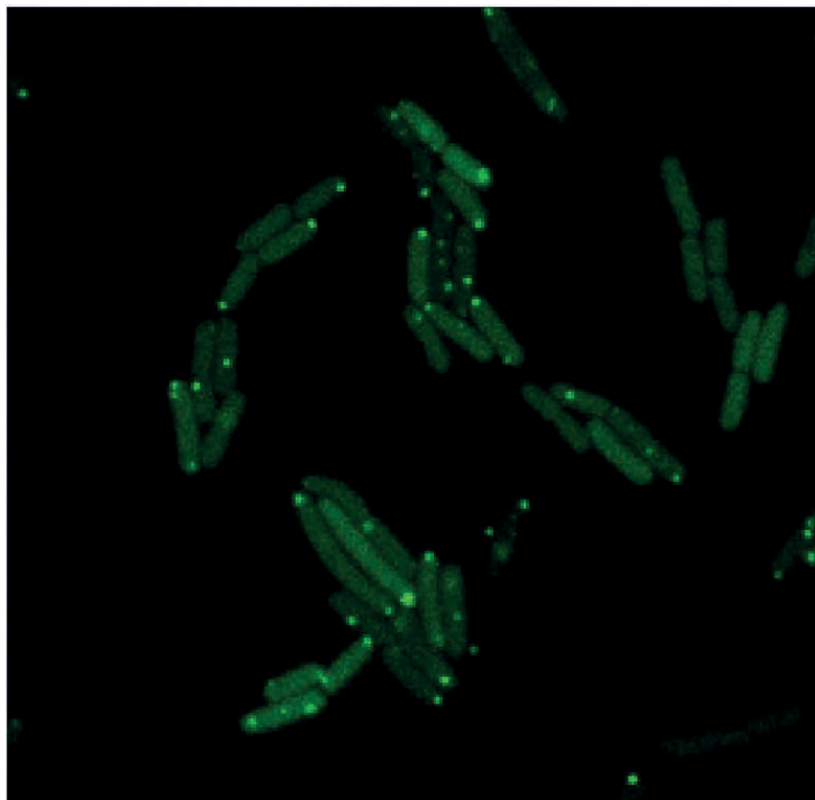


Fig 1. Example of a confocal microscopy image of *Escherichia coli* expressing the MS2-GFP and the target RNA. The cells are visible due the abundant MS2-GFP while the GFP-tagged target RNA appear as bright spots. The image was acquired at 35°C.

doi:10.1371/journal.pcbi.1005174.g001

Results and Discussion

Effects of Temperature on the $P_{lac/ara-1}$ Promoter under Various Inducer Concentrations

We studied how the distribution of durations between consecutive RNA productions in multiple, individual cells (afterwards denoted by transcription intervals) of the $P_{lac/ara-1}$ promoter differ with temperature. This study was conducted for each possible induction scheme of this promoter, so as to assess if the temperature-dependence of the initiation kinetics is inducer scheme-dependent. Those induction schemes are: a) 1 mM of IPTG and 0.1% of arabinose (denoted by Full), b) 1 mM of IPTG only (IPTG), and c) 0.1% of arabinose only (Ara). In each condition, we recorded multiple time series of 120 minutes in length with a 1 minute sampling interval for temperatures between 24 and 41°C (see [S1 Appendix](#) for details).

We first quantified how the mean and the standard deviation (sd) of the transcription intervals change with temperature. As the cell division time varies significantly between conditions, a procedure accounting for the uneven truncation of the intervals must be used. For this, the

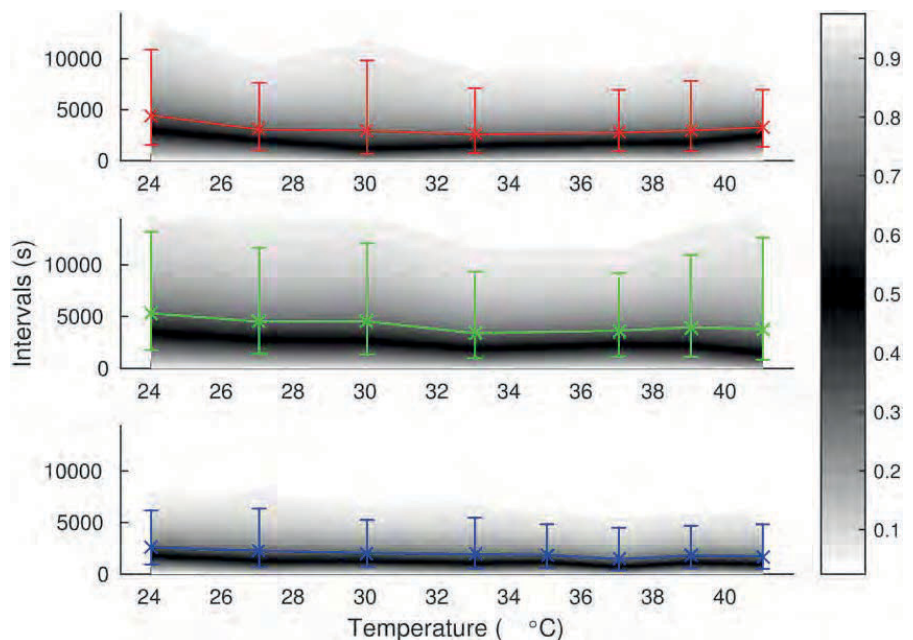


Fig 2. Distributions of transcription intervals in individual cells as a function of temperature for the $P_{lac/ara-1}$ promoter. The gray gradients represent quantiles of the interval distributions, as indicated by the color bar on the right. The crosses denote means, and the error bars represent the lower and upper standard deviations of the distributions. Cases from top to bottom: IPTG (red), Ara (green), Full (blue).

doi:10.1371/journal.pcbi.1005174.g002

right-censored (unobserved) transcription intervals at the end of each time series are communicated to the estimator as in [25]. First, we used gamma distribution as the model, as it can have any mean and sd independent of each other (a model must be assumed in order to estimate moments from the censored data). The results are tabulated in [S1 Table](#). The distributions are shown in [Fig 2](#). These data suggest that, the transcription interval duration changes by about a 2× factor along the range of temperatures tested, for each induction scheme. In addition, these changes appear to be nonlinear, and even non-monotonic in some cases. Meanwhile, the standard deviation of the transcription intervals tends to follow the mean, resulting in a coefficient of variation (c_v) of slightly above unity with a slight increase as temperature increases. The median-to-mean ratio is approximately constant with respect to temperature at about 0.61, which indicates that the temperature affects the whole distribution and not only the long intervals.

Next, we fit the data with the model of [Eq \(1\)](#) independently in each induction scheme and temperature, and tested if the on/off switching and which forward steps are most responsible for the observed transcription dynamics using the Bayesian information criterion (BIC) (see [Materials and Methods](#)). The results are summarized in [S2 Table](#). In all cases, we found evidence only for a single sequential rate limiting step. While there is weak evidence that this is true in all cases (ΔBIC lower bound (ΔBIC LB; see [S1 Appendix](#)) against any multi-step model is no less than 1.15), strong (statistically significant) evidence exists that this holds at least in certain temperature ranges (for all induction schemes, there exists a temperature with ΔBIC LB no less than 5.16). In addition, we also found evidence for the presence of on/off switching in

each induction scheme, as there is a condition where the lower bound for ΔBIC is no less than 9.20, providing strong evidence that the on/off switching plays a role, in at least certain temperature ranges, regardless of the induction scheme.

As the order of the steps with rates k_1 , k_2 , k_3 cannot be identified from an individual measurement by this methodology (see e.g. [25]), at this stage we cannot resolve if the rate limiting step occurs prior or after commitment to transcription. To achieve this, we combine the data from different temperatures, as models with equivalent changes in their transcription intervals can be rejected on the basis that they would require too intricate changes in their parameters between conditions (e.g. multiple parameters co-varying in a complex manner). Using this procedure, the order of the last two steps remains unknown, but the changes in each can be identified.

To combine the data, we fit the data jointly in each of the three induction schemes, such that the parameters k_{on}^{-1} , k_{off}^{-1} , k_1^{-1} , k_2^{-1} , and k_3^{-1} are either a) constant, b) vary linearly, or c) vary quadratically as a function of temperature (the quadratic curve representing a nonlinear relationship). Meanwhile, no model is imposed on the changes as a function of induction scheme, as these are likely nonlinear. Again, model selection is used to determine which parameters change significantly with temperature.

We found the preferred model to be the one where k_{on} , k_{off} , k_2 , and k_3 are constant and k_1^{-1} is a non-linear (Full and IPTG) or linear (Ara) as a function of the temperature. The ΔBICs between the second best and the best fitting models were about 1.32, 4.92, and 1.57 (ΔBIC LBs 1.31, 4.91, and 1.30) for Ara, IPTG, and Full, respectively. Meanwhile, we found similar results in models where the step with rate k_3 is removed, with ΔBICs of 1.32, 4.39, and 1.51 for Ara, IPTG, and Full, respectively, indicating that using the higher-order model does affect the identification of the temperature-dependence of the parameters.

In the best fitting models, the relationship between k_1^{-1} and temperature differs under different induction schemes. However, we tested if a similar change in this parameter could explain the changes in all cases, by deriving a combined model where this parameter follows a single function, up to a scale, in the three induction schemes. We found this new model to have smaller BIC with a ΔBIC of 10.6 when compared to the best unconstrained model, (ΔBIC LB of 8.80), which provides strong evidence that the temperature affects the promoter by regulating k_1^{-1} in a similar, nonlinear fashion, regardless of the induction scheme.

Finally, we tested if the changes in k_1^{-1} could be explained solely by the changes in RNA polymerase numbers (cf. R in Eq (1)). For this, we quantified in an independent measurement the relative abundance of the RpoC subunits using western blot analysis for 24, 37, and 41°C. We found that the numbers were 0.415 and 0.562 at 24 and 41°C, respectively, relative to that of at 37°C. By plugging these numbers in our model, we found strong evidence that the parameter k_1 does not follow the changes in RNA polymerase numbers (ΔBIC of 27.9, ΔBIC LB of 27.0), i.e. \tilde{k}_1 is not constant with temperature. Moreover, we found that if the above RNA polymerase numbers were the sole change caused by the temperature, the changes in the transcription intervals ought to be larger than what was observed *in vivo*. This suggests that there are other temperature-dependent changes in \tilde{k}_1 , which attenuates the effects of RNA polymerase number changes.

The means and sds of the transcription intervals resulting from the best fitting model (quadratic k_1^{-1} with a similar pattern in each induction scheme) are shown in Fig 3. In this model, the steps prior to transcription commitment vary between temperature conditions from 3060 to 5000 s, from 2310 to 3770 s, and from 1580 to 2590 s, while the duration of the process after transcription commitment is a constant equal to 399, 7.79, and 1.50 s for Ara, IPTG, and Full induction schemes, respectively. Interestingly, in each case, the temperature-independent post-

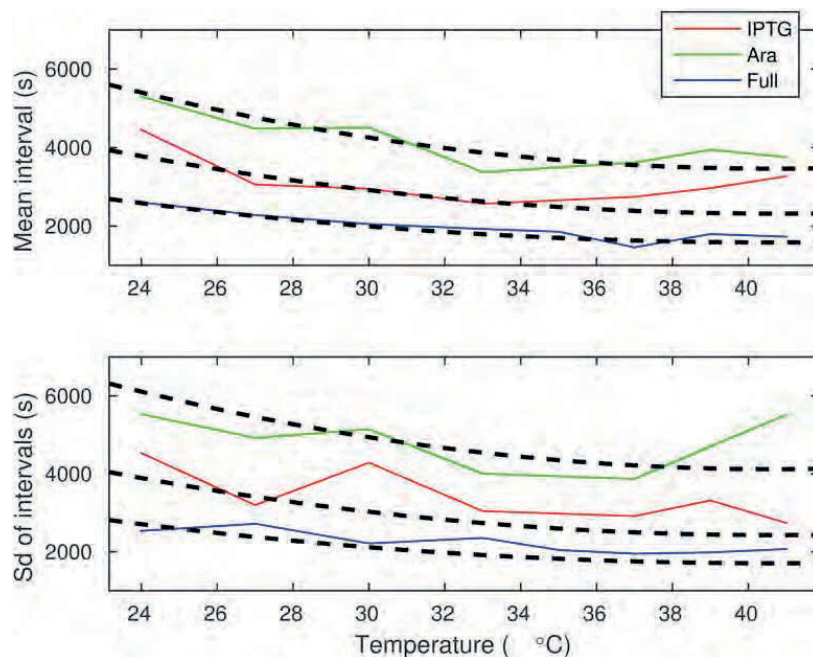


Fig 3. Mean and standard deviation of transcription intervals as a function of temperature for the $P_{lac/ara-1}$ promoter. The solid curves are those shown in Fig 2. The dashed curves represent the means and standard deviations of the best fitting models.

doi:10.1371/journal.pcbi.1005174.g003

commit step tends to be no more than 12% of the total duration of the mean interval between consecutive RNA production. Consequently, as the temperature modulates the longer lasting step in transcription, it allows for large changes in the transcription interval with temperature.

To assess whether the data fit the model as expected, we also performed Monte Carlo simulations to determine if the empirical data fits the model significantly worse than that generated from the appropriate model. For this, we generated a transcription interval from the best fitting model to correspond each sample extracted from the measurements, which was subsequently censored using the same procedure that constrains the empirical data (i.e. observation time is limited by the remaining cell lifetime, and observations occur at intervals of 60 s). This simulation was repeated for 1000 rounds, from which likelihoods were calculated. The fraction of rounds where the simulated data fit worse than the empirical data (cf. a p-value) were 0.230, 0.161, and 0.040 for Ara, IPTG, and Full, respectively, and 0.095 for all cases combined. As such, we found no evidence to reject our model (typically done e.g. when this probability is less than 0.01). The mean and sd for a single simulation is shown in S1 Fig (cf. the equivalent figure for the empirical data in Fig 3).

We compared our results on the relative durations of the pre- and post-commit steps with an independent method, namely, by extrapolating the corresponding values from a Lineweaver-Burk plot [24, 32]. For this, we used qPCR to measure the target gene activity and western blot to estimate the RNA polymerase concentrations in different media conditions (see S1 Appendix for details). We found the post-commit step to take up only about 22.5% (sd 14.6%) of the transcription duration for Full induction at 37°C. This confirms the previous result, as

the values reported above, derived from the microscopy measurements, are well within the 95% confidence interval (CI; $[-6.08, 51.1]\%$) of the qPCR/WB measurements.

The parameter values for the best fitting models are listed in [S3 Table](#) and some derived features of the kinetic of the on/off switching in [S4 Table](#). These parameter values suggest that the promoter remains most of the time unavailable for RNA polymerases to bind (small duty cycle). After this, it produces a small ($\ll 1$) burst, whose size is temperature-dependent. The mean interval between bursts appears to be unaffected by the removal of arabinose (cf. IPTG versus Full; as k_{on} is nearly constant and much smaller than k_{off}), while removing IPTG has more complex effects. These results are consistent with observed changes in the median-to-mean ratio.

Two studies, one using fluorescent in situ hybridization (FISH) [\[41\]](#) and the other using a yellow fluorescent protein fusion library for *E. coli* [\[10\]](#) to quantify the cell-to-cell diversity in RNA numbers from several promoters, have reported Fano factors ranging from 1 to 3. Also, in the work by Golding and colleagues [\[7\]](#), it was estimated that the Fano factor in transcript production could be as large as 4.1. These results were taken as indicative of bursty RNA production. However, recent works have shown that much of this cell-to-cell diversity in RNA numbers can be explained by other factors [\[42\]](#), such as RNA degradation [\[43\]](#) and stochastic partitioning of RNA molecules in cell division [\[44\]](#). As we base our study on the transcription intervals between consecutive RNA production events in individual cells, rather than the variability in RNA numbers, our analysis is less susceptible to such errors. In addition, only under specific conditions does a large Fano factor correspond to large burst sizes [\[45\]](#). As such, we do not expect our estimate of the average burst size to be in contradiction with these studies.

Effects of Temperature on the P_{tetA} Promoter

Even though the Lac-Ara-1 Promoter initiation kinetics responded similarly to temperature changes under all induction schemes, this response might not generalize to different promoters with different rate-limiting step configuration. To test this, we studied the effects of temperature changes on the transcription dynamics of the P_{tetA} promoter.

For this, using raw data from [\[11\]](#), we made use of the methodology in [\[25\]](#) in order to extract the single-cell transcription intervals of the P_{tetA} promoter at 24 and 37°C with 15 ng of aTc (Full induction). In each case, the time series were of about 60 minutes in length with 1 minute sampling. [S5 Table](#) tabulates the mean, sd, and c_r of the transcription intervals estimated from the data. The mean transcription interval changed by about a 2.5× factor.

Unlike the $P_{\text{lac/ara-1}}$ promoter, the P_{tetA} promoter appears to have sub-Poissonian dynamics, suggesting that its RNA production kinetics is mostly dictated by a sequence of multiple rate limiting steps, rather than the activation of the promoter. To confirm this, we fit the data with the model of [Eq \(1\)](#) and tested for on/off switching and number of rate limiting steps. The results, shown in [S6 Table](#), provide strong evidence for two or three rate limiting steps (ΔBIC LBs against any single step model were no less than 19.0). Meanwhile, no evidence of significant on/off switching is present (ΔBIC s against best on/off model are 1.14 (LB 0.387) for 24 and 5.74 (LB 5.34) for 37°C).

Finally, we fit the two data points jointly with a 3-step model, where one of the three parameters changes as a function of the temperature (more complex changes cannot be identified with only two temperature conditions). The results suggest that the two common rate limiting steps are about 121 s each in duration, while the changing step is 1910 and 543 s for 24 and 37°C, respectively. As there is no on/off switching, we cannot identify which steps are in the pre-commit and post-commit stages. It is however possible to determine that the relative

duration of the post-commit step must be either 11.3% or 94.4% for 24°C and 30.9% or 84.6% for 37°C.

To determine whether the post-commit step is the longer or shorter one, we extrapolated the relative durations of the post- and pre-commit steps from a Lineweaver-Burk plot [32] based on qPCR and WB measurements. At 37°C, we found the duration of the post-commit step to be about 77.1% (sd 6.64%) of the transcription duration with a 95% CI of [64.1,90.1]%, suggesting that, in the P_{tetA} promoter, the long, temperature-sensitive step occurs after transcription commitment.

Modeling the Effects of Changes in the Rate Limiting Steps

Having measured the dynamics of transcription initiation of $P_{lac/ara-1}$ and P_{tet} as a function of temperature and induction scheme, and assessed which parameter values are most responsible for the changes in the observed dynamics, we next make use of the model extracted from the data to investigate the ability of each parameter of the model in changing the dynamics of RNA production and the overall the range of behaviors possible with this model.

Response and bounds of mean, sd, and cv for parameter changes. We next explored the extent to which the dynamics of RNA production can change as these parameter values change. Namely, we used our model to study how changes in k_{on}^{-1} , k_{off}^{-1} , and k_1^{-1} affect the mean and standard deviation (sd) of the duration of the pre-commit step. The mean, sd, and c_v of this step cannot be freely varied when changing k_{on} or k_{off} only. Specifically, the mean is bounded to $[k_1^{-1}, \infty)$ while the variance is limited to $[k_1^{-2}, \infty)$. Meanwhile, c_v^2 is bounded to $[1, 1 + 2k_1/k_{off}]$ and to $[1, 1 + k_1/(2k_{on})]$ when changing k_{on} and k_{off} only, respectively. For changes in k_{off} , c_v is maximized at $k_{off} = k_{on}$, while for k_{on} , c_v monotonically increases with k_{on}^{-1} . In S2 Fig, these changes in mean and sd are exemplified by varying each parameter about $(k_{on}^{-1}, k_{off}^{-1}, k_1^{-1}, k_2^{-1}, k_3^{-1}) = (1, 1, 1, 0, 0)$ and computing the mean and sd.

Regardless of the dynamics of the post-commit step, the mean and sd of the total transcription interval are less tightly restricted when k_1 is allowed to vary when compared to the case where only k_{on} or k_{off} varies. This results from the fact that by varying k_{on} or k_{off} , the on/off switching can be disabled, but the forward step with rate k_1 remains rate limiting, while for varying k_1 , the entire pre-commit step can be made negligible. E.g., in Eq (1) the mean is limited to $k_2^{-1} + k_3^{-1}$ versus $k_1^{-1} + k_2^{-1} + k_3^{-1}$ (cf. mean and sd < 1 for small k_1^{-1} in S2 Fig).

On the mean transcription interval, the changes in k_{on} and k_{off} have an opposite effect (as the mean is $(1 + k_{on}^{-1}/k_{off}^{-1})k_1^{-1} + k_2^{-1} + k_3^{-1}$). Consequently, relative changes in both k_{on}^{-1} and k_{off}^{-1} induce an equal change in the mean (as quantified by the derivative $k_{on}^{-1}/k_{off}^{-1}k_1^{-1}$), while k_1^{-1} always induces a greater change (derivative of $(1 + k_{on}^{-1}/k_{off}^{-1})k_1^{-1}$); cf. the large mean limits in S2 Fig). Whether the above changes are smaller or larger than those induced by changes in the post-commit steps (k_2^{-1} or k_3^{-1}) is determined by the relative durations of the steps. Meanwhile, the reciprocity of k_{on} and k_{off} does not extend for sd, despite the fact that at the lower bound the models become equivalent. For large k_{on}^{-1} versus k_{off} , the sd always changes more with relative changes in the former. These changes also differ from those induced by k_1^{-1} . For large values, the transcription intervals can change more with either k_{on} or k_1 , depending on the parameters, but will never change most with k_{off} (cf. large sd limits in S2 Fig).

While the mean and sd of the transcription intervals vary monotonically with respect to the changes in the parameters, the noise can change in an intricate manner. This is demonstrated in Fig 4. Specifically, the c_v can be minimized/maximized for specific values of the means and sds of both the pre-commit and post-commit steps.

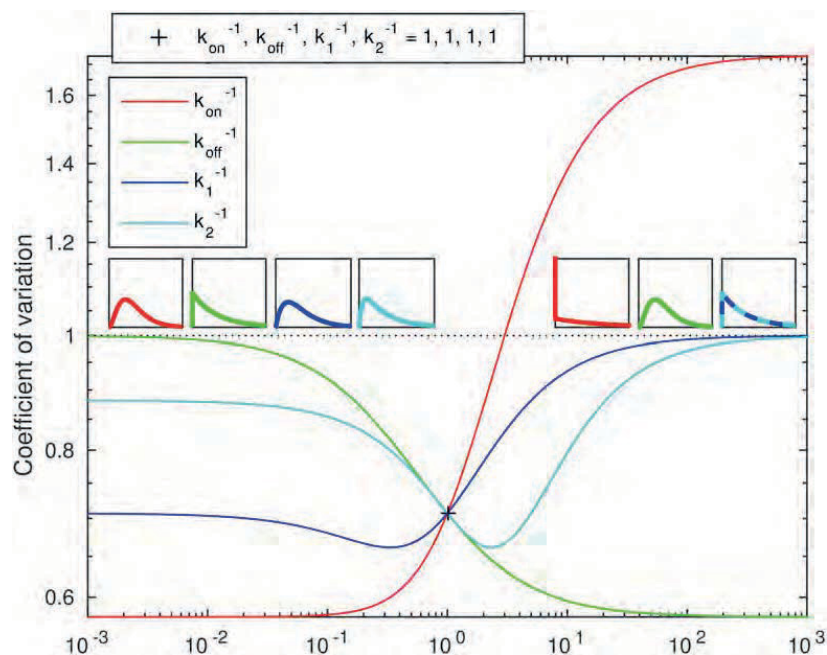


Fig 4. Coefficient of variation (c_v) of transcription intervals as a function of relative changes in the parameters k_{on}^{-1} , k_{off}^{-1} , k_1^{-1} , and k_2^{-1} . The horizontal axis corresponds to the relative value in the specified parameter. The dotted line denotes a c_v of unity. The c_v at the pivot parameter is $1/\sqrt{2}$. Left (right) insets represent the asymptotic transcription interval distributions for small (large) parameter values, their color indicating the varying parameter.

doi:10.1371/journal.pcbi.1005174.g004

In addition to the changes in c_v of the transcription intervals, we also studied the possible shapes of the distributions (cf. the insets of Fig 4) that can be attained by varying each parameter. The shape information is important as, for example, by varying k_{on} , a c_v of unity can be attained, but in this regime the distribution necessarily lacks fast intervals (cf. exponential distribution) and consists of a population of regular intervals mixed with large outliers.

As exemplified in the insets of Fig 4, for varying k_{on}^{-1} , the distribution of transcription intervals ranges from a regular (in other words, hypoexponential, a sequence of exponentials with rates k_1, k_2, k_3) with a low c_v ($c_v < 1$) to a mixed population with fast intervals combined with large outliers (highly noisy, $c_v > 1$). Meanwhile, by varying k_{off} , the distribution ranges from regular to an exponential ($c_v = 1$), as the effects of on/off switching become time-averaged. When varying k_1^{-1} , the distribution ranges from that of the post-commit steps alone to an exponential with time-averaged on/off switching. However, in the latter two cases, it is possible that a highly noisy distribution can be attained for some intermediate parameter values. Alternatively, a noise minimum might exist somewhere in the intermediate region in each of the three cases. When varying k_2^{-1} (or any other post-commit step), the distribution varies from a process lacking that particular step to a single exponential step with that particular rate, and a noise minimum always exists for some intermediate value where the means and sds of the steps are of specific values.

The lower bounds for the c_v of the transcription intervals are limited by the relative durations of the pre- and post-commit stages and number of steps in the post-commit stage. This is exemplified in [S3 Fig](#). Generally, for a post-commit stage consisting of n steps and having a relative duration of r (relative to the mean interval between consecutive transcription events), the lower bound is $(r - 1)^2 + r^2/n$ (cf. green curves in [S3 Fig](#)). Consequently, the lower bound for c_v for any value of r is $1/(n + 1)$ at $r^* = n/(n + 1)$, which occurs when the pre-commit step is exponential-like, and this step and all the n steps of the post-commit stage are of equal duration. As a result, the change in the c_v with changes in the rate limiting step durations is controlled by the number and relative durations of these steps during both post- and pre-commit stages.

Exploring the Model Using the Empirical Data of $P_{lac/ara-1}$ and P_{tetA} Promoters

Finally, we explored, in terms of the model, why the analysis identified some parameters as the most likely candidates for being responsible for the observed changes in RNA production dynamics with temperature and induction scheme, while other parameters were identified as not fit to explain these changes. In other words, we investigate the limitations of each parameter in changing the dynamics of RNA production. In this regard, note that the fitting procedure does not equate the transcription interval mean, standard deviation, nor the coefficient of variation of the model and data in particular, but maximizes the likelihood that the selected model generates the data.

[S4 Fig](#) shows the mean, sd, and c_v of the transcription intervals for the $P_{lac/ara-1}$ as a function of temperature for the Full induction case, while [Fig 5](#) shows the same variables in the relative parameter space about the parameters of the best-fitting model. The additional curves in [Fig 5](#) detail the behavior of the mean, sd, and c_v if other parameters were changing instead. For the $P_{lac/ara-1}$ promoter, the mean, sd, and c_v would not change significantly with changes in k_2 and k_3 . This holds as k_2 and k_3 are relatively small (not rate limiting) and relatively large changes in them barely affect the transcription intervals. Meanwhile, while varying k_{on} could result in the observed changes in the mean and sd, it could not explain the observed changes in the c_v of the data. This is due to the fact that the model features a large k_{on}^{-1} , implying that the noise can only be effectively tuned by varying the burst size k_1^{-1} / k_{off}^{-1} . Consequently, the changes observed in the data must be explained by variations in either k_{off}^{-1} or k_1^{-1} . We verified that these conclusions hold for the other model candidates as well (varying k_1^{-1} , k_{off}^{-1} , k_2^{-1} , and k_3^{-1}), as they could shift the model into different regions of the parameter space. In addition, we found the curves for IPTG to be similar, while for Ara, varying k_2 can modulate mean and sd, but cannot explain the changes in c_v .

Meanwhile, the corresponding curves displaying the changes in mean, sd, and c_v of the transcription intervals for the P_{tetA} promoter, are shown in [S5](#) and [S6 Figs](#). Here, both the mean and sd do not change significantly with changes in k_1 and k_3 , the smaller two of the rate limiting steps. Again, this is due to the fact that these steps are relatively smaller. Interestingly, k_1 and k_3 could be used to manipulate c_v appropriately, which is in contrast with the behavior observed in the $P_{lac/ara-1}$ promoter, and enabled by the more even distribution of the transcription interval durations to the pre- and post-commit stages.

Another reason why changing the identified parameters could be preferred over changing the other feasible parameters is that the two have opposite effects as a function of increasing temperature. Such results are found both in $P_{lac/ara-1}$, where k_{off} and k_1 have opposite effects on both mean and sd of the transcription intervals, and in P_{tetA} , where k_1 and k_2 (or k_3) have opposite effects on the c_v . Alternatively, there could be some physical constraints e.g. on the parameters ranges which are not considered in our models.

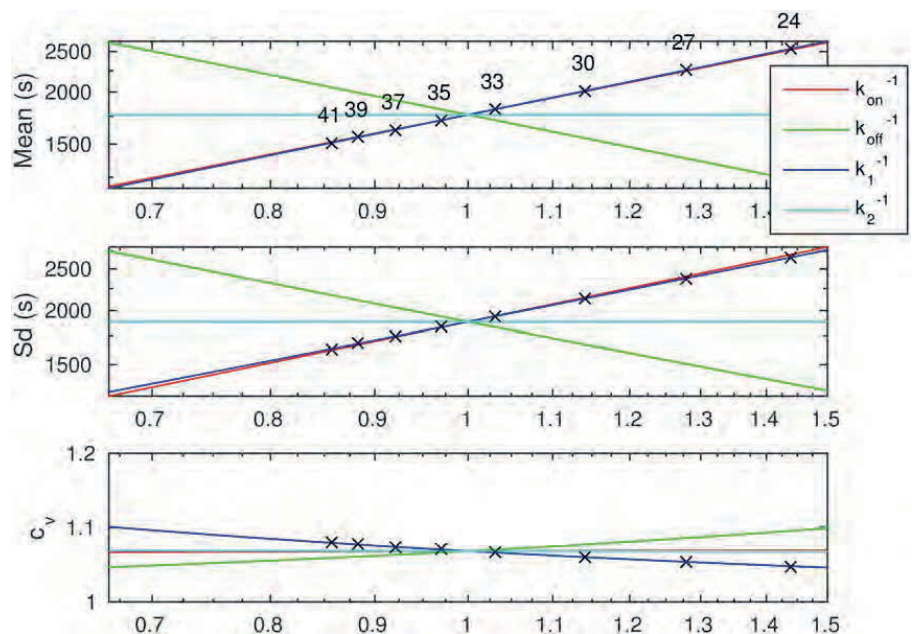


Fig 5. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals when varying each parameter independently around the parameter values found for the $P_{lac/ara-1}$ promoter, full induction. The numeric labels for the crosses indicate temperature ($^{\circ}\text{C}$). The data points are derived from the best fitting model. The curves for k_1^{-1} correspond to those shown in S4 Fig. The curve for k_{on}^{-1} is close to that of k_1^{-1} , k_1^{-1} , and k_2^{-1} for mean, sd, and c_v , respectively. The curve for k_3^{-1} is similar to that of k_2^{-1} in all cases and is thus omitted.

doi:10.1371/journal.pcbi.1005174.g005

Conclusion

We quantified how temperature affects the dynamics (rate, stochasticity, underlying steps, etc.) of the transcription initiation kinetics of the $P_{lac/ara-1}$ promoter in live individual *E. coli* cells. This was performed in three differing induction schemes of IPTG and arabinose, for a wide range of temperatures above and below the optimal growth temperature.

For this, we used statistical methods to compare detailed stochastic models fit to the empirical data in maximum likelihood sense. The selected models inform on the most likely way the changes in RNA production kinetics with temperature and induction scheme emerge from the changes in the rate limiting steps underlying transcription initiation. To overcome the limitations of the presently available methods of observing transcription dynamics in live cells, we performed a differential analysis of several measurements under a diverse set of conditions.

We found that not all steps in transcription initiation are affected equally by temperature changes: varying only some of them suffices to explain the changes found in the measured transcription intervals, regardless of the induction scheme. Specifically, nonlinear changes in the closed complex formation alone suffice to explain the observed changes in transcription dynamics of the $P_{lac/ara-1}$ promoter. By correlating these changes with variations in RNA polymerase numbers, we found that these can be only partly responsible for the observed changes in transcription dynamics, which indicates that temperature affects the interaction between the transcription start site and an RNA polymerase.

Next, we used similar methods to analyze P_{tetA} promoter under two different temperatures. We found that this promoter rate-limiting events occur at different stages of transcription initiation, resulting in a different, less noisy transcription kinetics shape. In agreement, we found that its response to temperature is not explained by modulating the closed complex formation step as for the $P_{lac/ara-1}$ promoter, but instead arises from changes in the open complex formation and/or promoter escape dynamics. Overall, this suggests that the patterns of rate-limiting step kinetics of *E. coli*'s promoters not only cause the genes to differ in RNA production rate and noise, but also in how responsive is their RNA production dynamics to temperature changes.

Finally, we used stochastic models to explore the possibilities of tuning transcription dynamics by varying each of the rate limiting steps. An advantage of modulating the identified steps over the other rate limiting steps was identified: it allows more flexibility in tuning both the mean and variance of the transcription intervals. In addition, in the region of the parameter space suggested by our data, the transcription dynamics is most sensitive to these particular changes, while other means of tuning the dynamics result in opposite changes in the response. This suggest that varying these particular steps might offer the promoters greater adaptability to temperature changes than if the transcription dynamics were tuned by other means.

Our study exemplifies how differential analysis of transcription intervals with statistical methods can inform on the underlying steps of transcription initiation which, at the moment, cannot be directly measured in live cells. We expect these techniques to be applicable, with small modifications only, to study similar processes such as translation or the behavior of genetic circuits.

Supporting Information

S1 Appendix. Extended Methods and Materials.

S1 Table. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals for the $P_{lac/ara-1}$ promoter. The table shows the condition, number of exact intervals (samples), number of right-censored intervals (R-samples), and the estimated mean, sd, and c_v of the intervals (considering exact and right-censored ones).
(CSV)

S2 Table. Presence or absence of on/off switching and number of identified steps in $P_{lac/ara-1}$ promoter. The table shows the condition, the existence or absence (yes/no) of on/off switching, the durations of significant steps (Steps 1 to 3), the second-best fitting model (Alt model), and the difference in BICs (ΔBIC) of the second-best and the best fitting models and its lower bound.
(CSV)

S3 Table. Estimated model parameters for $P_{lac/ara-1}$ promoter. The table shows the induction condition and the estimated parameter values. Here, $k_1^{-1} = AT^2 + BT + C$, where T is the temperature in degrees Celsius.
(CSV)

S4 Table. Statistics of the on/off switching for $P_{lac/ara-1}$ promoter. The table shows the condition, the fraction of time the gene is on (Duty cycle) and the average burst size and interval.
(CSV)

S5 Table. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals for the P_{tetA} promoter. The table shows the temperature condition, number of exact intervals (samples), number of right-censored intervals (R-samples), and the estimated mean, sd, and c_v of the intervals.
(CSV)

S6 Table. Presence or absence of on/off switching and number of identified steps in the P_{tetA} promoter. The table shows the temperature condition, the presence or absence of on/off switching (On/off), the durations of the significant steps (Steps 1 to 3), the second-best fitting model (Alt model), the difference in BICs (ΔBIC) of the second-best and the best fitting model, and its lower bound (LB).
(CSV)

S7 Table. List of promoters and measurement conditions used in the manuscript. The table shows an arbitrary condition ID, the promoter, concentrations of the inducers (IPTG (mM), Ara (L-arabinose; %), and aTc (anhydrotetracycline; ng/mL)), and the temperature ($^{\circ}C$).
(CSV)

S8 Table. List of all intervals between the production of consecutive transcripts used in the manuscript. The table shows an arbitrary interval ID, condition ID (a foreign key to), the lower (Interval LB) and upper bounds (Interval UB), and the ID of the interval that precedes in time (Previous Interval ID) for each interval observed in the cells under both promoters in each condition. The order of the rows has no particular meaning.
(CSV)

S1 Fig. Mean and standard deviation of transcription intervals generated using a Monte Carlo simulation as a function of temperature for the $P_{lac/ara-1}$ promoter. The dashed curves represent the means and standard deviations of the best fitting models.
(EPS)

S2 Fig. Mean and standard deviation (sd) of transcription intervals as a function of the relative change in k_{on}^{-1} , k_{off}^{-1} , and k_1^{-1} . The horizontal axis corresponds to the relative value in the specified parameter. The dashed curves are the asymptotes discussed in the text. The pivot parameters result in a mean of 2 and sd of $\sqrt{6}$.
(EPS)

S3 Fig. Lower bound for the coefficient of variation of transcription intervals as a function of the relative durations between pre- and post-commit stages. Blue points represent instances of random models. Green curves are the c_v lower bounds for different number of steps n in the post-commit stage: from top to bottom: 1, 2 (solid curve; the model in Eq (1)), 3, 4, and 5. The dotted black line is the absolute lower bound attained with a constant-duration post-commit stage ($n = \infty$).
(EPS)

S4 Fig. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals for the $P_{lac/ara-1}$ promoter under full induction, and the corresponding best fitting model. The data points correspond to those shown in S1 Table. Error bars denote one sd of the estimator uncertainty.
(EPS)

S5 Fig. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals for the P_{tetA} promoter under full induction, and the corresponding best fitting model. The data points correspond to those shown in S5 Table. Error bars denote one sd of the estimator uncertainty.
(EPS)

S6 Fig. Mean, standard deviation (sd), and coefficient of variation (c_v) of the transcription intervals when varying each parameter independently around the parameter values found

for the P_{tetA} promoter, full induction. The numeric labels for the crosses indicate temperature (°C). The data points are derived from the best fitting model. (EPS)

Author Contributions

Conceptualization: ASR.

Data curation: ASR AH JLP.

Formal analysis: AH JLP HT.

Funding acquisition: ASR AH SMDO JLP.

Investigation: SMDO VK AH VK.

Methodology: ASR AH SMDO JLP.

Project administration: ASR.

Resources: ASR.

Software: AH JLP.

Supervision: ASR.

Validation: SMDO VK HT JLP.

Visualization: ASR AH.

Writing – original draft: AH ASR SMDO.

Writing – review & editing: AH ASR SMDO.

References

1. Touhami A, Jericho M, Rutenberg AD. Temperature dependence of MinD oscillation in *Escherichia coli*: Running hot and fast. *J Bacteriol*. 2006; 188(21):7661±7667. doi: [10.1128/JB.00911-06](https://doi.org/10.1128/JB.00911-06) PMID: [16936014](https://pubmed.ncbi.nlm.nih.gov/16936014/)
2. Stricker J, Cookson S, Bennett MR, Mather WH, Tsimring LS, Hasty J. A fast, robust and tunable synthetic gene oscillator. *Nature*. 2008; 456(7221):516±519. doi: [10.1038/nature07389](https://doi.org/10.1038/nature07389) PMID: [18971928](https://pubmed.ncbi.nlm.nih.gov/18971928/)
3. Chandraseelan JG, Oliveira SMD, Hakkinen A, Tran H, Potapov I, Sala A, et al. Effects of temperature on the dynamics of the LacI-TetR-CI repressilator. *Mol BioSyst*. 2013; 9(12):3117±3123. doi: [10.1039/c3mb70203k](https://doi.org/10.1039/c3mb70203k) PMID: [24104727](https://pubmed.ncbi.nlm.nih.gov/24104727/)
4. Elowitz MB, Levine AJ, Siggia ED, Swain PS. Stochastic gene expression in a single cell. *Science*. 2002; 297(5584):1183±1186. doi: [10.1126/science.1070919](https://doi.org/10.1126/science.1070919) PMID: [12183631](https://pubmed.ncbi.nlm.nih.gov/12183631/)
5. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A. Regulation of noise in the expression of a single gene. *Nat Genet*. 2002; 31(1):69±73. doi: [10.1038/ng869](https://doi.org/10.1038/ng869) PMID: [11967532](https://pubmed.ncbi.nlm.nih.gov/11967532/)
6. Pedraza JM, van Oudenaarden A. Noise propagation in gene networks. *Science*. 2005; 307(5717):1965±1969. doi: [10.1126/science.1109090](https://doi.org/10.1126/science.1109090) PMID: [15790857](https://pubmed.ncbi.nlm.nih.gov/15790857/)
7. Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. *Cell*. 2005; 123(6):1025±1036. doi: [10.1016/j.cell.2005.09.031](https://doi.org/10.1016/j.cell.2005.09.031) PMID: [16360033](https://pubmed.ncbi.nlm.nih.gov/16360033/)
8. Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching. *Science*. 2008; 322(5900):442±446. doi: [10.1126/science.1161427](https://doi.org/10.1126/science.1161427) PMID: [18927393](https://pubmed.ncbi.nlm.nih.gov/18927393/)
9. Kaern M, Elston TC, Blake WJ, Collins JJ. Stochasticity in gene expression: From theories to phenotypes. *Nat Rev Genet*. 2005; 6(6):451±464. doi: [10.1038/nrg1615](https://doi.org/10.1038/nrg1615) PMID: [15883588](https://pubmed.ncbi.nlm.nih.gov/15883588/)
10. Taniguchi Y, Choi PJ, Li GW, Chen H, Babu M, Hearn J, et al. Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in single cells. *Science*. 2010; 329(5991):533±538. doi: [10.1126/science.1188308](https://doi.org/10.1126/science.1188308) PMID: [20671182](https://pubmed.ncbi.nlm.nih.gov/20671182/)

11. Muthukrishnan AB, Kandhavelu M, Lloyd-Price J, Kudasov F, Chowdhury S, Yli-Harja O, et al. Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucl Acids Res.* 2012; 40(17):8472±8483. doi: [10.1093/nar/gks583](https://doi.org/10.1093/nar/gks583) PMID: [22730294](https://pubmed.ncbi.nlm.nih.gov/22730294/)
12. Yu J, Xiao J, Run X, Lao K, Xie XS. Probing gene expression in live cells, one protein molecule at a time. *Science.* 2006; 311(5767):1600±1603. doi: [10.1126/science.1119623](https://doi.org/10.1126/science.1119623) PMID: [16543458](https://pubmed.ncbi.nlm.nih.gov/16543458/)
13. Chong S, Chen C, Ge H, Xie XS. Mechanism of transcriptional bursting in bacteria. *Cell.* 2014; 158(2):314±326. doi: [10.1016/j.cell.2014.05.038](https://doi.org/10.1016/j.cell.2014.05.038) PMID: [25036631](https://pubmed.ncbi.nlm.nih.gov/25036631/)
14. Ratkowsky DA, Olley J, McMeekin TA, Ball A. Relationship between temperature and growth rate of bacterial cultures. *J Bacteriol.* 1982; 149(1):1±5. PMID: [7054139](https://pubmed.ncbi.nlm.nih.gov/7054139/)
15. Bukau B. Regulation of the *Escherichia coli* heat-shock response. *Mol Microbiol.* 1993; 9(4):671±680. doi: [10.1111/j.1365-2958.1993.tb01727.x](https://doi.org/10.1111/j.1365-2958.1993.tb01727.x) PMID: [7901731](https://pubmed.ncbi.nlm.nih.gov/7901731/)
16. Gruber TM, Gross CA. Multiple sigma subunits and the partitioning of bacterial transcription space. *Annu Rev Microbiol.* 2003; 57(1):441±466. doi: [10.1146/annurev.micro.57.030502.090913](https://doi.org/10.1146/annurev.micro.57.030502.090913) PMID: [14527287](https://pubmed.ncbi.nlm.nih.gov/14527287/)
17. Lopez-Garcia P, Forterre P. DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles. *BioEssays.* 2000; 22(8):738±746. doi: [10.1002/1521-1878\(200008\)22:8<738::AID-BIES7>3.0.CO;2-5](https://doi.org/10.1002/1521-1878(200008)22:8<738::AID-BIES7>3.0.CO;2-5) PMID: [10918304](https://pubmed.ncbi.nlm.nih.gov/10918304/)
18. Bertrand-Burggraf E, Schnarr M, Lefevre JF, Daune M. Effect of superhelicity on the transcription from the tet promoter of pBR322. Abortive initiation and unwinding experiments. *Nucl Acids Res.* 1984; 12(20):7741±7752. PMID: [6387626](https://pubmed.ncbi.nlm.nih.gov/6387626/)
19. Buc H, McClure WR. Kinetics of open complex formation between *Escherichia coli* RNA polymerase and the *lac* UV5 promoter. Evidence for a sequential mechanism involving three steps. *Biochemistry.* 1985; 24(11):2712±2723. doi: [10.1021/bi00332a018](https://doi.org/10.1021/bi00332a018) PMID: [3896304](https://pubmed.ncbi.nlm.nih.gov/3896304/)
20. Roe JH, Burgess RR, Record MT Jr. Temperature dependence of the rate constants of the *Escherichia coli* RNA polymerase-*AP_R* promoter interaction: Assignment of the kinetic steps corresponding to protein conformational change and DNA opening. *J Mol Biol.* 1985; 184(3):441±453. PMID: [3900414](https://pubmed.ncbi.nlm.nih.gov/3900414/)
21. Ryals J, Little R, Bremer H. Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *J Bacteriol.* 1982; 151(2):879±887. PMID: [6178724](https://pubmed.ncbi.nlm.nih.gov/6178724/)
22. Kandhavelu M, Hakkinen A, Yli-Harja O, Ribeiro AS. Single-molecule dynamics of transcription of the *lac* promoter. *Phys Biol.* 2012; 9(2):026004. doi: [10.1088/1478-3975/9/2/026004](https://doi.org/10.1088/1478-3975/9/2/026004) PMID: [22475987](https://pubmed.ncbi.nlm.nih.gov/22475987/)
23. Kandhavelu M, Lloyd-Price J, Gupta A, Muthukrishnan AB, Yli-Harja O, Ribeiro AS. Regulation of mean and noise of the in vivo kinetics of transcription under the control of the *lac/ara-1* promoter. *FEBS Lett.* 2012; 586(21):3870±3875. doi: [10.1016/j.febslet.2012.09.014](https://doi.org/10.1016/j.febslet.2012.09.014) PMID: [23017207](https://pubmed.ncbi.nlm.nih.gov/23017207/)
24. 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(3):203±214. doi: [10.1093/dnares/dsw009](https://doi.org/10.1093/dnares/dsw009) PMID: [27026687](https://pubmed.ncbi.nlm.nih.gov/27026687/)
25. Hakkinen A, Ribeiro AS. Characterizing rate limiting steps in transcription from RNA production times in live cells. *Bioinformatics.* 2016; 32(9):1346±1352. doi: [10.1093/bioinformatics/btv744](https://doi.org/10.1093/bioinformatics/btv744) PMID: [26722120](https://pubmed.ncbi.nlm.nih.gov/26722120/)
26. Peccoud J, Ycart B. Markovian Modelling of Gene Product Synthesis. *Theor Pop Biol.* 1995; 48:222±234.
27. McClure WR. Mechanism and control of transcription initiation in prokaryotes. *Annu Rev Biochem.* 1985; 54:171±204. doi: [10.1146/annurev.bi.54.070185.001131](https://doi.org/10.1146/annurev.bi.54.070185.001131) PMID: [3896120](https://pubmed.ncbi.nlm.nih.gov/3896120/)
28. Saecker RM, Record MT Jr, deHaseth PL. Mechanism of bacterial transcription initiation. *J Mol Biol.* 2011; 412(5):754±771. doi: [10.1016/j.jmb.2011.01.018](https://doi.org/10.1016/j.jmb.2011.01.018) PMID: [21371479](https://pubmed.ncbi.nlm.nih.gov/21371479/)
29. Golding I, Cox EC. Eukaryotic transcription: What does it mean for a gene to be 'on'? *Curr Biol.* 2006; 16(10):R371±R373. doi: [10.1016/j.cub.2006.04.014](https://doi.org/10.1016/j.cub.2006.04.014) PMID: [16713947](https://pubmed.ncbi.nlm.nih.gov/16713947/)
30. 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. *Nucl Acids Res.* 2001; 29(18):3873±3881. doi: [10.1093/nar/29.18.3873](https://doi.org/10.1093/nar/29.18.3873) PMID: [11557820](https://pubmed.ncbi.nlm.nih.gov/11557820/)
31. Kontur WS, Saecker RM, Capp MW, Record MT Jr. Late Steps in the Formation of *E. coli* RNA Polymerase-APR Promoter Open Complexes: Characterization of Conformational Changes by Rapid [Perturbant] Upshift Experiments. *J Mol Biol.* 2008; 376(4):1034±1047. doi: [10.1016/j.jmb.2007.11.064](https://doi.org/10.1016/j.jmb.2007.11.064) PMID: [18191943](https://pubmed.ncbi.nlm.nih.gov/18191943/)
32. McClure WR. Rate-limiting steps in RNA chain initiation. *Proc Natl Acad Sci USA.* 1980; 77(10):5634±5638. doi: [10.1073/pnas.77.10.5634](https://doi.org/10.1073/pnas.77.10.5634) PMID: [6160577](https://pubmed.ncbi.nlm.nih.gov/6160577/)
33. Vogel U, Jensen KF. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J Bacteriol.* 1994; 176(10):2807±2813. PMID: [7514589](https://pubmed.ncbi.nlm.nih.gov/7514589/)

34. Larson MH, Mooney RA, Peters JM, Windgassen T, Nayak D, Gross CA, et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science*. 2014; 344(6187):1042±1047. doi: [10.1126/science.1251871](https://doi.org/10.1126/science.1251871) PMID: [24789973](https://pubmed.ncbi.nlm.nih.gov/24789973/)
35. Hakkinen A, Kandhavelu M, Garasto S, Ribeiro AS. Estimation of fluorescence-tagged RNA numbers from spot intensities. *Bioinformatics*. 2014; 30(8):1146±1153. doi: [10.1093/bioinformatics/btt766](https://doi.org/10.1093/bioinformatics/btt766) PMID: [24389660](https://pubmed.ncbi.nlm.nih.gov/24389660/)
36. Lehmann EL, Casella G. Theory of point estimation. 2nd ed. New York, NY: Springer-Verlag; 1998.
37. Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995; 90(430):773±795. doi: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
38. Tran H, Oliveira S, Goncalves N, Ribeiro A. Kinetics of the cellular intake of a gene expression inducer at high concentrations. *Mol BioSyst*. 2015; 11:2579±2587. doi: [10.1039/c5mb00244c](https://doi.org/10.1039/c5mb00244c) PMID: [26223179](https://pubmed.ncbi.nlm.nih.gov/26223179/)
39. Maeda H, Fujita N, Ishihama A. Competition among seven *Escherichia coli* σ subunits: relative binding affinities to the core RNA polymerase. *Nucl Acids Res*. 2000; 28(18):3497±3503. doi: [10.1093/nar/28.18.3497](https://doi.org/10.1093/nar/28.18.3497) PMID: [10982868](https://pubmed.ncbi.nlm.nih.gov/10982868/)
40. Chamberlin MJ. The selectivity of transcription. *Annu Rev Biochem*. 1974; 43:721±775. doi: [10.1146/annurev.bi.43.070174.003445](https://doi.org/10.1146/annurev.bi.43.070174.003445) PMID: [4605337](https://pubmed.ncbi.nlm.nih.gov/4605337/)
41. So L, Ghosh A, Zong C, Sepulveda LA, Segev R, Golding I. General properties of transcriptional time series in *Escherichia coli*. *Nat Genet*. 2011; 43(6):554±560. doi: [10.1038/ng.821](https://doi.org/10.1038/ng.821) PMID: [21532574](https://pubmed.ncbi.nlm.nih.gov/21532574/)
42. Jones DL, Brewster RC, Phillips R. Promoter architecture dictates cell-to-cell variability in gene expression. *Science*. 2014; 346(6216):1533±1536. doi: [10.1126/science.1255301](https://doi.org/10.1126/science.1255301) PMID: [25525251](https://pubmed.ncbi.nlm.nih.gov/25525251/)
43. Bernstein JA, Khodursky AB, Lin PH, 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 USA*. 2002; 99(15):9697±9702. doi: [10.1073/pnas.112318199](https://doi.org/10.1073/pnas.112318199) PMID: [12119387](https://pubmed.ncbi.nlm.nih.gov/12119387/)
44. Huh D, Paulsson J. Random partitioning of molecules at cell division. *Proc Natl Acad Sci USA*. 2011; 108(36):15004±15009. doi: [10.1073/pnas.1013171108](https://doi.org/10.1073/pnas.1013171108) PMID: [21873252](https://pubmed.ncbi.nlm.nih.gov/21873252/)
45. Raj A, Peskin CS, Tranchina D, Vargas DY, Tyagi S. Stochastic mRNA synthesis in mammalian cells. *PLoS Biol*. 2006; 4(10):e309. doi: [10.1371/journal.pbio.0040309](https://doi.org/10.1371/journal.pbio.0040309) PMID: [17048983](https://pubmed.ncbi.nlm.nih.gov/17048983/)

Temperature-Dependent Model of Multi-Step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements

Samuel M. D. Oliveira[✉], Antti Häkkinen[✉], Jason Lloyd-Price,
Huy Tran, Vinodh Kandavalli, and Andre S. Ribeiro

[✉]These authors contributed equally to this work.

S1 Appendix: Extended Materials and Methods

Cells, Plasmids, Media, and Chemicals

The strain *E. coli* DH5 α -PRO (identical to DH5 α -Z1) [1], generously provided by I. Golding (Baylor College of Medicine, Houston, TX), was used to express the target and reporter plasmids. The strain genotype is *deoR*, *endA1*, *gyrA96*, *hsdR17*(rK⁻ mK⁺), *recA1*, *relA1*, *supE44*, *thi-1*, Δ (*lacZYA-argF*)U169, Φ 80 δ *lacZ* Δ M15, F⁻, λ^- , P_{N25}/*tetR*, P_{lacIq}/*lacI*, SpR. Importantly, this strain contains two constitutively overexpressed genes, *lacI* and *tetR*, under the control of P_{lacIq} and P_{N25} promoters, respectively, ensuring stable tight transcription regulation [1].

Two bacterial systems were used. The first is the mentioned strain containing: i) a medium-copy vector PROTET-K133 with the reporter gene P_{LtetO-1}-MS2d-GFP, which produces the dimeric fusion protein MS2d-GFP; and ii) a single-copy F-plasmid pIG-BAC, carrying the target gene P_{lac/ara-1}-mRFP1-MS2-96bs with a coding region for a monomeric red fluorescent protein (mRFP) followed by an array of 96 MS2 binding sites [2]. The second is a modified version of the original strain with the following differences: i) the low-copy vector pZS12 carries the reporter gene P_{LlacO-1}-MS2-GFP; and ii) the single-copy F-plasmid vector pIG-BAC carries the target gene P_{tetA}-mRFP1-MS2-96bs [3]. The activity of the promoters P_{LtetO-1} and P_{tetA} is regulated by the repressor tetracycline (TetR) and the inducer anhydrotetracycline (aTc). Meanwhile, the activity of the promoter P_{LlacO-1} is regulated by the LacI repressor and the inducer IPTG, and the activity of the promoter P_{lac/ara-1} is regulated by both LacI and AraC repressors and the inducers IPTG and L-arabinose.

All strains were grown in Lysogeny Broth (LB) medium, supplemented with the appropriate antibiotics (35 μ g/mL kanamycin and 34 μ g/mL chloramphenicol for the first strain, and 100 μ g/mL ampicillin and 34 μ g/mL chloramphenicol

for the second strain). Antibiotics were from Sigma-Aldrich (USA). The composition of LB was: 10 g/L of tryptone (Sigma Aldrich, USA), 5 g/L of yeast extract (LabM, UK) and 10 g/L of NaCl (LabM, UK).

Finally, in order to obtain a set of medium conditions where differences between intracellular RNAP concentrations are maximized while differences in growth rates are minimized, we followed the procedure established in [4]. Namely, we carried out measurements in modified LB medium that have lower tryptone and yeast extract concentrations (by 0.25 or 0.5 fold), which reduces intracellular RNAP concentrations accordingly [4].

Induction of Target and Reporter Genes

Cell cultures were diluted in LB from overnight cultures to OD_{600} of 0.05, and kept at 37 °C at 250 RPM in a shaker until reaching mid-logarithmic phase with an OD_{600} of 0.3. After that, cells containing the promoters $P_{lac/ara-1}$ and $P_{LtetO-1}$ were induced with 0.1% L-arabinose and 1 mM IPTG for target activation, and 100 ng/mL aTc for reporter activation. Cells containing the target promoter P_{tetA} and the reporter promoter $P_{LlacO-1}$ were induced with 15 ng/mL aTc for target activation and 1 mM IPTG for reporter activation. In both cases, for the cells to produce sufficient MS2-GFP for the detection of target RNAs, the reporter and target genes were induced 50 minutes prior to the measurements, while keeping cells shaking at 250 RPM in the incubator at the appropriate temperature (24, 27, 30, 33, 35, 37, 39, or 41 °C). In the case of $P_{lac/ara-1}$, the induction was complemented by adding 1 mM IPTG 10 minutes prior to microscopy. In the end, cells were collected by centrifugation at $8000 \times g$ for 1 minute, and diluted in fresh LB medium. For this, 5 μ L of cells were added to an agarose pad (Sigma Life Science, USA), and placed into a temperature chamber (Biopetechs, FCS2) at the appropriate temperature for image acquisition.

Microscopy

The imaging was performed using a Nikon Eclipse (Ti-E, Nikon) inverted microscope, equipped with a 100 \times Apo TIRF (1.49 NA, oil) objective and a C2+ (Nikon) confocal laser-scanning system. Images were captured with the aid of a motorized stage. To visualize fluorescent-tagged RNA spots, we used a 488 nm laser (Melles-Griot) and an emission filter (HQ514/30, Nikon). Time-lapse fluorescence images were taken once per minute for 120 or 60 minutes. The software used for image acquisition was NIS-Elements (Nikon), and the images were analyzed using a custom software, described below.

Image Analysis

The fluorescence microscopy images were processed as follows. First, consecutive images in the time series were aligned such that the cross-correlation of

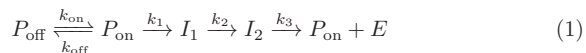
fluorescence intensities is maximized. Next, a region occupied by each cell during the time series is manually annotated. After this, the locations, dimensions, and orientation of each cell in each frame are found by principal component analysis and the assumption that the fluorescence inside the cells is uniform [5]. Cell lineages were constructed using CellAging, which associates segments in consecutive frames based on their overlapping areas [6].

Next, the intensity of each cell is fit with a surface, which is a quadratic polynomial of the distance from the cell border, in least-deviations sense [7]. This surface is taken to represent the cellular background intensity resulting from the abundant, unbound MS2-GFPs, and is subtracted to obtain the foreground intensity. The foreground intensity is fit with a set of Gaussian surfaces, in least-deviations sense, with decreasing heights until the heights are in the 99% confidence interval of the background noise (estimated assuming a normal distribution and using median absolute deviation) [7]. The Gaussians are taken to represent fluorescent RNA spots, the volume under each representing the total spot intensity.

Since the lifetime of a MS2-GFP-tagged RNA is much longer than the cell division time [8, 3], the cellular foreground intensity is expected to be an increasing curve, with a jump corresponding to an appearance of a new tagged RNA. The jump positions are estimated using a specialized curve fitting algorithm [9]. Any observed interval between two consecutive RNA productions is recorded for further analysis as-is, while the interval occurring after the last observed production is rendered right censored [10]. These right censored data improve the accuracy and avoid underestimating the transcription duration [10], as the exactly observed intervals tend to lack more longer intervals than shorter ones.

Modeling Transcription Initiation and Transcription Intervals

We assume the model of transcription initiation specified in Eq (1) of the main manuscript. This model is a submodel of the more general model proposed in [10], which in turn combines the models of [11] and [12]. The reactions are:



where P_{off} , P_{on} , I_1 , and I_2 represent the different states of the promoter, and E represents a produced elongation complex (can be taken to approximate produced RNAs). The mechanistic details are further discussed in the main manuscript. We expect the dynamics of the above reactions to follow the stochastic chemical kinetics [13, 14].

While the phenotypic distribution of a cell population with respect to a particular gene is characterized by the RNA and protein numbers, the contribution of transcription is best characterized by the distribution of produced transcripts in a given period of time, as the former is affected by the latter along with other processes such as degradation and dilution (due to cell division) of the transcripts.

Meanwhile, the distribution of the number of produced transcripts is intimately related to the distribution of consecutive transcription intervals:

$$F_{E(t)}(k) = \mathbb{P}[E(t) \leq k] = \mathbb{P}\left[\sum_{i=1}^{k+1} \tau_i > t\right] = 1 - \left(f_{\tau}^{*k} * F_{\tau}\right)(t) \quad (2)$$

where $F_E(t)$ is the cumulative density of $E(t)$, f_{τ} and F_{τ} are the probability density function and cumulative density of the intervals between the production of consecutive transcripts τ_i (assumed to be independent and identically distributed), $f * g$ is the convolution of f and g , and f^{*k} is the k :th convolution power of f .

For exponential $\tau \sim \mathcal{E}(\lambda)$ with rate λ , the produced RNA numbers are Poisson distributed $E(t) \sim \mathcal{P}(\lambda t)$, in which case it is said that the RNAs are produced according to a Poisson process, or more concisely, that the transcription is Poissonian. Regardless of the interval distribution, the two moments of the two distributions are related in the long-term limit [15]:

$$\begin{aligned} \lim_{t \rightarrow \infty} \frac{\mathbb{E}[E(t)]}{t} &= \mathbb{E}[\tau]^{-1} \\ \lim_{t \rightarrow \infty} \frac{\mathbb{V}\text{ar}[E(t)]}{t} &= \mathbb{V}\text{ar}[\tau] \mathbb{E}[\tau]^{-3} \end{aligned} \quad (3)$$

that is, the mean number of RNA produced per unit time equals the inverse of the transcription interval mean, while the Fano factor of the produced RNA numbers equals the squared coefficient (variance over squared mean) of variation of the transcription intervals. Conveniently, the latter equals always unity for Poissonian process. The long-term limit assumption is necessary such that short-term memory effects (which are present for a non-Poissonian transcription process) of the transcription process vanish.

Note that the long-term limit assumption is only necessary to link the moments of the transcription interval distribution to that of the RNA numbers as specified above, and not in determining the appropriate transcription process from the measurement data—the link in Eq (2) holds for any t .

The transcription intervals of the model of Eq (1) are conveniently written using the following functional equation:

$$f_{\tau} = \left(\sum_{k=0}^{\infty} \frac{k_1}{k_{\text{off}} + k_1} \left(\frac{k_{\text{off}}}{k_{\text{off}} + k_1} \right)^k f_{k_{\text{off}} + k_1}^{*k+1} f_{k_{\text{on}}}^{*k} \right) * f_{k_2} * f_{k_3} \quad (4)$$

where $f_k(t) = k \exp(-kt)$ is the probability density function of the waiting time of a reaction with rate k . The parenthesized expression arises from the random number of visits to P_{off} prior to transitioning to I_1 , and the latter two terms from the remaining reactions. The expression for f_{τ} can be simplified by manipulating it in the Laplace space ($X \mapsto \mathbb{E}[e^{tX}]$) [10]. The result can be

written as:

$$f_\tau(t) = \sum_{j=1}^n \frac{k_{\text{on}} - p_j}{k_{\text{on}}} \left(\prod_{i=1}^n \frac{p_i}{p_i - p_j} \right) p_j \exp(-p_j t) \quad (5)$$

where $p_1, p_2 = \frac{k_{\text{off}} + k_1 + k_{\text{on}}}{2} \pm \frac{\sqrt{(k_{\text{off}} + k_1 - k_{\text{on}})^2 + 4 k_{\text{on}} k_{\text{off}}}}{2}$
 $p_3 = k_2, p_4 = k_3$

provided that $k_{\text{on}}, p_1, \dots, p_n$ are distinct. The singularities $p_i = p_j$ can be removed, and a more general result can be found in our previous article [10]: essentially, in such case, the density is a mixture of Erlang densities instead of the exponential ones as above.

From the above equation, several choices of parameters can result in an identical distribution of transcription intervals (and, consequently, RNA numbers). For example, k_2 and k_3 can be interchanged. This implies that while the best fitting distribution for a set of data can be found, the order of the steps k_2 and k_3 , or in fact, the exact values of any of the parameters, cannot be identified without additional information. In the manuscript, we exploit the details of how the parameters change to provide such additional information to identify $k_{\text{off}}, k_{\text{on}},$ and k_1 .

The other properties of the transcription intervals can be derived from the density, e.g. the mean and variance of the transcription intervals can be integrated from f_τ :

$$\begin{aligned} \mathbb{E}[\tau] &= \left(1 + \frac{k_{\text{off}}}{k_{\text{on}}}\right) k_1^{-1} + k_2^{-1} + k_3^{-1} \\ \text{Var}[\tau] &= \left(\left(1 + \frac{k_{\text{off}}}{k_{\text{on}}}\right)^2 + 2 \frac{k_1}{k_{\text{on}}} \frac{k_{\text{off}}}{k_{\text{on}}} \right) k_1^{-2} + k_2^{-2} + k_3^{-2} \end{aligned} \quad (6)$$

which can alternatively be derived from the results of Peccoud and Ycart [12] using Eq (3), which links the moments of the long-term RNA distribution to those of the transcription intervals.

The duty cycle, average burst size, and average burst interval of the on/off switching loop are $k_{\text{on}} / (k_{\text{on}} + k_{\text{off}})$, k_1 / k_{off} , and $k_{\text{on}}^{-1} + k_{\text{off}}^{-1}$, respectively. Here, the duty cycle is the fraction of time the gene spends in the on versus off state, the burst size is the number of RNAs produced prior to turning off, and the burst interval is the duration between such bursts.

In the temperature-dependent models, each model parameter changes as a function of temperature according to a polynomial function:

$$k_x(T)^{-1} = \sum_{j=0}^p a_{k_x^{-1},j} T^j \quad (7)$$

where $a_{k_x^{-1},j}$ is the order- j coefficient for the parameter k_x and T is temperature. In practice, we consider polynomials up to the second order $p = 2$, as an

order- p polynomial (or higher) can always pass through to $p + 1$ data points. We do not expect the polynomial models to provide particular insight, rather, parameters with a model of orders 0, 1, and 2, indicate that the parameter is independent of, varies linearly, or in a nonlinear fashion, respectively, with temperature.

Model Fitting and Selection

The models are fit using censored time intervals between the production of consecutive transcripts extracted from live-cell measurements (intervals are available in S8 table). The censoring is necessary, as production intervals longer than cell division cannot be observed, resulting in underestimation of the true transcription intervals, and as it improves the accuracy of the parameter estimation by properly accounting for the effects of finite sampling rate (60 s sampling interval) [10]. With censoring, rather than observing the production intervals τ_i , we observe bounds for each interval: $\tau_i \in [x_i, y_i]$.

The models are fit in a maximum likelihood sense. The maximum likelihood estimate is:

$$\begin{aligned}\hat{\boldsymbol{\theta}} &\doteq \arg \max_{\boldsymbol{\theta}} \mathbb{P}[\tau_1 \in [x_1, y_1], \dots, \tau_m \in [x_m, y_m] | \boldsymbol{\theta}] \\ &= \arg \max_{\boldsymbol{\theta}} \underbrace{\sum_{i=1}^m \log(F_{\tau}(y_i | \boldsymbol{\theta}) - F_{\tau}(x_i | \boldsymbol{\theta}))}_{\tilde{\ell}(\boldsymbol{\theta})}\end{aligned}\tag{8}$$

where $\boldsymbol{\theta}$ represents a vector of the parameters to be estimated. If multiple models are to be fit together with independent data sets, the likelihoods sum as for the different samples above. The parameter vector $\boldsymbol{\theta}$ contains the appropriate set of the polynomial coefficients $a_{k_x, j}$ that determine the model parameters for each temperature T . Here, the objective $\tilde{\ell}$ is some function that is equal up to some additive constant to the logarithm of the likelihood function ℓ .

In general, the maximum likelihood objective is not guaranteed to feature attractive properties such as convexity or unimodality, but it is smooth almost everywhere and in practice well behaved, provided that the model is somewhat correct. The optimization was performed using a general-purpose derivative-free nonlinear optimization algorithm [16]. To counter the convergence of the optimization procedure to a local maximum, we used 1,000 random restarts, with each parameter being generated from an unit-interval uniform distribution. The parameters were scaled to have a mean equal to that of the data, assuming that the data were exponential.

The distribution of the estimated parameters or any model feature derived from them can be estimated using the delta method. It can be shown that a mapping applied to the maximum likelihood estimate converges in distribution to that applied to the true parameter such that [17]:

$$\sqrt{m} \left(\mathbf{g}(\hat{\boldsymbol{\theta}}) - \mathbf{g}(\boldsymbol{\theta}) \right) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \mathbf{g}_{\boldsymbol{\theta}}(\boldsymbol{\theta}) \mathcal{I}(\boldsymbol{\theta})^{-1} \mathbf{g}_{\boldsymbol{\theta}}(\boldsymbol{\theta})^T \right)\tag{9}$$

for any $\mathbf{g}(\boldsymbol{\theta})$ that is continuous almost everywhere. Here, $\mathbf{g}_{\boldsymbol{\theta}}$ is the Jacobian of \mathbf{g} , $\mathcal{I}(\boldsymbol{\theta})$ is the Fisher information matrix, and $\cdot \xrightarrow{d} \mathcal{N}(\mathbf{b}, \mathbf{c})$ represents a convergence in distribution to a normal distribution with a mean of \mathbf{b} and covariance \mathbf{c} . For practical purposes, the Jacobian $\mathbf{g}_{\boldsymbol{\theta}}(\boldsymbol{\theta})$ can be approximated with that of at the parameter estimate $\mathbf{g}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$, and the Fisher information at the true parameter $\boldsymbol{\theta}$ can be approximated with the observed information $-\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ at the parameter estimate, where $\ell(\boldsymbol{\theta})$ is the logarithm of the likelihood function and $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}$ its Hessian. In this work, $\mathbf{g}_{\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ is computed using automatic differentiation and $\ell_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}}) = \tilde{\ell}_{\boldsymbol{\theta}\boldsymbol{\theta}}(\hat{\boldsymbol{\theta}})$ is computed numerically.

As the models with more parameters fit never worse than those with less, a scheme to penalize the excess degrees of freedom in the models is required. For this, we use the Bayesian information criterion (BIC) [18]. The BIC is computed according to [18]:

$$\text{BIC} \doteq -2\ell(\hat{\boldsymbol{\theta}}) + k \log(n) \quad (10)$$

where $\ell(\hat{\boldsymbol{\theta}})$ is the log-likelihood at the maximum likelihood estimate, k is the number of parameters, and n is the number of samples. In general, $\ell(\hat{\boldsymbol{\theta}})$ and n are not known when some of the data are censored. Instead, we know $\tilde{\ell}(\hat{\boldsymbol{\theta}})$, the log-likelihood up to some additive constant, and n is known to be in some range, as each censored interval can contain information worth of 0 to 1 samples (the specific value depending on both the sample and the true model, and as such, cannot be determined). However, the additive constant vanishes when comparing two BICs, so the difference of two BICs can be estimated as:

$$\Delta\widehat{\text{BIC}}_{1,2} \doteq \widehat{\text{BIC}}_1 - \widehat{\text{BIC}}_2 = -2 \left(\tilde{\ell}_1(\hat{\boldsymbol{\theta}}_1) - \tilde{\ell}_2(\hat{\boldsymbol{\theta}}_2) \right) + (k_1 - k_2) \log(\hat{n}) \quad (11)$$

where $\hat{n} \doteq 1 n_i + 0.5 n_r$ is an estimate of the effective number of samples. As indicated above, in this work, each of the n_i interval censored samples is assumed to be worth of 1 samples, as the sampling intervals are relatively short compared to the transcription intervals, and each of the n_r right censored sample is assumed to be worth of 0.5 samples.

Finally, a conservative lower bound for $\Delta\widehat{\text{BIC}}_{1,2}$ can be derived:

$$\Delta\text{BIC LB}_{1,2} \doteq \min_{n \in [n_i, n_i + n_r]} -2 \left(\tilde{\ell}_1(\hat{\boldsymbol{\theta}}_1) - \tilde{\ell}_2(\hat{\boldsymbol{\theta}}_2) \right) + (k_1 - k_2) \log(n) \quad (12)$$

which guarantees that invalid conclusions are not drawn due to the inaccuracy of the approximation, and allows a degree of inaccuracy in the former.

qPCR of Target Gene Activity

The activity of the target genes were also analyzed using quantitative PCR (qPCR). Cells containing the target plasmids were grown at various LB media [4] at 37 °C, and induced with their respective inducers (0.1% L-arabinose and 1 mM IPTG for $P_{\text{lac/ara-1}}$ -mRFP1-96BS, and 15 ng/mL aTc for P_{tetA} -mRFP-96BS), as described above. Cells were collected by centrifugation at $8000 \times g$

for 5 minutes. Twice the cell culture volume of RNA protect reagent (Qiagen) were added to the reaction tube, following the addition of Tris EDTA lysozyme buffer (pH 8.0) for enzymatic lysis. The total RNA from cells was isolated by the RNeasy kit (Qiagen), according to the manufacturer instructions. The concentration of RNA was quantified by Nanovue plus spectrophotometer (GE Healthcare). To remove the residual DNA, the samples containing the total isolated RNA samples were treated with DNase. Following that, iSCRIPT reverse transcription super mix was added for cDNA synthesis. Next, the cDNA samples were mixed with the qPCR master mix, containing iQ SYBR Green supermix (Biorad), with specific primers for the target and reference genes. The reaction was carried out in triplicates with a total reaction volume of 20 μ L. For quantifying the target gene, we used mRFP1 primers (forward: 5' TACGACGCCGAGGTCAAG 3' and reverse: 5' TTGTGGGAGGTGATGTCCA 3') and for the reference gene, we used the 16S RNA primers (forward: 5' CGTCAGCTCGT-GTTGTGAA 3' and reverse: 5' GGACCGCTGGCAACAAAG 3'). The qPCR experiments were performed using a MiniOpticon Real time PCR system (Biorad). The following conditions were used during the reaction: 40 cycles at 95 °C for 10 s, 52 °C for 30 s, and 72 °C for 30 s for each cDNA replicate. We used no-RT controls and no-template controls to crosscheck non-specific signals and contamination. PCR efficiencies of the reactions were greater than 95%. The data from CFX Manager TM software was used to calculate the relative gene expression and its standard error [19].

Western Blot for RNA Polymerase Quantification

To quantify RNA polymerase abundance in DH5 α -PRO strain at the different media, we measured the amount of RpoC subunits by western blot. Cells were grown until reaching mid-logarithmic phase, and harvested by centrifugation at 8000 $\times g$ for 1 minute. After that, cell lysate was treated with the B-PER bacterial protein extraction reagent (Thermo scientific), containing protease inhibitors, and incubated at room temperature for 10 minutes. The samples were centrifuged at 15000 $\times g$ for 10 minutes, after which the supernatant was collected. Next, the total protein samples were diluted to the 4 \times lamellae sample loading buffer, containing β -mercaptoethanol, and boiled for 5 minutes at 95 °C. The samples containing about 30 μ g of total protein were resolved by 4 to 20% TGX stain free precast gels (Biorad). Proteins were separated by electrophoresis and then electro-transferred on the PVDF membrane. Membranes were blocked with 5% non-fat milk and incubated with primary RpoC antibodies of 1:2000 dilutions (Biolegend) overnight at 4 °C, followed by the HRP-secondary antibodies 1:5000 dilutions (Sigma Aldrich) for 60 minutes at room temperature. For detection of the RpoC protein, chemiluminescence reagent (Biorad) was used. Images were generated by the Chemidoc XRS system (Biorad). Band quantification was done by using the Image Lab software (version 5.2.1).

References

- [1] Lutz R, Bujard H. Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. Nucl Acids Res. 1997;25(6):1203–1210.
- [2] Golding I, Paulsson J, Zawilski SM, Cox EC. Real-time kinetics of gene activity in individual bacteria. Cell. 2005;123(6):1025–1036.
- [3] Muthukrishnan AB, Kandhavelu M, Lloyd-Price J, Kudasov F, Chowdhury S, Yli-Harja O, et al. Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. Nucl Acids Res. 2012;40(17):8472–8483.
- [4] 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(3):203–214.
- [5] Kandhavelu M, Hakkinen A, Yli-Harja O, Ribeiro AS. Single-molecule dynamics of transcription of the lar promoter. Phys Biol. 2012;9(2):026004.
- [6] Hakkinen A, Muthukrishnan AB, 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(13):1708–1709.
- [7] Hakkinen A, Kandhavelu M, Garasto S, Ribeiro AS. Estimation of fluorescence-tagged RNA numbers from spot intensities. Bioinformatics. 2014;30(8):1146–1153.
- [8] Golding I, Cox EC. RNA dynamics in live *Escherichia coli* cells. Proc Natl Acad Sci USA. 2004;101(31):11310–11315.
- [9] Hakkinen A, Ribeiro AS. Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data. Bioinformatics. 2015;31(1):69–75.
- [10] Hakkinen A, Ribeiro AS. Characterizing rate limiting steps in transcription from RNA production times in live cells. Bioinformatics. 2016;32(9):1346–1352.
- [11] McClure WR. Rate-limiting steps in RNA chain initiation. Proc Natl Acad Sci USA. 1980;77(10):5634–5638.
- [12] Peccoud J, Ycart B. Markovian Modelling of Gene Product Synthesis. Theor Pop Biol. 1995;48:222–234.
- [13] McQuarrie DA. Stochastic approach to chemical kinetics. J Appl Prob. 1967;4(3):413–478.
- [14] Gillespie DT. A rigorous derivation of the chemical master equation. Physica A. 1992;188(1–3):404–425.

- [15] Cox DR. Renewal theory. London, UK: Methuen; 1962.
- [16] Nelder JA, Mead R. A simplex method for function minimization. Comput J. 1965;7(4):308–313.
- [17] Lehmann EL, Casella G. Theory of point estimation. 2nd ed. New York, NY: Springer-Verlag; 1998.
- [18] Kass RE, Raftery AE. Bayes factors. J Am Stat Assoc. 1995;90(430):773–795.
- [19] Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C_T}$ method. Methods. 2001;25(4):402–408.

UNPUBLISHED MANUSCRIPT – STUDY IV

Chromosome and plasmid-borne lacO3O1 promoters differ in sensitivity to critically low temperatures

N.S.M. Goncalves*, S.M.D. Oliveira*, L. Martins, R. Neeli-Venkata, J. Reyelt, J.M. Fonseca, J. Lloyd-Price, H. Kranz, and A.S. Ribeiro

Under Revision. *Equal contributions.

doi: -

Publication reprinted with the permission of the copyright holders.

Chromosome and plasmid-borne *lacO301* promoters differ in sensitivity to critically low temperatures

Nadia S.M. Goncalves^{1,*}, Samuel M.D. Oliveira^{1,*}, Leonardo Martins^{1,2}, Ramakanth Neeli-Venkata¹, Jan Reyelt³, Jose M. Fonseca², Jason Lloyd-Price^{4,5}, Harald Kranz³, and Andre S. Ribeiro^{1,2, +}

¹ Laboratory of Biosystem Dynamics, Multi-Scaled Biodata Analysis and Modelling Research Community and BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, PO Box 553, 33101 Tampere, Finland.

² CA3 CTS/UNINOVA. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal.

³ Gene Bridges, Im Neuenheimer Feld 584, 69120 Heidelberg, Germany.

⁴ Biostatistics Department, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

⁵ Infectious Disease and Microbiome Program, Broad Institute, Cambridge, MA 02142, USA

⁺ Corresponding author. Email: andre.ribeiro@tut.fi, Tel: +358408490736, Fax: +358331154989

Running title: Temperature sensitivity of transcription

Keywords: *In vivo* transcription; single-RNA; chromosome integration; plasmid integration; critically low temperatures.

*Equal contributions

Summary

Bacteria undergo genome-wide expression changes following temperature shifts. Using in vivo single-RNA detection, we study, in *Escherichia coli*, whether these changes differ between chromosome-integrated and single-copy plasmid-borne genes. We show that, for P_{LacO3O1} , while the induction range, fold change, and initial response to temperatures downshifts are similar, at critically low temperatures, the chromosome-integrated promoter kinetics becomes weaker and noisier. Dissection of the initiation kinetics indicates longer duration of the states preceding open complex formation when chromosome-integrated, suggesting enhanced super-coiling. Measurements at 10 °C, 30 °C, and with Gyrase and Topoisomerase I inhibitors, show longer escape times from DNA super-coiling at low temperatures, although Gyrase and Topoisomerase I are active. Similarly, long-lasting buildups are found at 30 °C in energy-depleted cells. We find no long-term consequences of low temperatures, as transcription rates are restored by raising temperature. Simulations of a stochastic, multi-step model of P_{LacO3O1} dynamics with empirical parameter values and a temperature-dependent escape rate from super-coiled states support the conclusions. We conclude that, when chromosome-integrated, P_{LacO3O1} has long-lasting super-coiled states at critically low temperatures that enhance sensitivity to shifts to critically low temperatures. We hypothesize that *E. coli* may be capitalizing this phenomenon to directly encode temperature-sensitive global transcriptional programs.

Introduction

Escherichia coli has evolved sophisticated regulatory programs to adapt to fluctuating environments that involve the precise regulation of the expression of many specific genes (Arsene et al., 2000; Kannan et al., 2008). In general, this regulation occurs at the stage of transcription initiation (Browning and Busby, 2016), and is performed, e.g., by transcription factors (Brewster et al., 2014; McClure, 1985) and σ factors (Cho et al., 2014; Dong and Schellhorn, 2009; Farewell et al., 1998; Rouvière et al., 1995). Chromosomal DNA compaction, caused by supercoiling (Holmes and Cozzarelli, 2000; Stuger et al., 2002) and regulated by nucleoid-associated proteins (NAPs) (Dillon and Dorman, 2010; Postow et al., 2004), is expected to have both local and global effects on gene expression (Dillon and Dorman, 2010; Peter et al., 2004; Pruss and Drlica, 1989). Namely, changing the degree of compaction will affect many genes' expression level, but it does so such that, in some genes expression increases while in others it decreases (Dorman, 2006; Dorman and Dorman, 2016; McClure, 1985; Travers and Muskhelishvili, 2005).

E. coli's genome has approximately 4000 chromosomal genes (Blattner et al., 1997), and extra-chromosomal DNA in plasmids. In general, genes in plasmids are associated to, among other, antibiotic resistance (Davies and Davies, 2010). When beneficial, over time, these tend to be integrated into the chromosome, likely increasing stability and reducing the metabolic burden of plasmid maintenance (Ochman et al., 2000; Rankin et al., 2011).

While the chromosome has topologically constrained segments that allow a buildup of supercoiling (Hardy and Cozzarelli, 2005; Higgins, 2016; Postow et al., 2004; Rovinskiy et al., 2012), plasmids lack discrete constraints, i.e. only have transient ones, e.g., due to transient protein binding (Chong et al., 2014; Leng et al., 2011). Exceptions to this are, e.g., plasmids encoding

membrane associated proteins that, due to anchoring to the membrane (Boeke and Model, 1982; Deng et al., 2004; Pruss and Drlica, 1986; Lynch and Wang, 1993), can form longer lasting constraints, plasmids carrying tandem copies of one or two DNA-binding sites (Fulcrand et al., 2016; Leng et al., 2011), and plasmids carrying the T7 promoter, when expressed in topA mutant strains (Samul and Leng, 2007). These specific topologically constrained segments in plasmids are more ‘efficient’ than transient ones in causing the emergence of negative and positive supercoiling in transcription (respectively, behind and ahead of the RNA polymerase (Deng et al., 2005; Liu and Wang, 1987)).

Further, *in vitro* measurements suggest that, when supercoiling buildups emerge in plasmids due to transient constraints, they diffuse freely in opposite directions and annihilate one another, due to the lack of the segment-based constraints (Chong et al., 2014). This may be one of the reasons for significant differences in the transcription kinetics of genes chromosome-integrated and genes carried in plasmids (Chong et al., 2014). Nevertheless, it is worth noticing that *in vivo* measurements suggest that, prior to annihilation, transient supercoiling changes can influence transcription rates of both plasmid-borne and chromosome-integrated promoters (Moulin et al., 2005; Rahmouni and Wells, 1992; Samul and Leng, 2007).

Temperature shifts affect DNA supercoiling directly (Goldstein and Drlica, 1984; López-García and Forterre, 2000) as well as indirectly, e.g., by affecting the interactivity between NAPs and chromosomal DNA (Amit et al., 2003). It is therefore not surprising that the activity of virtually all chromosomal genes in *E. coli* is affected by temperature downshifts (Jones et al., 1987). Nevertheless, the means by which such global ‘shut-downs’ occur remain unknown.

Another process affecting the kinetics of transcription that is also temperature-dependent is the sequence-dependent process of promoter escape (Kapanidis et al., 2006), the stage at which the RNA polymerase is freed from the promoter and moves downstream towards the elongation region of the DNA template (Hsu, 2002). The stronger is the binding between the RNA polymerase and the promoter, the longer it usually takes for the polymerase to be released from the promoter and begin elongation, due to the multiple failures in escape (Hsu, 2002). Recent evidence suggests that for this escape to be successful, the RNA polymerase needs to pull a sufficient amount of downstream DNA into itself (so as to reach its active center), which involves the breakage of interactions between the RNA polymerase and promoter, and between the RNA polymerase and initiation factors (Kapanidis et al., 2006).

Given the above, we hypothesized that genes carried on plasmids and chromosome-integrated genes can differ in sensitivity to temperature shifts. To test this, we compare quantitatively the effects of temperature shifts on the *in vivo* kinetics of transcription of the P_{LacO3O1} promoter, when on a plasmid and when chromosome-integrated (Experimental Procedures). For this, we use the MS2-GFP RNA tagging technique in *E. coli*, along with a recently proposed methodology to resolve the rate-limiting steps governing the *in vivo* dynamics of initiation of prokaryotic promoters (similar to established steady-state assays to resolve the *in vitro* dynamics) (Lloyd-Price et al., 2016), to characterize, with single-RNA sensitivity, the RNA production dynamics by these constructs at various temperatures, as well in the presence of Gyrase and Topoisomerase I inhibitors and of DNP-based energy depletion. Finally, we make use of stochastic modeling to show that the observed differences in transcription kinetics between chromosome and plasmid

integrated promoters at low temperatures are consistent with currently accepted models of transcription initiation that account for the effects super-coiling build-ups, provided that such low temperatures result in the hindrance to release of DNA super-coiling.

To our knowledge, this study is the first to compare, at the single RNA level, the transcription kinetics of a promoter on a plasmid and in the chromosome as a function of temperature. Further, we study this process at critically low temperatures (below 23 °C), a regime in which most cellular processes exhibit significant differences due to, e.g., globally-altered transcription rates (Jones et al., 1987) and increased cytoplasmic viscosity (Oliveira et al., 2016a).

Results

To study the differences in the dynamics of plasmid and chromosome integrated promoters as a function of temperature, we compare the kinetics of RNA production under the control of the $P_{LacO3O1}$ promoter when integrated into the chromosome with the kinetics of RNA production under the control the same promoter but integrated into a single-copy F-plasmid. We note that, in both cases, the reporter is identical and carried by a multi-copy plasmid (see Experimental Procedures). Also, in all measurements, for each condition, we performed three or more biological repeats. Since in no case did we find statistically significant differences between the repeats, all results are obtained from the cumulative number of cells (composed of all cells from all three replicates).

The $P_{LacO3O1}$ -BS chromosome construct (Figures S1 and S2) is the first engineering of a chromosome-integrated gene coding for an RNA with multiple MS2-GFP binding sites in *E. coli*, which allow a quantification of target RNA molecules from microscopy images (Golding et al., 2005).

It is noted that this method of RNA counting based on MS2-GFP tagging of RNAs (Golding et al., 2005) provides an ‘integer-valued absolute number’ of RNA molecules in a single cell (Golding and Cox, 2004; Kandavalli et al., 2016; Lloyd-Price et al., 2016; Mäkelä et al., 2017; Mäkelä et al., 2011; Muthukrishnan et al., 2014; Oliveira et al., 2016b). The term ‘integer-valued’ is here used due to these numbers being extracted from distributions of spot intensities (Experimental Procedures).

To obtain such numbers, one needs to determine the fluorescence intensity of 1 tagged RNA. This is achieved by determining the fluorescence intensity of each RNA ‘fluorescent spot’ from the images captured by microscopy. This intensity is obtained from the intensity of the smallest spots, as these should correspond to individual RNA molecules. The number of RNAs of ‘larger’ spots is then calculated by dividing their intensity with the intensity of individual RNAs (Golding et al., 2005). This method is described in detail in Experimental Procedures, section ‘RNA counting from fluorescent spots’ (Häkkinen et al., 2013).

To strengthen the argument that the smallest spots correspond to a single RNA, we extracted this intensity from non-induced cells containing a single spot. In addition, we also merged all data from different measurements at different induction levels and again detected the smallest peak of fluorescence intensity arising from spots (as described in Experimental Procedures, section ‘Image Analysis’). This peak intensity was found to be indistinguishable (in a statistical sense) from the one detected using the previously described methodology. Further, to ensure that the MS2-

GFP do not form spots of such intensity in the absence of target RNAs, we assessed if spots of similar intensity existed in cells not carrying the target gene. ‘Fake spots’ were found to occur very rarely (Experimental Procedures, section ‘Control test of the RNA counting method’ and in Supporting Material, Table S1, condition “no target gene”).

It is also of relevance to consider that the target RNA molecules, once tagged by MS2, due to the robustness of the fluorescent properties of the tagging proteins, do not degrade nor lose fluorescence intensity for the duration of measurements reported here (Experimental Procedures) (Golding and Cox, 2004; Tran et al., 2015). We verified these past results, as they were obtained at different temperatures. From the new measurements (Table S2 in Supporting Material), we conclude that, during the measurement period, the fluorescence of tagged RNAs do not decrease significantly over time (gradually or abruptly), in either construct or temperature, in agreement with those previous reports (Golding and Cox, 2004; Tran et al., 2015). In addition, we conclude that tagged RNAs from the chromosome integrated gene and from the plasmid-borne gene have nearly indistinguishable lifetime times from one another, with these lifetimes differing only with temperature. The methodology used to obtain the values in Table S2 are described in Experimental Procedures, section “Half-life of the fluorescence intensity of MS2-GFP tagged RNA molecules”.

In addition, we also verified whether the time for the RNA to become fully tagged did not differ between strains or temperature conditions. We verified that, in all cases, the total fluorescence of individual tagged RNAs was already maximal at the first moment it was detected, meaning that the time for its tagging was always under 1 minute (measurements were taken once per minute). This result is in agreement with previous measurements (Tran et al., 2015).

Finally, in the calculations below we assume that, for both strains (see ‘Experimental Procedures’), only one copy of the target promoter is present in each cell at any given time. This assumption/approximation is based on two facts. First, F-plasmids replicate at the same time (Cooper and Keasling, 1998) or shortly after (Keasling et al., 1991) the chromosome (which also ensures that differences in RNA production between the two strains is not due to differences in copy-numbers). Second, we performed measurements (Experimental Procedures, section “Number of promoter copies during the cell lifetime”) which showed that, 1 hour after the start of the measurements, at 10 °C, only ~15% of the cells had 2 nucleoids, while at 30 °C only ~12% of the cells had 2 nucleoids (in both cases, more than 600 cells were analyzed). Finally, we determined the plasmid copy numbers using RT-qPCR (Experimental Procedures, section ‘Plasmid copy number calculation using RT-qPCR’). These showed that the pBELO absolute copy number equaled 1.00 and 1.02 for the conditions 10°C and 30°C, respectively (Figure S9 and Table S12 in Supporting Material), validating our assumption.

In addition to this, since the two strains are not subject to the same antibiotics (only the cells carrying the target gene in the single-copy plasmid are subject to chloramphenicol), we tested whether the two strains produce non-distinguishable levels (in a statistical sense) of MS2-GFP reporter proteins (as differences in these could result in different ability to count target RNAs). To test this, inducing only the reporter gene, we measured the total background fluorescence intensity of individual cells in both strains (which is almost exclusively due to MS2-GFP reporters) and then compared the two distributions of single-cell background fluorescence intensity by performing a Kolmogorov-Smirnov (KS) test of statistical significance. We found no significant difference between the two distributions. Namely, we obtained a p-value of 0.46, which is much

larger than 0.01 (only when smaller than 0.01 do we reject the null hypothesis that the two sets of data are from the same distribution).

Induction of gene expression is similar in the chromosome and plasmid-integrated constructs

We first verified if this construct is functional and responsive to the inducer, IPTG, and then compared its functioning to the single copy F-plasmid construct (Figures S1 and S2) (Tran et al., 2015). As mentioned in the Experimental Procedures, each of these constructs was inserted in *E. coli* strain BW25993 cells (Datsenko and Wanner, 2000).

For this, first, we verified that there are no significant differences between the growth curves of the strains carrying each of the constructs (Figure S5 in Supporting Material). Next, for each construct and IPTG concentration, we quantified the RNA numbers in live cells at 30 °C by microscopy imaging 1 hour after induction of the target promoter by IPTG (Experimental Procedures), and then image processing cells and RNA-GFP spots within (Experimental Procedures). We find that the mean RNA numbers per cell, relative to maximum induction (1 mM IPTG) exhibit a similar fold change in both constructs (~4 for the plasmid and ~5 for the chromosome construct), as shown in Figure 1, which is specific to the $P_{LacO3O1}$ promoter (see, e.g. (Kandavalli et al., 2016)).

Table S3 shows the P values of KS tests (Massey, 1951) comparing all pairs of distributions of absolute integer-valued RNA numbers per cell obtained in each induction condition, for the plasmid, and for the chromosome construct, when at 30°C. We find that these numbers differ significantly (in a statistical sense) with increasing IPTG concentration in the plasmid construct. In the chromosome construct, these numbers differ in behavior in the absence versus the presence of IPTG, meaning that, while responsive to the inducer, this construct is less sensitive to varying IPTG concentrations. Nevertheless, it is possible to conclude that both constructs are active and responsive to IPTG, while the plasmid construct being more sensitive. Further, for both, these numbers do not increase significantly beyond 500 μ M IPTG, indicating that 1 mM IPTG suffices for full induction, in agreement with previous studies (e.g. (Kandavalli et al., 2016; Lloyd-Price et al., 2016)).

Overall, the results suggest that the same model of transcription (reactions 1-3 in Experimental Procedures) is applicable to both constructs, aside from the values in one or more rate constants (e.g. tuning k_{cc} or k_{unlock} suffices to account for changes in mean RNA numbers per cell as a function of induction).

Transcription by the chromosome-integrated construct is noisier at lower temperatures

We next study whether the chromosome and plasmid constructs are affected differently by temperature changes. We measured the RNA numbers in cells subject to full induction (1 mM of IPTG) by microscopy, for both constructs at various temperatures (30°C, 27°C, 23°C, 20°C, 16°C and 10°C). Cells were induced and kept at the specific temperature for 60 minutes prior to imaging. From the absolute integer-valued RNA numbers in each cell, we calculated the mean and squared coefficient of variation (CV^2) of the RNA numbers in single cells for each condition (Table 1).

Tables S4 and S5 show the P values of KS tests comparing the distributions of RNA numbers in individual cells between conditions for the chromosome and the plasmid (Table S4) constructs, and between these constructs at each temperature (Table S5). This test assesses if the RNA production kinetics of these constructs differ with temperature and from each another, in a statistical sense. For P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

From Table 1 and Table S4, we find that P_{LacO3O1} , when integrated into the single-copy plasmid, is highly responsive to temperature decreases only until 23 °C. Below this, changes in RNA numbers are only significant for temperature shifts wider than those considered in Table S4 (e.g. $p < 0.01$ for 23 °C and 10 °C, not shown in Table S4). Interestingly, this behavior is in line with previous reports for a P_{TetA} promoter and for a Lac/Ara-1 promoter on single-copy plasmids (Oliveira et al., 2016b).

Meanwhile, from Table 1 and Table S4, when chromosome-integrated, P_{LacO3O1} activity decreases significantly for a wider range of temperatures. Differences are detectable between all pairs of ‘neighboring’ conditions, except between 16 °C and 10 °C.

These results are supported by those in Table S5. The P values of the KS tests indicate that, below 23 °C, the plasmid and chromosome constructs differ in all temperature conditions. Meanwhile, for 23 °C and above, they only differ at 30 °C. From this and Table 1, we conclude that the activity of the chromosome-integrated promoter is more heavily reduced as temperature is lowered because it remains sensitive to a wider range of temperature shifts.

To validate these results, we used RT-PCR (Experimental Procedures) to obtain the mean RNA numbers relative to the control (30°C) in cells subject to full induction (1 mM of IPTG) at 23 °C, 16 °C and 10 °C. Results are shown in Figure S6 in Supporting Material. Meanwhile, from Table 1, we calculated the same quantities in the microscopy measurements. We find that both the chromosome and plasmid constructs exhibit the same qualitative behavior as temperature decreases when measured by microscopy and RT-PCR.

Finally, we assessed if the weaker transcriptional activity of the chromosome-integrated promoter at the lowest temperatures could be explained by changes in the known heterogeneities in the spatial distribution of RNAPs (Bratton et al., 2011) or of the nucleoid. For this, we performed measurements of RNAP intracellular spatial distributions as described in Experimental Procedures (section “Tuning intracellular RNAP concentrations and nucleoid staining with DAPI”). However, this hypothesis is rejected by the fact that measurements at 10 °C and 30 °C (Figure S7 in Supporting Material) show no significant differences in the spatial distribution of RNAPs or in nucleoid size.

In this regard, we performed two tests for cells with a chromosome-integrated promoter. First, since the mean RNA numbers per cell in induced cells at 10 °C (Table 1) appears to be smaller than in non-induced cells at 30 °C (Table S1), we tested whether this difference is statistically significant. For that, we performed a KS test between the distributions of RNA numbers in individual cells in the two conditions. We obtained a p-value of 0.99 and, thus, we conclude that the null hypothesis that the two sets of data are from the same distribution cannot be rejected (p-value larger than 0.01). In other words, the RNA numbers in the two conditions do not differ, in a statistical sense. As such, induced cells at 10 °C produce at least as much RNAs as non-

induced cells at 30 °C. Second, we tested whether the induction mechanism is functional at 10 °C, i.e. whether the differences in produced RNA numbers are significant (in a statistical sense) between induced and non-induced cells. For that, we performed a KS test between the distributions of RNA numbers in individual cells in the two conditions (Figure 2 and Table S1). We found the p-value to be much smaller than 0.01, from which we conclude that the two distributions can be distinguished, in a statistical sense. Thus, there is tangible induction at 10 °C.

Mean relative time prior to the commitment to transcription increases in the chromosome-integrated construct at low temperature

To investigate the cause for the different behavior of the two constructs at the lower temperatures, we assessed whether the changes in kinetics with decreasing temperature occur *prior to* or *following* commitment to open complex formation. For this, we estimated the mean fraction of time between consecutive transcription events taken by the steps preceding (t_{prior}) and following (t_{after}) the commitment to open complex formation (Kandavalli et al., 2016; Lloyd-Price et al., 2016) at the highest and lowest temperature conditions (Experimental Procedures, section ‘Relative mean duration prior to and following commitment to transcription’).

In short, this method established in (Lloyd-Price et al., 2016) (also used in (Kandavalli et al., 2016)), similar to established steady-state assays to investigate the *in vitro* dynamics of transcription in bacteria (McClure, 1985), is based on the assumption that, increasing the concentration of active RNAP molecules in the cell, one should increase the rate of RNA production. Importantly, this increase in RNA production rate occurs due to the increased rate of the steps prior to the start of the open complex formation (which depend on RNAP numbers, see reaction 1 in Experimental Procedures), while the rate of the steps following the start of open complex formation is not altered. Given this, one can assume that, for infinite RNAP numbers, the rate-limiting step of transcription are solely the steps after the open complex formation.

To estimate the duration of these steps, one needs to plot the inverse of the RNA production rate against the inverse of the RNAP concentrations, for various conditions differing in the concentration of RNAP in the cells, thus obtaining a Lineweaver–Burk plot, and then fit a line to obtain the estimated rate of RNA production for infinite concentration of RNAP. This method is valid if the increase in the rate of RNA production is linear with the increase in RNAP concentration, within the range of conditions used to obtain the data points (Lloyd-Price et al., 2016).

Importantly, in (Lloyd-Price et al., 2016), it was found that, for a certain range of media richness conditions (here obtained by differing Glycerol concentrations), the cell growth rates can remain unaltered, while the RNAP numbers increase as the richness increases (resulting in higher RNAP intracellular concentrations), which is the condition needed to use a Lineweaver–Burk plot.

Thus, first, we tested if the differences in Glycerol concentration of the M9 medium used to alter intracellular RNAP concentrations (Experimental Procedures), do not affect the cell growth rate. Cells were grown in media with 0.2, 0.4, 0.6 and 0.8% of Glycerol, denoted as 0.5X, 1X, 1.5X and 2X, respectively. For each of these media conditions, growth rates were assessed from the OD₆₀₀ over time by a spectrophotometer (Muthukrishnan et al., 2014). For cells with either construct, we found no significant differences in growth rates with changes in media (Figure S8 in Supporting Material).

In addition, to show that these changes in the media conditions also do not generate differences in the cells' DNA spatial organization, we obtained the DAPI signal in each cell for each of the 0.5X, 1X, 1.5X and 2X conditions, for cells at 30 °C and 10 °C. Then we tested (KS-tests), whether the single-cell distributions of these fluorescence levels differed between conditions. Results in Table S6 show that the distributions cannot be distinguished, in a statistical sense (except when comparing, at 10 °C, the 0.5X and the 2.0X conditions directly).

Next, for each condition, we obtained the RNAP levels relative to a 'control' condition (Experimental Procedures). This is obtained by, first, subtracting the mean absolute fluorescence values measured in the 0.5X condition from the values obtained in each of the other conditions. Afterward, we obtained the ratio between each of these differences (namely, for the 1.5X and the 2.0X conditions) and the difference in the 1X (control) condition. Finally, we obtained the inverse of these two values. These calculations, whose results are shown in Table S7 along with the KS-tests shown in Table S8, demonstrate that the RNAP concentrations differ in cells in different media, as expected. Interestingly, the degree of this change with media composition appears to be temperature dependent.

Subsequently, we measured the RNA production rates of both constructs by RT-PCR for each media condition (1.5X and 2X) relative to the 1X condition. Results in Table S9 show significant differences with media, temperature, and construct, as expected from the microscopy measurements and the present and previous data on the effects of changing media richness on *in vivo* RNA production rates (Kandavalli et al., 2016; Lloyd-Price et al., 2016).

Relevantly, it is the relative free RNAP concentration in the cells that determines the transcription rate. Here, we measure RNAP concentrations from the RNAP-GFP signal intensity (Experimental Procedures). Meanwhile, previous studies have shown that, in some cases, RNAP foci (here RNAP-GFP), may not be adequate in reflecting the ability of RNAP to synthesize mRNA (Cabrera and Jin, 2006).

To determine whether, in the conditions studied here, the relative differences in RNAP-GFP concentration between conditions (measured here and shown in Supporting Table S7) are a good proxy for the relative differences in free RNAP concentrations and, thus, can be used to assess the ability of RNAP to synthesize mRNA in the cells, we need to test if, for the range of media richness used, the inverse of the RNA production rates (Supporting Table S9) change linearly with the inverse of the RNAP concentrations (Kandavalli et al., 2016; Lloyd-Price et al., 2016). I.e., a Lineweaver–Burk plot (Lineweaver and Burk, 1934) should exhibit a line, since this is evidence that, for the conditions tested, both the relative free RNAP concentrations can be approximately assessed from the total RNAP concentrations, as well as that no factors other than the changes in the free RNAP concentration affect the promoter of interest.

Lineweaver–Burk plots are shown for the plasmid and chromosome constructs in Figure S10 in Supporting Material. In each, the above-mentioned linear relationship was tested by a likelihood ratio test, to determine whether the small deviations from linearity are statistically significant (here we used a weighted total least-squares algorithm for fitting a straight line, WTLS (Krystek and Anton, 2007)).

From Figure S10, the linear relationship can be observed for both chromosome and plasmid constructs. In agreement, the results of the linearity tests show that, within this range of RNAP con-

centrations, the linear model cannot be rejected in either case ($P > 0.25$). Overall, we find that for the conditions used here, the relative free RNAP concentrations in the cells are well-approximated by the total RNAP concentrations, and there are no significant other factors affecting the initiation dynamics of $P_{lacO3O1}$.

Finally, from the values in Table S9, along with the RNAP relative concentrations shown in Table S7, we estimated the mean time-length (relative to the mean duration of the intervals between transcription events) of transcription initiation prior ($\frac{t_{prior}}{\Delta t}$) and after ($\frac{t_{after}}{\Delta t}$) the initiation of the open complex formation, for each construct and temperature condition. The methodology to perform these estimations was first described in (Lloyd-Price et al., 2016) and later used in (Kandavalli et al., 2016) and (Mäkelä et al., 2017) (see Experimental Procedures, section ‘Relative mean duration prior to and following commitment to transcription’ for a detailed description of this methodology). Results in Table 2 show that, in all 4 conditions, the most rate-limiting events occur *after* commitment to open complex formation.

From Table 2, in the plasmid construct, in agreement with previous *in vitro* measurements for the synthetic Lac-UV5 promoter, the reduction in RNA production rate with lowering temperature is mostly due to a reduction in the rate of the events after the commitment to open complex formation. Meanwhile, in the chromosome construct, we observe the opposite. I.e., at 10 °C, the mean fraction of time spent *prior* to commitment to open complex formation increased significantly. This suggests that the events whose rates were most reduced occur prior to commitment to open complex formation, provided that the RNA production rate decreases with lowering temperature (as is the case, given the data in Table 1).

Local DNA supercoiling in the chromosome-integrated gene drives the differences between constructs

We observed an increased time-length of the events *preceding* the open complex formation in the chromosome-integrated construct at lower temperatures. Our model of transcription initiation (Experimental Procedures, reactions 1-3) allows for this to occur, provided that decreasing temperature leads to significant changes in specific rate constants. In particular, it can occur when decreasing temperature causes a decrease in the rate of unlocking (k_{unlock}) of the promoter from the locked state (reaction 3 in Experimental Procedures) or in the rate of unbinding of a repressor from the promoter (k_{ON} , reaction 2 in Experimental Procedures), or both.

A third reasonable possibility is that decreasing temperature leads to a significant modification in the kinetics of closed complex assembly-disassembly that results in an increased relative duration of the closed complex formation, e.g. due to reduced k_1 or k_2 , or instead increased k_{-1} . However, this would result in reduced noise in RNA production (Lloyd-Price et al., 2016; Mannerström et al., 2011), and thus reduced CV^2 in RNA numbers in individual cells (because, at 30 °C, most time between transcription events is spent in open complex formation, as shown in Table 2). The data on CV^2 in RNA numbers in Table 1 disproves this hypothesis.

To increase the relative duration of the events preceding the open complex by tuning the physics of the locking/unlocking process (due to supercoiling), the unlocking step must last longer on average, equivalent to reduced k_{unlock} in the model. This modification results in increased noise in RNA production (Golding et al., 2005), increasing the CV^2 in RNA numbers in individual cells,

which was observed in Table 1. Since a decrease of k_{ON} with decreasing temperature would have a similar effect, the CV^2 in RNA numbers cannot distinguish between these possibilities. In any case, both of them result in increased ‘escape time’, from two different ‘forms’ of ‘OFF’ states, during which the promoter is not available to RNA polymerases.

The increase in ‘escape time’ at critically low temperatures could be due to several effects, including the efficiency of repressors (see e.g. (Oliveira et al., 2015)), DNA packaging (known to differ between plasmid and chromosomes (Higgins and Vologodskii, 2015), or DNA super-coiling (López-García and Forterre, 2000) (known to affect both packaging (Stuger et al., 2002) and transcription (Blot et al., 2006; Chong et al., 2014; Fulcrand et al., 2016; Marr et al., 2008))). However, the change in the efficiency of the repressors should affect both the chromosome and plasmid constructs similarly, while we observe divergent behaviors between these two constructs at the lowest temperatures, with the plasmid-borne one being unable to turn ‘OFF’ its RNA production as efficiently as the chromosome-integrated construct (Table 1). Because of this, we conclude that the stronger decrease in the chromosome construct in RNA production rate with lowering temperature (at the lowest temperature conditions tested) should be due to an increased amount of time required to remove the promoter from the ‘Locked’ state, which does not occur in the plasmid construct.

Finally, to determine if the divergence in behavior between the two constructs is mostly associated with DNA packaging and/or super-coiling, we measured the nucleoid sizes in cells with one nucleoid (Experimental Procedures). If changing temperature (within the ranges in Table 1) affects DNA packaging significantly, we expect differences in the mean or variability of the size of the nucleoids. However, we found no significant differences between 10 °C and 30 °C (Table S10). Similar results were reported in (Oliveira et al., 2016a). Thus, we discard effects on DNA packaging as the main cause of the differences between chromosome and plasmid responses to the temperature shifts.

Given all of the above, we hypothesize that the divergence in behaviors of chromosome and plasmid-integrated gene with lowering temperature is due to an increased rate of accumulation of local DNA super-coiling in the chromosome-integrated gene, which enhances the strength of Locked states, resulting in increased escape times (reactions 1 and 3 in Experimental Procedures).

To test this, we performed several experiments. First, in cells with the chromosome construct and in cells with the plasmid construct, we measured at 10 °C, from the moment when the target gene was activated, the RNA numbers in individual cells every 15 minutes for 90 minutes (for each time point, new cells were taken from the original culture). If the weaker activity of the chromosome-integrated promoter is due to an increased propensity to be in the Locked state due to the accumulation of DNA super-coiling, we expect these cells to have their transcription activity blocked after a few events. At a population level, this will result in a sharp decrease in the mean rate of RNA production, sometime after the start of the measurements. Meanwhile, in cells with the plasmid construct, we expect a constant RNA production rate over time, due to the lack of accumulation of local DNA super-coiling (Chong et al., 2014).

Results in Figure 2A confirm these predictions. Cells at 10 °C with the chromosome construct only exhibit production in the first 30 minutes. Cells with the plasmid construct exhibit an approximately constant RNA production rate throughout the measurement period.

We also performed measurements in cells with the plasmid construct and with the chromosome construct at 30 °C. Given the similar behaviors previously observed in this condition (Table 2), we expect approximately constant RNA production rates in both constructs. Results in Figure 2B confirm this.

Next, we compared the two constructs at 30 °C, when subjecting cells to Novobiocin, a repressor of Gyrase activity (Experimental Procedures) (Chong et al., 2014; Gellert et al., 1976). Gyrase releases positive supercoiling (Drlica, 1992), but not negative supercoiling (Wang, 1996). According to the twin-supercoiled-domain model (Liu and Wang, 1987), which predicts that negative/positive supercoils should accumulate in the absence of supercoil-relaxing enzymes, we expect cells with the chromosome construct to exhibit a similar behavior as when at 10 °C. Meanwhile, cells with the plasmid construct should exhibit a constant rate of transcription over time (Ching et al., 2014). Figure 3A confirms these predictions.

We note that, in both strains, the gene *acrA* is present, and thus, Novobiocin is not expected to affect cell division rates (Ma et al., 1995). To test this, we measured cell growth rates by OD₆₀₀ for varying Novobiocin concentrations (0, 50, 75, 100 and 150 ng µl⁻¹). We found the growth rates not to differ significantly between the conditions (data not shown). Note that it is also possible to state that 100 ng µl⁻¹ Novobiocin concentration suffices to affect (but not halt) the transcription rate of the chromosome-integrated construct (compare the results for this construct in Figures 2B and 3A, at 30 °C).

Subsequently, we observed cells carrying the chromosome and construct when subject to Novobiocin and at 10 °C. Results are shown in Figure 3B. Comparing to the previous results (Figures 2A and 3A), we find that transcription in cells carrying the chromosome is more strongly blocked when combining the two factors. I.e. while at 10 °C alone and subject to Novobiocin alone, RNA numbers increase by a factor of 4 (from 0.25 to 1) in a period of 90 min., when subjecting cells to 10 °C and to Novobiocin, the RNA numbers increase only by a factor of 2 (from 0.5 to 1) in the same period of time. Meanwhile, regarding cells carrying the plasmid, we observed exactly the same RNA production as in Figure 2A, meaning that, for these cells, Novobiocin has no effect at 10 °C, confirming the results for the same cells at 30 °C (Figure 3A).

Given this, we suggest that the transcription activity of the chromosome-integrated gene at 10 °C is hampered by an increased difficulty in executing the unblocking of the DNA from supercoiled states that is not directly caused by a loss of functionality of Gyrases.

Next, we observed the dynamics of transcription of cells with the chromosome-integrated gene at 30 °C subject to Topotecan, a repressor of Topoisomerase I activity (Chen and Liu, 1994; Patel et al., 1998) (Experimental Procedures). Topoisomerase I releases negative, but not positive supercoiling (Wang, 1996). Again, we expect these cells to exhibit a similar behavior as when at 10 °C, which Figure 3C confirms.

In addition, we observed cells with chromosome integrated gene subject to Topotecan when at 10 °C. From Figure 3D, transcription is again blocked more strongly than in cells at 10 °C not subject to Topotecan and then in cells at 30 °C subject to Topotecan. Namely, while in the latter two conditions, RNA numbers increased by a factor of 4 (from 0.25 to 1) in a period of 90 min., when subjecting cells to 10 °C and to Topotecan the RNA numbers increase only by a factor of 2 (from 0.5 to 1) in the same period of time.

These results again suggest that the activity of the chromosome-integrated gene at 10 °C is being hampered by an increased difficulty in unblocking the DNA from supercoiled states, rather than directly due to a loss of functionality of Topoisomerases I. Finally, from Supporting Table S15, showing the results of the KS tests between the distributions of RNA numbers in individual cells at 10 °C and 30 °C, when subject to Novobiocin or Topotecan, confirm that the distributions differ with temperature, in a statistical sense.

Finally, as seen in Figures 3A-3D, in cells carrying the single-copy F-plasmid, the presence of Topotecan causes the same qualitative behavior as the presence of Novobiocin (at 30 °C and at 10 °C).

Promoter release from supercoiling buildup is similarly hampered if cellular energy is depleted

If the release from DNA supercoiling buildups in the chromosome-integrated construct at low temperatures is an ‘energy deficiency’ problem due to the low temperatures (namely, the energy required for the necessary endothermic reactions to occur is expected to be much higher in such conditions), it should be possible to mimic the observed phenomena by, instead of lowering temperature, depleting cells of their energy via DNP treatment (Parry et al., 2014) (Experimental Procedures). In particular, we expect cells subject to this, even at 30 °C, to become less able to maintain the chromosome integrated promoter active over time, similar to when at 10 °C.

We tested this by subjecting cells to DNP for 90 minutes prior to imaging (Experimental Procedures). As expected, we observe a similar RNA production dynamics (Figure 4) as in untreated cells at 10 °C with a chromosome-integrated $P_{LacO3O1}$ (Figure 2A). Namely, beyond 30 minutes, there is little to no transcription. This similarity further confirms that the activity of the chromosome-integrated promoter at 10 °C is being hampered by an increased difficulty in unblocking the DNA from supercoiled states.

Low temperatures have no long-term consequences on transcription blocking by DNA supercoiling

We expect that, if the release from DNA supercoiling buildup in the chromosome construct at low temperatures is an ‘energy deficiency’ problem due to the low temperatures, changing temperature again to near-optimal conditions (e.g. 30 °C) should restore the cells’ ability to relax DNA supercoiling. We thus tested, for cells with a chromosome-integrated $P_{LacO3O1}$, whether shifting temperature, first from ‘high’ (30 °C) to ‘low’ (10 °C) and then from ‘low’ (10 °C) to ‘high’ (30 °C), results in smooth transitions in the release of blocked promoters and corresponding changes in RNA production rates.

Results in Figure 5 on the relative mean RNA numbers in cells over time show that this is the case. Clearly, while the cells are at 30 °C there is constant RNA production. Once temperature is shifted to 10 °C, after 15-30 minutes, little to no RNA production is observed (similar to the results in Figure 2A). More importantly, as soon as high temperatures are restored (to 30 °C), RNA production is restored to nearly the original rate.

This result reinforces the conclusion that the activity of the chromosome integrated promoter at 10 °C is hampered by an increased difficulty in unblocking the promoter from supercoiled states, and that, once temperature is restored to near-optimal numbers, so is the RNA production rate.

Note that, following the shift from 30 °C to 10 °C, it does not follow a transient of ~15-30 minutes of reduced transcription activity that is visible in Figures 2, 3, and 4. This is because here, when the shift occurs, the cells already contain sufficient IPTG for full transcription while, in the other experiments, the inducer was added immediately before the microscopy measurements began, and thus, a transient time to reach ‘quasi-equilibrium’ RNA production rates is expected, due to the non-negligible time to intake inducers from the media (Kapanidis et al., 2006) particularly at low temperatures. In the case of IPTG, this transient is expected to be ~15-30 minutes long (Tran et al., 2015), in agreement with the results in Figures 2A, 3A-D, and 4.

Stochastic modeling also suggests increased long-lasting super-coiled states at critically low temperatures to be the cause for enhanced sensitivity to shifts to critically low temperatures

We tested whether the increase in the expected time to release promoters from a supercoiling state across the cell population is, in accordance with current stochastic models of transcription (Kandavalli et al., 2016; Lloyd-Price et al., 2016), a plausible explanation for the change with decreasing temperature in the measured average RNA numbers over time for cells with the chromosome integrated promoter (Figure 2). For this, we use the stochastic model of transcription initiation described in Experimental Procedures (reactions 1-3), derived from multiple studies, including genome-wide studies of variability in transcript counts (Bernstein et al., 2002; Taniguchi et al., 2010) and studies of the transcription dynamics of individual genes (Kandavalli et al., 2016; Lloyd-Price et al., 2016).

All parameter values (Table S11, in Supporting Material) are from single-cell, single-RNA empirical data on the activity of Lac derivative promoters (Chong et al., 2014; Lloyd-Price et al., 2016). Mean RNA polymerase numbers are set so as to correspond to the intracellular RNAP concentration reported in (Lloyd-Price et al., 2016). Finally, based on the results above, we assume that the rate constant most responsible for the increase in $\frac{t_{prior}}{\Delta t}$ as temperature is lowered (Table 2) is a decrease in k_{unlock} . The remaining rate constants are, for simplicity, left unchanged.

For each value of k_{unlock} tested, we performed 500 independent 75 minutes long simulations, every 15 minutes as in the experiments (Figure 2). The values of k_{unlock} were selected as follows: the highest value, corresponding to high temperatures (30 °C), is reported in (Chong et al., 2014). This value was then gradually lowered until the mean number of RNAs per cell at the end of the measurement period was similar to that observed in cells at 10 °C. We then assessed if the model was able to reproduce the observed RNA numbers over time at both high and low temperatures, and if there is a gradual behavioral change between these extreme conditions.

Initial numbers of all molecular species were set to zero, with the exception of P_{ON} (set to 1, corresponding to one active promoter per cell), RNAP (as noted above), and RNA. Initial RNA numbers were drawn randomly from a Poisson distribution with rate 0.7 RNA/cell to match observed spurious RNA production events. We observed that this number did not differ with temperature, as expected, since prior to moment ‘0’ cells were kept at the same temperature (30 °C) in both cases.

In Figure 6, we compare the results of the model with those in Figures 2A (10 °C) and 2B (30 °C) for the chromosome-integrated promoter. For simplicity, as noted above, we ignore the first

time moment of the empirical data (0 minutes following induction) as, at this stage, the cells do not yet have fully active transcription (Tran et al., 2015). This removes the need to model the intake process for the inducers (for an example of how to model this process see (Tran et al., 2015)).

Results in Figure 6 support the earlier conclusions. The accuracy with which the model reproduces the results from the measurements suggests that the difference in mean RNA production rates over time between cells with the chromosome-integrated promoter at critically low (10 °C) and at high (30 °C) temperatures can be explained by a reduced ability to release chromosome-integrated promoters from the effects of DNA super-coiling when at critically low temperatures.

Finally, we note that, setting k_{unlock} to ‘infinite’ in reaction 3 (which is equivalent to having a model that does not allow the promoter to become locked) results in a similar behavior to that observed for the plasmid-borne construct, and, therefore, to the chromosome integrated promoter at 30 °C (data not shown).

Discussion

Temperature-driven changes in genomic DNA supercoiling are expected to be one of *E. coli*’s mechanisms of sensing and responding to temperature shifts (Los, 2004). A previous study (Chong et al., 2014) using detailed single-cell, single-RNA detection methods found that, at optimal temperatures, DNA supercoiling buildup on DNA segments by transcription eventually stops transcription initiation, which can be resumed upon release of the supercoiling by Gyrase. This buildup only has significant effects in chromosomal genes, since plasmids lack discrete topological constraints, allowing the negative and positive supercoiling emerging in transcription to diffuse freely in opposite directions, until nullifying one another (Chong et al., 2014).

We studied this phenomenon in the P_{LacO3O1} promoter, under full induction, as a function of temperature (this promoter, due to the lacking of the O2 site and due to being under full induction is not expected to form significant discrete topological constraints (Fulcrand et al., 2016)). We showed that, depending on whether it is chromosome-integrated or carried on a plasmid, there will be major differences in its response to temperature downshifts. Namely, at lower temperatures, similar to when Gyrase or Topoisomerase I are inhibited, the release of transcription due to the buildup of supercoiling is hampered in the chromosome-integrated construct. Consequently, P_{LacO3O1} suffers a transcriptional shutdown at critically low temperatures, which does not occur when it is plasmid-borne. Overall, these results show that promoters’ in *E. coli* can differ in their degree of sensitivity to shifts to critically low temperatures, depending on whether they are chromosome-integrated and when they are plasmid-borne.

Our conclusion is supported by two independent sets of experiments. First, we dissected the kinetics of transcription of P_{LacO3O1} by measuring RNA production kinetics in cells at 10 °C and 30 °C with various RNA polymerase concentrations. These showed that only the chromosome-integrated P_{LacO3O1} exhibits a wide reduction in the speed of transcription at the stages prior to initiation of the open complex formation. More precisely, we showed that it remained more often in the locked state, as expected if transcription is blocked. Secondly, we conducted temporal measurements of RNA numbers in individual cells at 10 °C and 30 °C in the presence and ab-

sence of the Gyrase or Topoisomerase I inhibitors, Novobiocin and Topotecan. The results from these measurements also supported our hypothesis.

Simulations of a stochastic model with realistic parameter values for $P_{LacO3O1}$ further supported our explanation for the observed behavioral changes with temperature in the chromosome-integrated construct and for the behavioral differences between the chromosome and plasmid-borne constructs at low temperatures. Not only was the model capable of matching the transcript numbers of cell populations over time, at both normal and low temperatures, but this was achieved by tuning a single parameter: the expected time to unlock a promoter locked due to super-coiling.

Importantly, we found evidence that the blocking of the release of $P_{LacO3O1}$ from supercoiled buildup states at low temperatures is not due to the significant loss of functionality of Gyrases or Topoisomerases I. Specifically, we observed that blocking the activity of either of these proteins in cells at 10 °C further reduced transcription reactivation. This is agreement with the fact that these proteins are believed to have evolved to act in response to cold-shock (Jones et al., 1987; Yamanaka, 1999).

Consequently, the blocking of the transcription at low temperatures in the chromosome-integrated gene is likely due to an energy-associated increase in difficulty to release the promoters from the blocked state, rather than due to a reduced functionality of Gyrases or Topoisomerases I. This hypothesis is supported by our results from two additional measurements that showed that subjecting cells to energy depletion via DNP treatment results in a similar transcription dynamics as when at low temperatures, and subjecting cells to consecutive shifts between high and low temperatures results in smooth transitions in the transcription dynamics.

Another evidence for the existence of this promoter blocking in some cells at low temperatures is the very high cell-to-cell variability in RNA numbers in cells at 10 °C with a chromosomally integrated promoter (Table 1). This variability one hour after the start of induction indicates that, while on average only a small amount of RNAs have been produced per cell, this production occurs in a few cells of the population (in agreement with the existence of ‘locked’ promoters in the remaining cells), meaning that those cells with ‘active promoters’ will have produced, most likely, a biologically significant number of RNAs (~1 to 4) (Taniguchi et al., 2010). Nevertheless, it is worth mentioning that such biological significance is expected to differ significantly between genes, depending on various factors. I.e. in some genes, only wide changes in resulting protein numbers suffice to cause phenotypic changes, while for other genes small changes suffice (see e.g. (Choi et al., 2008)).

Interestingly, these results suggest that *E. coli* may not have evolved efficient gene expression regulation mechanisms to make transcription immune to temperature fluctuations (future studies may reveal whether this is a ‘genome-wide’ phenomenon). This may be due to the cumbersome energy requirements involved, that led this organism to capitalize instead on the opportunity to directly leverage the molecular dynamics characteristics at these temperatures to directly encode a temperature-dependent global transcriptional program. Here, we provided evidence that this global program, if existing, could be composed of (at least) the gene’s localization (chromosomal or on a plasmid) and its initiation dynamics. Interestingly, this is in contrast to the complex signaling cascades used by *E. coli* to respond to other environmental stimuli.

We do not know whether this phenomenon of transcriptional halting at low temperatures of $P_{LacO3O1}$, when integrated into the chromosome, is enhanced by the known overexpression at low temperatures of H-NS and similar NAP proteins present in the nucleoid, which, in these conditions, appear to selectively inhibit early step(s) in transcription initiation by binding to the promoter and acting as transcriptional repressors (Ueguchi and Mizuno, 1993). However, the fast recovery of the kinetics of RNA production under the control of $P_{LacO3O1}$ observed in Figure 5, when changing temperature from 10 °C and 30 °C, could be an indication that these proteins are not involved in the phenomenon observed. Similarly, we also do not know whether there is any influence from stringent response mechanisms. Studies of the roles of, e.g., *dsbA* and *ppGpp*, may prove to be of value to determine whether, e.g., the biophysical phenomena here reported are affected by these mechanisms.

Overall, particularly if future studies show that this phenomenon occurs on various genes, we expect our results to contribute to a better understanding of the dynamics of natural genetic circuits at critically low temperatures and to near-future efforts in Synthetic Biology. For example, our results suggest that the behavior of synthetic circuits integrated into plasmids could be more robust to critically low temperatures. On the other hand, if the aim is to avoid leaky RNA production at such temperatures, the implementation should be on the chromosome. Finally, our findings may be of relevance to the understanding of the evolutionary process of genetic circuits, i.e. the factors here studied could be one of the many evolutionary forces behind the localization of genes, not only whether they are integrated on the chromosome or on a plasmid, but also where in the chromosome are they located.

Experimental Procedures

Cells and Plasmids

E. coli strain BW25993 (*lacI^q* *hsdR514* Δ *araBAD*_{AH33} Δ *rhaBAD*_{LD78}) (Datsenko and Warner, 2000) was used to carry the target and reporter genes. The target gene, controlled by $P_{LacO3O1}$, codes for an array of 48 binding sites (Kandavalli et al, 2016; Goncalves et al, 2016) for the modified viral coat protein MS2-GFP (Peabody, 1993; Peabody, 1997; Golding and Cox, 2004). $P_{LacO3O1}$, inducible by IPTG, was engineered from the *E. coli* native *lac* promoter, by removing the O2 repressor binding site downstream of the transcription start site (Oehler et al., 1990). Due to the lacking of the site O2 in $P_{LacO3O1}$, and provided full induction, we expect it not to form significant topological constraints (Fulcrand et al, 2016). Further, also due to lacking the O2 site, the repression strength of LacI on this promoter is expected to be 2-3 fold weaker than on the wild-type promoter (Oehler et al., 1990).

To compare RNA production rates of $P_{LacO3O1}$ when in a single-copy plasmid and when in the chromosome, two strains were engineered from the original BW25993. One strain carries a single copy full F-plasmid (~11 kbp) (Goncalves et al., 2016), pBELOBAC11 (the target plasmid), unknown to form long-lasting bounds to the membrane and originally responsible for the expression of transient DNA-binding proteins (Hayakawa et al., 1985; Mori et al., 1986). In this, we inserted the target gene, $P_{LacO3O1}$, coding for the bindings sites for MS2-GFP (Figure S1 in Supporting Material). In the other strain, the target gene under the control of $P_{LacO3O1}$ was integrated into the *lac* locus of *E. coli*'s genome using Red/ET recombination (performed by Gene Bridges,

Heidelberg, Germany) (Figure S2 in Supporting Material). Additional information is provided in Supplemental Tables S13 and S14.

Both strains were also transformed with the medium copy reporter plasmid pZA25-GFP (Nevo-Dinur et al. 2011) (kind gift from Orna Amster-Choder, Hebrew University of Jerusalem, Israel), coding for the reporter protein MS2-GFP under the control of the P_{BAD} promoter. The multiple MS2-GFP binding sites in the target RNAs and the strong binding affinity of each site allow target RNAs to appear as bright spots soon after produced (see example images in Figure S3 in Supporting Material) (Golding and Cox, 2004). Namely, their maximum fluorescence is reached rapidly (always less than 1 min) (Tran et al., 2015)) and, once reached, it remains constant for several hours (Tran et al., 2015), facilitating their counting, due to lack of interference from RNA degradation (Golding et al., 2004; Tran et al., 2015).

We note that while the strain carrying the target gene in single-copy F-plasmid also contains a native Lac promoter in the chromosome (and thus, has higher number of LacI binding sites overall than the strain carrying the chromosome integrate target gene since the original Lac promoter was replaced by the target one), both strains overexpress LacI, thus reducing the possibility that there would be a significant effect due to ‘shortage’ of repressors in the strain carrying the F-plasmid. Further, our measurements were all conducted under full induction (except to obtain the induction curves), which further reduces any possibility of effects of differences in number of available repressors.

Chemicals, Growth Conditions, and Induction of the Reporter and Target Gene

From single colonies on LB agar plates, cells were cultured in LB medium with the appropriate concentration of antibiotics and incubated overnight at 30 °C and 250 rpm. These overnight cultures were then diluted to an initial 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), which was incubated at 37 °C with a 250 rpm agitation, until reaching an OD_{600} of ~ 0.3.

Regarding antibiotic controls, as we used the same reporter (plasmid-borne) in two strains differing in the location of the target gene, we also use the same antibiotic as a control (kanamycin). As a control for the target gene (when on the single-copy plasmid), we used chloramphenicol.

Next, to induce the expression of the reporter MS2-GFP proteins, 0.4% of L-Arabinose (Sigma-Aldrich, USA) was added and cells were incubated at 37 °C for 30 min with 250 rpm agitation. At this point, cells were incubated at the respective temperature (30 °C, 27 °C, 23 °C, 20 °C, 16 °C and 10 °C) (Innova® 40 incubator, New Brunswick Scientific, USA), for 15 min with agitation, to allow adaptation to a new temperature condition before activating the target gene. For all conditions, the temperature of the culture was constantly measured during the experiments using a thermometer, allowing to ensure that the culture had reached and kept the desired temperature. Following full induction of the target gene (1000 μ M IPTG, Sigma-Aldrich, USA), cells were incubated at the respective temperature for 1 hour, prior to image acquisition or RT-PCR measurements.

To obtain the induction curve of the target gene (Figure 1), cells were incubated with different concentrations of IPTG (0, 50, 100, 250, 500 and 1000 μ M) for 1 hour, at 30 °C, before imaging.

To compare RNA production of the chromosome and plasmid genes at 10 °C and at 30 °C, cells were grown as previously described and incubated at 10 °C and 30 °C, respectively, for 15 minutes prior the induction of the target gene. Next, 1000 μ M of IPTG was added. To determine the target RNA production level of both constructs, while keeping the temperature of the cultures constant, samples were taken right before adding the IPTG (time 0), and afterward for every 15 minutes, for a total of 90 minutes.

We also quantified RNA numbers in cells with repressed Gyrase activity and with repressed Topoisomerase I activity. For that, we used, respectively, Novobiocin and Topotecan (Chen and Liu, 1994; Gellert et al., 1976; Patel et al., 1998). Cells with the target and reporter systems were grown as described below in the section ‘Growth Conditions and Induction of the Reporter and Target Gene’. Following induction of the reporter gene, cells were incubated at the appropriate temperature (10°C or 30°C), at 250 rpm for 15 minutes, prior to induction of the target gene. Afterward, 1000 μ M of IPTG and 100 ng μ l⁻¹ of Novobiocin or 100 μ M of Topotecan were added to the cells.

To determine the RNA levels in cells treated with 2,4-Dinitrophenol (DNP) (known to uncouple the oxidative phosphorylation, thus resulting in the depletion of Adenosine triphosphate) (de Boer, Bakker, Weyer, and Gruber, 1976) the growth and activation of the reporter genes were carried out as previously described. Next, 1000 μ M of IPTG and 200 μ M of DNP were added to the media and cells were incubated at 30°C.

Tuning intracellular RNAP concentrations and nucleoid staining with DAPI

To measure intracellular concentrations of RNA polymerases (RNAP), we used *E. coli* RL1314 strain (a kind gift from Robert Landick, University of Wisconsin-Madison, USA), carrying GFP tagged RNAPs (RNAP-GFP) (Bratton et al., 2011). To change intracellular RNAP concentrations in *E. coli* RL1314 cells, we followed the strategy proposed in (Lloyd-Price et al, 2016). Shortly, it consists of placing cells in media with differing richness that, for a limited range of media richness, results in different intracellular RNAP concentrations without significant differences in cell growth rates between conditions (Lloyd-Price et al, 2016). Further, it was shown that, in that range, the RNA production rate in the cells changes hyperbolically with the RNAP concentrations (i.e. the inverse of this rate changes linearly with the inverse of the RNAP concentration) (Lloyd-Price et al, 2016), as expected from standard models of transcription in *E. coli* obtained from studies using *in vitro* measurement techniques (McClure, 1985).

To obtain cell populations with significantly different mean RNAP concentrations, instead of LB medium with various concentrations of tryptone and yeast extract as in (Lloyd-Price et al, 2016), we used M9 medium and supplemented it with different concentrations of Glycerol. As we set M9 medium supplemented with 0.4% Glycerol as our ‘standard’ medium, this condition is here denoted as ‘1X’. When the Glycerol concentration was increased by 50 % relative to the ‘control’ 1X condition, the altered media was denoted as ‘2X’.

To ensure the validity of this methodology, we studied cell growth rates as a function of Glycerol concentrations. Results are shown in Figure S6 in Supporting Material, confirming that changes in Glycerol concentration did not affect cell growth rates significantly.

The observed changes in fluorescence levels (see example image in Figure S4A in Supporting Material) with varying media richness are consistent with RT-PCR (*rpoC* transcript levels) and plate reading measurements (Lloyd-Price et al., 2016).

Finally, we assessed if the weaker transcriptional activity of the chromosome-integrated promoter at the lowest temperatures could be explained by changes in the known heterogeneities in the spatial distribution of RNAPs or of the nucleoid. For this, starting from a single colony, cells were incubated overnight at 30 °C with agitation and aeration, in LB medium supplemented with 35 µg ml⁻¹ Kanamycin. The overnight culture was then diluted to an initial OD₆₀₀ of 0.05 in the respective M9 medium (0.5X, 1X, 1.5X and 2X), supplemented with Kanamycin, and grown at 37 °C until an OD₆₀₀ of 0.3. At this point, cells were transferred and incubated at the appropriate temperature (10 °C, 16 °C, 23 °C and 30 °C), with agitation, for 75 minutes. Afterward, cells were fixed with 3.7 % formaldehyde, at room temperature, and then centrifuged. To visualize the nucleoid, cells were re-suspended in 1 X PBS, and 4',6-Diamidino-2-Phenylindole (DAPI) (0.2 µg ml⁻¹) was added to this suspension (Figure S4B) (Wery et al., 2001). Next, cells were incubated at room temperature, for 20 minutes, and then washed twice with PBS. Cells were then re-suspended in PBS, and 3 µl aliquots of these samples were placed on 2% agarose pads for imaging.

Results of these tests are reported in the Results section 'Transcription by the chromosome-integrated construct is noisier at lower temperatures'.

RT-PCR

Two sets of RT-PCR measurements were conducted. One to validate the microscope measurements at different temperatures for both target genes and the other to produce τ plots. In both, the target gene is activated as described above and cells were grown as for the microscopy measurements. The reporter gene was not activated, as it was not necessary.

To determine the fold change of the target gene as a function of RNAP concentrations, the *E. coli* strain BW25993 was grown in 0.5X, 1X, 1.5X, and 2X media, as described in the section 'Growth Conditions'. These measurements were conducted in cells kept at 10 °C and 30 °C (the lowest and highest temperature conditions in the microscopy measurements, respectively).

One hour after induction of the target gene, cells were fixed by adding the RNAprotect bacteria reagent (Qiagen, Germany), followed by enzymatic lysis with Tris-EDTA Lysozyme (15 mg ml⁻¹) buffer (pH 8.3). From the lysates, the RNA content was isolated using the RNeasy purification kit (Qiagen) according to 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 found to be intact, with clear bands for the 16S and 23S ribosomal RNA. The RNA yield (~2 µg µl⁻¹) and the A260/A280 nm ratio were determined by a Nanovue Plus Spectrophotometer (GE Healthcare Life Sciences, USA). The ratio obtained (2.0-2.1) is indica-

tive of a highly purified RNA. To remove DNA contamination, the samples were treated with DNaseI (Thermo Scientific, USA) following the manufacturer instructions. The cDNA was synthesized from 1 µg of RNA using the iScript Reverse Transcription Supermix (Biorad, USA) according to the manufacturer instructions.

cDNA samples (10 ng µl⁻¹) were mixed with the qPCR master mix containing iQ SYBR Green Supermix (Biorad, USA) with primers (200nM) for the target and reference genes. The 16S rRNA was used as a reference. Since the sequences of MS2-GFP binding sites consist of many repeats, for this we use the sequence in between the promoter sequence and these sites sequences (Supplementary Figure S2). Namely, the primers set for the target mRNA (mCherry) and reference (16S rRNA) genes were: mCherry (Forward: 5' CACCTACAAGGCCAAGAAGC 3', Reverse: 5' TGGTGTAAGTCCTCGTTGTGG 3'), 16S rRNA (Forward: 5' CGTCAGCTCGTGTGTGAA 3', Reverse: 5' GGACCGCTGGCAA CAAAG 3').

The qPCR experiments were performed using a Biorad MiniOpticon Real-Time PCR System (Biorad, USA). The thermal cycling protocol used was 40 cycles of 95 °C for 10s, 52 °C for 30s, and 72 °C for 30s, with the fluorescence being read after each cycle. All reactions were performed in 3 replicates per condition. The PCR efficiencies of these reactions were greater than 95%. No-RT and no-template controls were used to crosscheck non-specific signals and contamination. The Cq values generated by the CFX ManagerTM Software were used to calculate the fold changes in the target gene, normalized to the reference gene, and its standard error using Livak's 2^{-ΔΔCT} method (Livak and Schmittgen, 2001).

Microscopy

To image cells, cells with the target and reporter genes were grown as described above (section 'Growth Conditions and Induction of the Reporter and Target Gene'). After, cells were pelleted and re-suspended in ~100 µl of the remaining media. Three microliters of cells were placed on a 2% agarose gel pad of M9 medium and kept in between the microscope slide and a coverslip. It took, on average, ~3 minutes to move cells from the incubator to the microscope and start the observation. This time includes the assembly of the microscope imaging-chamber containing the slides and cells.

Cells were visualized by a Nikon Eclipse (Ti-E, Nikon) inverted microscope with a 100x Apo TIRF (1.49 NA, oil) objective. Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system, with a pinhole size set to 1.2 AU. For 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 fluorescence from RNAP-GFP were visualized by a 488 nm laser (Melles-Griot) and an HQ514/30 emission filter (Nikon). Epifluorescence images, for visualization of DAPI-stained nucleoids, were taken by a mercury lamp excitation and a DAPI filter cube (EX 340-380, DM 400, BA 435-485, Nikon). Phase contrast images were taken (for cells segmentation) by an external phase contrast system and DS-Fi2 CCD camera (Nikon). Size of the phase contrast images was 2560×1920 pixels, in which a pixel corresponds to 0.048 µm. Phase contrast and confocal images were taken once and simultaneously by Nis-Elements software (Nikon).

Image Analysis

Cell segmentation from the images was performed by the software “iCellFusion” (Santinha et al., 2015) (Supporting Material, example Figures S3, and S4). iCellFusion first performs automatic cell segmentation from phase contrast images. The results were then manually improved. Next, iCellFusion conducts automatic inter-modal image alignment between the phase-contrast and the corresponding fluorescence images (see example Figure S7 in Supporting Material).

Next, detection of RNA-MS2-GFP fluorescence ‘spots’ inside cells was automatically done in each segmented cell by the software “CellAging” (Häkkinen et al., 2013). Spots were detected by the Kernel Density Estimation (KDE) method using a median filter. An example of cell segmentation and spot detection results are shown in Figures S7 (A) and (B), respectively.

Nucleoid (red channel) and the RNAP (green channel) intensity levels were obtained from the images by iCellFusion as well. For these, Principal Component Analysis (PCA) (Abdi and Williams, 2010) was used to normalize the major and minor axes lengths and the center coordinates of each cell, in order to plot the intensity distribution of the selected fluorescence levels along the major cell axis (by dividing each cell in normalized bins along the major axis). Example images of cells with fluorescent RNAPs and nucleoids are shown in Figures S7 (C) and S10 (D), respectively.

Nucleoid segmentation is made using the algorithm described in (Oliveira et al., 2016a), where a Gaussian approximation is applied. To detect the presence of one nucleoid or two separated nucleoids, we applied the Gradient Path Labelling algorithm (Mora et al., 2011). This selection was manually inspected and corrected. After nucleoid detection and segmentation, principal component analysis was used to obtain the position, dimension and orientation of the nucleoid in each cell (Oliveira et al., 2016a).

To quantify the integer-valued RNA numbers in each cell from the total fluorescence intensity of the spots within, we used the method proposed in (Golding et al., 2005). Briefly, the intensity of a single RNA molecule is set to correspond to the first peak of the distribution of total spots’ intensity in multiple cells. Next, the number of tagged RNAs in each spot in each cell is estimated by dividing its intensity by that of the first peak.

RNA counting from fluorescent spots

We used the method described in (Häkkinen et al., 2013) to quantify the number of MS2-GFP-tagged mRNA molecules in individual cells. First, ‘RNA-spots’ are detected as described in the previous section. Then, the cellular background fluorescence intensity is subtracted from the intensity of each fluorescent ‘RNA-spot’ light intensity, accounting for the number of pixels occupied by the RNA-spot. The resulting RNA-spot intensities are given in arbitrary units (a.u.) and, as such, do not inform on how many RNAs the spot actually contains.

Thus, the next step is to estimate the intensity of a single RNA-spot. For this, given several measurements of RNA-spots intensities, a histogram of the results can be plotted, and the intensity corresponding to the first “peak” of the histogram should be selected. The integer-valued absolute number of RNA molecules in each spot is then calculated by dividing the spot intensity with the intensity of the first peak of the histogram, followed by a rounding to the nearest integer.

Control test of the RNA counting method

In order to determine whether the smallest spots detected by microscopy and image analysis of the cells correspond to a single RNA, we assessed if cells not carrying the target gene while carrying the reporter produced any spots of similar intensity. As shown in Table S1, such ‘Fake spots’ were found to occur only very rarely and, importantly, in much lower numbers than in cells carrying the target gene but non-induced, both in the case of chromosome integration and plasmid-borne.

This difference between the two latter conditions and the first condition shows that ‘leaky’ RNA production is the main responsible for ‘RNA spot detection’ in the two latter conditions, rather than being the appearance of ‘fake spots’ due to MS2-GFP ‘clumpiness’.

Model of transcription kinetics

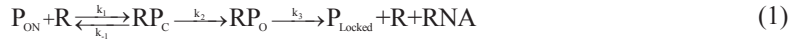
First, we consider chromosome-integrated promoters. We assume the model of transcription whose rates have been empirically validated in (Lloyd-Price et al., 2016) and (Chong et al., 2014). The model includes active transcription, a repression mechanism, and accounts for the effects of DNA super-coiling.

Active transcription (deHaseth et al., 1980; Saecker et al., 2011) consists of a multi-step process (reaction 1) that starts with the binding of an RNA polymerase (R) to a free, active promoter (P_{ON}). This results in the formation of an unstable closed complex, RP_c (McClure, 1985). The reversibility of this step allows multiple occurrences of the closed complex formation between the occurrences of two transcription events.

Once a closed complex successfully leads to the initiation of an open complex (RP_o) formation, the process becomes nearly irreversible (Lutz and Bujard, 1997). Once the DNA double helix is opened and the open complex is formed, the RNA polymerase moves to elongation (clearing the promoter for additional initiation events), and finally, it completes RNA production (Bertrand-Burggraf et al., 1984; Buc and McClure, 1985; Chamberlin, 1974; Chong et al., 2014; Golding and Cox, 2004; Golding et al., 2005; McClure, 1980; Lutz and Bujard, 1997; McClure, 1985; deHaseth et al., 1998; Peccoud and Ycart, 1995; Saecker et al., 2011). Promoter clearing may be preceded by the production of a few short RNA transcripts (<10 nt) from abortive initiation events (deHaseth et al., 1998; Hsu, 2009), but only in rare promoters, with short-living open complexes, are these events rate limiting (Hsu, 2002). As such, these events are not represented here.

Note that the steps in reaction 1 do not represent elementary transitions, but rather the effective rates of rate-limiting steps (Lloyd-Price et al., 2016). E.g., k_1 is the rate at which R binds to the promoter, but its value includes the influence of the time spent in non-specific bindings to the DNA and 1D diffusive searches (Bai et al., 2006). We refer to this collection of events as the closed complex formation. Meanwhile, k_2 is the rate at which a series of steps is completed once initiated (deHaseth et al., 1998). As mentioned above and in accordance with the literature (McClure, 1980), we refer to these steps as the open complex formation. Finally, note that, in

reaction 1, when the forward rate of a given step is much faster than its backward rate, the negligible reversibility is not represented:



Aside from the process of active transcription, there is a repression mechanism, due to the action of LacI molecules (Lloyd-Price et al., 2016; Lutz and Bujard, 1997), accounted for by reaction 2:



Note that we expect that, when the promoter is under full induction (which is the condition that we focus on in the present study), the effects of repression will be very mild in the overall process of RNA production.

Finally, in reaction 1, following a transcription event, it is assumed that the promoter changes directly into a ‘locked’ state, making it unable to initiate a new closed complex until released by the action of Gyrase or Topoisomerase I, depending on the nature of the supercoiled state (Chong et al., 2014; Golding et al., 2005). This model of locking by DNA super-coiling is, to an extent, ‘simplistic’, since this is a cumulative process that does not necessarily occur after each transcription event (as the model assumes), only occurring after “one to a few” events (Chong et al., 2014) (i.e. it is a stochastic, cumulative process). However, as shown in the results section, this approximation does not affect the ability of the model to match the empirical data for the entire range of conditions tested, suggesting that it takes very few events for locking to occur (as the results reported in (Chong et al., 2014) suggest).

The process of ‘unlocking’ from super-coiling states is modeled by reaction 3 (Gyrase and Topoisomerase I are not explicitly represented due to their large copy numbers (Chong et al., 2014)):



The rate of this reaction, k_{unlock} , is assumed the most temperature-dependent variable of this system of three reactions.

Meanwhile, in the case of plasmid-borne promoters, we assume a similar model but with no super-coiling effects, in agreement with (Chong et al., 2014). The model of transcription of plasmid-integrated promoters can thus be obtained from the above by allowing the promoter to change to a P_{ON} state following a transcription event, or equivalently by setting k_{unlock} to ‘infinite’.

In normal conditions, the subsequent steps of elongation, termination and RNA release (deHaas et al., 1978; Harden et al., 2016; Mooney et al., 2005; Raffaele et al., 2005), are much faster than initiation (Epshtein and Nudler, 2003; Erie et al., 1993; Greive and Von Hippel, 2005; Herbert et al., 2006; Mäkelä et al., 2011; Proshkin et al., 2010) in both plasmids and in the chromosome. Thus, k_3 is assumed ‘infinitely’ fast. Further, regardless of the duration of these steps; note that rate-limiting steps in elongation should not affect the mean of the time-length of the intervals between consecutive RNA productions, only their variance (Mannerström et al., 2011).

Finally, simulations of the stochastic models of gene expression were performed by SGNS (Lloyd-Price et al., 2012), a simulator of chemical reaction systems whose dynamics is driven by the Stochastic Simulation Algorithm (Gillespie, 1977) but that allows for multi-time-delayed

reactions (Russel and Zhu, 2006). SGNS also allows hierarchical, interlinked compartments to be created, destroyed and divided at runtime, a feature used to generate dynamically independent model cells that can differ in any of the model features, such as an initial number of molecular components, to lifetimes, to rate-constant values, etc.

Relative mean duration prior to and following commitment to transcription

We use a recently developed technique to dissect the *in vivo* kinetics of transcription initiation in live *E. coli* cells (Kandavalli et al., 2016; Lloyd-Price et al., 2016), based on previous *in vitro* techniques (Buc and McClure, 1985; McClure, 1985; 1980). This method allows us to estimate the mean fraction of time between consecutive transcription events that is taken by the steps *following* the commitment to the open complex formation.

Given our model of transcription, let t_{prior} be the mean time for a *successful* closed complex formation, i.e. it is the mean time-length of all events prior to the commitment of the RNAP to open complex formation. As such, t_{prior} includes the expected time in OFF state as well as the time taken by multiple (failed) attempts to form a stable closed complex. Note that the kinetics of these steps depends on the RNAP intracellular concentrations.

Meanwhile, the remaining time to produce an RNA is denoted t_{after} , and includes all steps *following* the commitment to the open complex formation, e.g. isomerization (Saecker et al., 2011), and prior to commitment to transcription elongation. Relevantly, the kinetics of these steps does not depend on the RNAP intracellular concentrations. Given this, the mean time interval between consecutive RNA productions (Δt) is:

$$\Delta t = t_{\text{prior}} + t_{\text{after}} \quad (4)$$

Since only t_{prior} is affected by a change in RNAP concentration, the new mean interval between consecutive RNA productions after such a change can be written:

$$\Delta t^{\text{new}} = t_{\text{prior}}^{\text{new}} + t_{\text{after}} \quad (5)$$

Where:

$$t_{\text{prior}}^{\text{new}} = S^{-1} \times t_{\text{prior}} \quad (6)$$

with

$$S = \frac{[RNAP]^{\text{new}}}{[RNAP]} \quad (7)$$

From this, one can write:

$$\frac{\Delta t^{\text{new}}}{\Delta t} = \frac{S^{-1} \times t_{\text{prior}} + t_{\text{after}}}{t_{\text{prior}} + t_{\text{after}}} \quad (8)$$

Next, assume that in the new condition the cells contain an infinite number of RNAPs (same assumption as in (Lloyd-Price et al., 2016)). Given this, S^{-1} becomes null and, from equation 8:

$$\frac{t_{\text{after}}}{t_{\text{prior}} + t_{\text{after}}} = \frac{\Delta t^{\text{new}} (RNAP = \infty)}{\Delta t} \quad (9)$$

Given this, from the normalized mean interval between RNA productions and the inverse of the relative RNAP concentration for a few conditions differing in RNAP concentrations (section ‘Tuning intracellular RNAP concentrations’ in Experimental Procedures), one can extrapolate the ratio $\frac{\Delta t^{new}(RNAP=\infty)}{\Delta t}$, thus obtaining $\frac{t_{after}}{\Delta t}$. This extrapolation, done here with the method of Weighted Total Least Squares (Krystek and Anton, 2007) with error in both coordinates, is valid if Δt changes linearly with changes in the inverse of the RNAP concentrations (shown to be valid within a certain range of media richness (Lloyd-Price et al., 2016; Kandavalli et al., 2016)). From this, one can also obtain the fraction of time prior to commitment to the open complex formation:

$$\frac{t_{prior}}{\Delta t} = 1 - \frac{\Delta t^{new}(RNAP=\infty)}{\Delta t} \quad (10)$$

Here, the empirical values of Δt are estimated from the inverse of the RNA production rates obtained by RT-PCR, as in (Kandavalli et al., 2016). While these values are relative to a reference gene (here 16S RNA), the ratio between rates (r) in the ‘new’ and ‘control’ conditions equal the inverse of the ratio between the time intervals between consecutive RNA productions in individual cells, since RNA degradation rates do not differ with media richness within the range employed here (Chen et al., 2015).

Confidence Intervals

Using Fieller's theorem (Fieller, 1954), we can derive the 90% CI of the ratio between the reference condition B and the tested condition A (Motulsky, 1995) as follows:

$$90\%CI = \left(\frac{mean(A)}{mean(B)} \right) \pm \left(t^* \times \left(\frac{mean(A)}{mean(B)} \right) \times \sqrt{\left(\frac{(\sigma_A)^2}{mean(A)^2} \right) + \left(\frac{(\sigma_B)^2}{mean(B)^2} \right) - 2 \left(\frac{(\sigma_{A,B})}{mean(A) * mean(B)} \right)} \right) \quad (11)$$

Where t^* is the critical Student's t-value for the degrees of freedom of the sum of A and B samples minus 2. Since we perform two-side tests, we search the table for the critical values for a probability of 0.95.

Number of promoter copies during the cell lifetime

In general, we assume that there is only one copy of the target promoter in a cell at all times, for both strains. To determine the extent to which this assumption is accurate, we measured the fraction of cells containing two chromosomes, 1 hour after the start of the microscopy measurements. Since, in *E. coli*, single-copy F-plasmids replicate at the same time (Cooper and Keasling, 1998) or shortly after (Keasling et al., 1991) the chromosome, we only measured chromosome numbers, both in cells carrying the gene of interest in the chromosome as well as in cells carrying the gene of interest in the single copy F-plasmid.

For this, cells of both strains were grown as described in Experimental Procedures. Next, they were fixed and stained with DAPI to assess the location and size of nucleoids in live cells (Experimental Procedures, section ‘‘Tuning intracellular RNAP concentrations and nucleoid staining with DAPI’’).

From images taken 1 hour after the start of the microscopy measurements, cells and nucleoids were segmented as described in Experimental Procedures (section ‘Image Analysis’). Results are reported in the third paragraph of the Results section. An example image of DAPI-stained nucleoid(s) is shown in Figure S7-D in Supporting Material.

Plasmid copy number calculation using RT-qPCR

To assess the plasmid copy number, we followed the method described in (Lee et al., 2006). Briefly, it allows the absolute quantification of plasmid copy number using a standard curve that correlates the copy number of a gene with the C_T (cycle threshold) value obtained from RT-qPCR measurements. For this, the separate detection of the plasmid and the host strain chromosomal DNA is required, which is achieved by using two primer sets, one specific for a single copy gene present in the plasmid and the other present in the host strain chromosome.

Here, we used two sets of primers, one for the Chloramphenicol acetyltransferase (*cat*) gene (a single copy gene on the pBELO plasmid), and another for the *lacI* gene (a single copy gene on the *E. coli* BW25993 chromosome). For obtaining the curve, these genes were inserted in the same plasmid (gene ratio 1:1), which then can be used to calculate the copy number of the plasmid of interest. Since *cat* and *lacI* are single-copy genes, the plasmid copy number is the ratio of *cat* to *lacI*.

Cells carrying the plasmid-borne gene were grown as described in the section “Growth Conditions and Induction of the Reporter and Target Gene” of the Experimental Procedures. After the induction of the target gene (at 10°C and at 30°C), the total DNA content of the cells was extracted using the QIAamp DNA Mini kit (Qiagen) following the method for bacterial cells as per manufacturer’s instructions.

For the construction of the standard curves, we used the plasmid pCA24N-ligase (Kitagawa et al., 2005; Wilson et al., 2013), which carries the coding sequence for both the *cat* and *lacI* genes. This plasmid was purified from an overnight culture grown in LB medium at 37°C, using the QIAprep Spin Miniprep kit (Qiagen).

The DNA concentration of both samples (the plasmid and the extracted DNA from *E. coli* BW25993 cells) was determined using the Qubit 4 Fluorimeter (ThermoFisher Scientific) and the Qubit 1X dsDNA HS kit (ThermoFisher Scientific).

A 10-fold serial dilution of the pCA24N-ligase plasmid, ranging from 10^4 to 10^9 copies μl^{-1} , was performed to obtain the standard curves for the *cat* and *lacI* genes. After determining the plasmid concentration in each of these dilutions, the plasmid copy number was calculated from (Lee et al., 2006; Whelan et al., 2003):

$$\text{DNA(copy)} = \frac{6.02 \times 10^{23} (\text{copy/mol}) \times \text{DNA amount (g)}}{\text{DNA length (dp)} \times 660 (\text{g/mol/dp})} \quad (12)$$

The real-time qPCR was performed using a Biorad MiniOpticon Real-Time PCR System (Biorad, USA). The RT-qPCR reaction mixture contained the iQ SYBR Green Supermix (Biorad,

USA), the DNA template, and the primers for the target (*cat*) and reference (*lacI*) genes at a final concentration of 200nM. The primers for the *cat* gene were (Forward: 5' ATTACATTCTTGCCCGCC 3' and Reverse: 5' CACCGTAACACGCCACATC 3') and for the *lacI* were (Forward: 5' ACCAGGATGCCATTGCTGTG 3' and Reverse: 5' TTTATGCCAGCCAGCCAGAC 3'), with the amplicon size being 209 and 221, respectively. The thermal cycling protocol used was: denaturation at 95°C for 3 min, followed by 40 cycles of 95°C for 10s, 58°C for 30s and 72°C for 30s, with the fluorescence being read at the end of each cycle. For all the samples tested, (the DNA extracted from *E. coli* BW25993 cells and for each of the pCA24N-ligase serial dilution), the reaction was conducted in triplicates, with a final volume of 25 µl. No-template controls were used to crosscheck for non-specific signals and contamination. After the amplification, a melting curve analysis, with a temperature gradient of 0.5°C/s from 65°C to 95°C, was performed to confirm the specificity of the amplification. The C_T values generated by the CFX Manager™ Software were then used to obtain the standard curve for both genes, where these C_T values were plotted against the logarithm of their initial template copy numbers, determined using equation (12). In addition, the C_T obtained from the DNA extracted from *E. coli* BW25993 cells was used, along with the standard curve to determine the pBELO plasmid copy-number.

Each standard curve was generated using a linear regression of the plotted data. From the slope of each standard curve, the amplification efficiency (E) was determined from (Lee et al., 2006; Rasmussen, 2001):

$$E = 10^{-1/\text{slope}} - 1 \quad (13)$$

The standard curves for *cat* and *lacI*, ranging from 10⁴ to 10⁹ copies µl⁻¹, are presented in Figure S9 in Supporting Material. Both curves were linear, in the tested range, with a R² > 0.997 and 0.999, respectively. The slopes of each standard curve were -3.12 and -3.08, for *cat* and *lacI*, respectively. From these, we determined the amplification efficiency for both genes, with the results being 1.10 for both *cat* and *lacI*.

The absolute quantification of the plasmid copy number was determined using the standard curves in Figure S9. The copy number of the *cat* and *lacI* genes in the total DNA extracted from *E. coli* BW25993 cells was determined from the corresponding standard curve, using the C_T values. The plasmid copy number of the pBELO was then calculated by dividing the copy number of *cat* by the copy number of *lacI*, given that both are single copy genes of pBELO and BW25993 chromosome, respectively. Thus, the ratio of *cat* to *lacI* is equal to the plasmid copy number of pBELO. The results from this quantification are shown in Table S12 and described in the Results section.

Half-life of the fluorescence intensity of MS2-GFP tagged RNA molecules

We observed several cells with a single MS2-GFP tagged RNA molecule for 1 hour, since the moment the tagged RNA was first detected. Next, we fitted the intensity of each such RNA over time with a decaying exponential function and inferred the degradation rate of its fluorescence intensity. We obtained the decaying rates, calculated the mean of these rates and then converted it into the mean half-life of MS2-GFP tagged RNAs.

The results in Table S2 in Supporting Material agree with previous analyses of the coat protein of bacteriophage, MS2, which showed that the MS2 binding sites on the RNA are constantly occupied by MS2-GFP proteins provided that these exist in sufficient abundance, resulting in the ‘immortalization’ of the target RNA due to isolation from RNA-degrading enzymes for time-series microscopy measurements of 1-2 hours (see e.g. (Golding et al., 2005)).

Author Contributions

ASR conceived the study. NSMG performed experiments. LM and SMDO performed image and data analysis. SMDO and RN-V designed genetic constructs. JR and HK performed genome integrations. JMF assisted in data analysis. ASR and SMDO created models. ASR, NSMG, SMDO, LM, and JL-P wrote the manuscript, which was revised by all authors. All authors performed research, contributed to the experimental design, discussion, and paper writing.

Acknowledgments

We thank Sofia Startceva and Huy Tran for valuable advices on data analysis and models. Work supported by Portuguese Foundation for Science and Technology FCT/MCTES (SFRH/BD/88987/2012 to L.M.), Vilho, Yrjö and Kalle Vaisala Foundation (to S.M.D.O.), Tampere University of Technology President’s Graduate Programme (R.N.-V.), FCT Strategic Program (UID/EEA/00066/203 to J.M.F.), Academy of Finland (295027 to A.S.R.), Academy of Finland Key Project Funding (305342 to A.S.R.), and Jane and Aatos Erkko Foundation (610536 to A.S.R.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. The authors declare that they have no conflict of interest.

References

- Abdi, H., and Williams, L.J. (2010) Principal Component Analysis. *Wiley Interdiscip Rev Comput Stat* 2(4), 433–459.
- Amit, R., Oppenheim, A.B., and Stavans, J. (2003) Increased Bending Rigidity of Single DNA Molecules by H-NS, a Temperature and Osmolarity Sensor. *Biophys J* 84(4), 2467–2473.
- Arsène, F., Tomoyasu, T., and Bukau, B. (2000) The heat shock response of *Escherichia coli*. *Int. J Food Microbiol* 55(1), 3–9.
- Bai, L., Santangelo, T.J., and Wang, M.D. (2006) Single-Molecule Analysis of Rna Polymerase Transcription. *Annu Rev Biophys Biomol Struct* 35, 343–360.
- Bernstein, J.A., Khodursky, A.B., Pei-Hsun, L., 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. *Proc Natl Acad Sci U. S. A.* 99(15), 9697–9702.
- Bertrand-Burggraf, E., Lefèvre, J.F., and Daune, M. (1984) A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322. *Nucleic Acids Res* 12(3), 1697–1706.

1180 Blattner, F.R., Plunkett III, G., Bloch, C.A., Perna, N.T., Burland, V., Riley, M., Collado-Vides,
1181 J., Glasner, J.D., Rode, C.K., Mayhew, G.F., Gregor, J., Davis, N.W., Kirkpatrick, H.A.,
1182 Goeden, M.A., Rose, D.L., Mau, B., and Shao, Y. (1997) The complete genome sequence
1183 of *Escherichia coli* K-12. *Science* 277(5331), 1453–1462.

1184 Blot, N., Mavathur, R., Geertz, M., Travers, A., and Muskhelishvili, G. (2006) Homeostatic reg-
1185 ulation of supercoiling sensitivity coordinates transcription of the bacterial genome.
1186 *EMBO Rep* 7(7), 710–715.

1187 Boeke, J.D., and Model, P. (1982) A prokaryotic membrane anchor sequence: Carboxyl terminus
1188 of bacteriophage ϕ 1 gene III protein retains it in the membrane. *Proc Natl Acad Sci U. S.*
1189 *A.* 79(17), 5200–5204.

1190 Bratton, B.P., Mooney, R.A., and Weisshaar, J.C. (2011) Spatial distribution and diffusive mo-
1191 tion of RNA polymerase in live *Escherichia coli*. *J Bacteriol* 193(19), 5138–5146.

1192 Brewster, R.C., Weinert, F.M., Garcia, H.G., Song, D., Rydenfelt, M., and Phillips, R. (2014)
1193 The transcription factor titration effect dictates level of gene expression. *Cell* 156(9),
1194 1312–23.

1195 Browning, D.F., and Busby, S.J.W. (2016) Local and global regulation of transcription initiation
1196 in bacteria. *Nat Rev Microbiol* 14(10), 638–650.

1197 Buc, H., and McClure, W.R. (1985) Kinetics of open complex formation between *Escherichia*
1198 *coli* RNA polymerase and the lacUV5 promoter. Evidence for a sequential mechanism in-
1199 volving three steps. *Biochemistry* 24(11), 2712–2723.

1200 Cabrera, J.E., and Jin, D.J. (2006) Active transcription of rRNA operons is a driving force for the
1201 distribution of RNA polymerase in bacteria: Effect of extrachromosomal copies *rrnB* on
1202 the in vivo localization of RNA polymerase. *J Bacteriol* 188(11), 4007–4014.

1203 Chamberlin, M.J. (1974) The selectivity of transcription. *Annu Rev Biochem* 43(1), 721–775.

1204 Chen, A.Y., and Liu, L.F. (1994) DNA topoisomerases: Essential enzymes and lethal targets.
1205 *Annu Rev Pharmacol Toxicol* 34, 191–218.

1206 Chen, H., Shiroguchi, K., Ge, H., and Xie, X.S. (2015) Genome-wide study of mRNA degrada-
1207 tion and transcript elongation in *Escherichia coli*. *Mol Syst Biol* 11(1), 781.

1208 Cho, B.-K., Kim, D., Knight, E.M., Zengler, K., and Palsson B.O. (2014) Genome-scale recon-
1209 struction of the sigma factor network in *Escherichia coli*: topology and functional states.
1210 *BMC Biol* 12:4.

1211 Choi, P.J., Cai, L., Frieda, K., and Xie, X.S. (2008) A stochastic single-molecule event triggers
1212 phenotype switching of a bacterial cell. *Science* 322(5900), 442–446.75.

1213 Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014) Mechanism of Transcriptional Bursting in Bac-
1214 teria. *Cell* 158(2), 314–326.

1215 Cooper, S., and Keasling, J.D. (1998) Cycle-specific replication of chromosomal and F-plasmid
1216 origins. *FEMS Microbiol Lett* 163(2), 217–222.

1217 Datsenko, K., and Wanner, B. (2000) One-step inactivation of chromosomal genes in *Escherichia*
1218 *coli* K-12 using PCR products. *Proc Natl Acad Sci U. S. A.* 97(12), 6640–5.

1219 Davies, J., and Davies, D. (2010) Origins and evolution of antibiotic resistance. *Microbiol Mol*
1220 *Biol Rev* 74(3), 417–433.

1221 de Boer, H.A., Bakker, A.J., Weyer, W.J., and Gruber, M. (1976) The role of energy-generating
1222 processes in the degradation of guanosine tetraphosphate, ppGpp, in *Escherichia coli*. *Bi-*
1223 *ochim Biophys Acta (BBA)-Nucleic Acids Protein Synth* 432(3), 361–368.

1224 deHaseth, P.L., Lohman, T.M., Burgess, R.R. and Record, M.T. (1978) Nonspecific interactions
1225 of *Escherichia coli* RNA polymerase with native and denatured DNA: differences in the
1226 binding behavior of core and holoenzyme. *Biochemistry* 17(9), 1612–1622.

1227 deHaseth, P.L., Zupancic, M.L. and Record, M.T. (1998) RNA polymerase-promoter interac-
1228 tions: the comings and goings of RNA polymerase. *J Bacteriol* 180(12), 3019–3025.

1229 Deng, S., Stein, R. A., and Higgins, N.P. (2004) Transcription-induced barriers to supercoil dif-
1230 fusion in the *Salmonella typhimurium* chromosome. *Proc Natl Acad Sci U. S. A.* 101(10),
1231 3398–3403.

1232 Deng, S., Stein, R.A., and Higgins, N.P. (2005) Organization of supercoil domains and their re-
1233 organization by transcription. *Mol Microbiol* 57(6), 1511–1521.

1234 Dillon, S.C., and Dorman, C.J. (2010) Bacterial nucleoid-associated proteins, nucleoid structure
1235 and gene expression. *Nat Rev Microbiol* 8(3), 185–195.

1236 Dong, T., and Schellhorn, H.E. (2009) Global effect of RpoS on gene expression in pathogenic
1237 *Escherichia coli* O157:H7 strain EDL933. *BMC Genomics* 10:349.

1238 Dorman, C.J. (2006) DNA supercoiling and bacterial gene expression. *Sci Prog* 89(Pt 3-4), 151–
1239 166.

1240 Dorman, C.J., and Dorman, M.J. (2016) DNA supercoiling is a fundamental regulatory principle
1241 in the control of bacterial gene expression. *Biophys Rev* 8(Suppl 1), 89–100.

1242 Drlica, K. (1992) Control of bacterial DNA supercoiling. *Mol Microbiol* 6(4), 425–433.

1243 Epshtein, V., and Nudler, E. (2003) Cooperation Between RNA Polymerase Molecules in Tran-
1244 scription Elongation. *Science* 300(5620), 801–805.

1245 Erie, D.A., Hajiseyedi, O., Young, M.C., and von Hippel, P.H. (1993) Multiple RNA pol-
1246 ymerase conformations and GreA: control of the fidelity of transcription. *Science*
1247 262(5135), 867–873.

1248 Farewell, A., Kvint, K., and Nyström, T. (1998) Negative regulation by RpoS: A case of sigma
1249 factor competition. *Mol Microbiol* 29(4), 1039–1051.

1250 Fieller, E.C. (1954) Some Problems in Interval Estimation. *J R Stat Soc Ser B* 16(2), 175–185.

1251 Fulcrand, G., Dages, S., Zhi, X., Chapagain, P., Gerstman, B.S., Dunlap D., and Leng, F. (2016)
 1252 DNA supercoiling, a critical signal regulating the basal expression of the lac operon in
 1253 *Escherichia coli*. *Sci Rep* 6, 1–12.

1254 Gellert, M., O’Dea, M.H., Itoh, T., and Tomizawa, J.I. (1976). Novobiocin and coumermycin
 1255 inhibit DNA supercoiling catalyzed by DNA gyrase. *Proc Natl Acad Sci U. S. A.* 73(12),
 1256 4474–4478.

1257 Gillespie, D. T. (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem*
 1258 81(25), 2340–2361.

1259 Golding, I., and Cox, E.C. (2004) RNA dynamics in live *Escherichia coli* cells. *Proc Natl Acad*
 1260 *Sci U. S. A.* 101(31), 11310–11315.

1261 Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005) Real-time kinetics of gene activi-
 1262 ty in individual bacteria. *Cell* 123(6), 1025–1036.

1263 Goldstein, E., and Drlica, K. (1984) Regulation of bacterial DNA supercoiling: plasmid linking
 1264 numbers vary with growth temperature. *Proc Natl Acad Sci U. S. A.* 81(13), 4046–4050.

1265 Goncalves, N.S.M., Martins, L., Tran, H., Oliveira, S.M.D., Neeli-Venkata, R., Fonseca, J.M.,
 1266 Ribeiro, A.S. (2016) In vivo single-molecule dynamics of transcription of the viral T7 Phi
 1267 10 promoter in *Escherichia coli*. Proceedings of the 8th International Conference on Bio-
 1268 informatics, Biocomputational Systems and Biotechnologies (BIOTECHNO 2016), June
 1269 26–30, 2016, Lisbon, Portugal, pp. 9–15. ISBN: 978–1–61208-488-6.

1270 Goncalves, N.S.M., Oliveira, S.M.D., Kandavalli, V., Fonseca, J.M., and A.S. Ribeiro. (2016)
 1271 Temperature dependence of leakiness of transcription repression mechanisms of *E. coli*.
 1272 In: Bartocci E, P Lio, N Paoletti, editors. Lecture Notes in Bioinformatics. Springer Berlin
 1273 Heidelberg. pp. 342–343.

1274 Greive, S.J., and von Hippel, P.H. (2005) Thinking quantitatively about transcriptional regula-
 1275 tion. *Nat Rev Mol Cell Biol* 6, 221–232.

1276 Häkkinen, A., Muthukrishnan, A.-B., Mora, A., Fonseca, J.M., and Ribeiro, A.S. (2013) CellAg-
 1277 ing: a tool to study segregation and partitioning in division in cell lineages of *Escherichia*
 1278 *coli*. *Bioinformatics* 29(13), 1708–1709.

1279 Harden, T.T., Wells, C.D., Friedman, L.J., Landick, R., Hochschild, A., Kondev, J., and Gelles,
 1280 J. (2016) Bacterial RNA polymerase can retain $\sigma 70$ throughout transcription. *Proc Natl*
 1281 *Acad Sci U. S. A.* 113(7), 602–607.

1282 Hardy, C.D., and Cozzarelli, N.R. (2005) A genetic selection for supercoiling mutants of *Esche-*
 1283 *richia coli* reveals proteins implicated in chromosome structure. *Mol Microbiol* 57(6),
 1284 1636–1652.

1285 Hayakawa, Y., Murotsu, T., and Matsubara, K. (1985) Mini-F protein that binds to a unique re-
 1286 gion for partition of mini-F plasmid DNA. *J Bacteriol* 163(1), 349–354.

1287 Herbert, K.M., La Porta, A., Wong, B.J., Mooney, R.A., Neuman, K.C., Landick, R., and Block,
1288 S.M. (2006) Sequence-Resolved Detection of Pausing by Single RNA Polymerase Mole-
1289 cules. *Cell* 125(6), 1083–1094.

1290 Higgins NP. (2016) Species-specific supercoil dynamics of the bacterial nucleoid. *Biophys Rev*
1291 8(Suppl 1): 113-121.

1292 Higgins, N.P., and Vologodskii, A.V. (2015) Topological Behaviour of Plasmid DNA. *Microbiol*
1293 *Spectr* 3(2), 1–49.

1294 Holmes, V.F., and Cozzarelli, N.R. (2000) Closing the ring: links between SMC proteins and
1295 chromosome partitioning, condensation, and supercoiling. *Proc Natl Acad Sci U. S. A.*
1296 97(4), 1322–1324.

1297 Hsu, L. (2009) Monitoring abortive initiation. *Methods* 47(1), 25–36.

1298 Hsu, L.M. (2002) Promoter clearance and escape in prokaryotes. *Biochim Biophys Acta - Gene*
1299 *Struct Expr* 1577(2), 191–207.

1300 Jones, P.G., VanBogelen, R.A., and Neidhardt, F.C. (1987) Induction of proteins in response to
1301 low temperature in *Escherichia coli*. *J Bacteriol* 169(5), 2092–2095.

1302 Kandavalli, V.K., Tran, H., and Ribeiro, A.S. (2016) Effects of σ factor competition are promot-
1303 er initiation kinetics dependent. *Biochim Biophys Acta (BBA) - Gene Regul Mech*
1304 1859(10), 1281–1288.

1305 Kannan, G., Wilks, J.C., D.M., Fitzgerald, Jones, B.D., BonDurant, S.S. and Slonczewski, J.L.
1306 (2008) Rapid acid treatment of *Escherichia coli*: transcriptomic response and recovery.
1307 *BMC Microbiol* 8(1), 1-13.

1308 Kapanidis, A. N., Margeat, E., Ho, S.O., Kortkhonjia, E., Weiss, S., Ebright, R.H. (2006) Initial
1309 Transcription by RNA Polymerase Proceeds Through a DNA-Scrunching Mechanism.
1310 *Science* 314(5802), 1144-1147.

1311 Keasling, J.D., Palsson, B.Ø., and Cooper, S. (1991) Cell-cycle-specific F plasmid replication:
1312 regulation by cell size control of initiation. *J. Bacteriol* 173(8), 2673–2680.

1313 Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., and
1314 Mori, H. (2005) Complete set of ORF clones of *Escherichia coli* ASKA library (A com-
1315 plete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA Res*
1316 12(5), 291–299.

1317 Krystek, M., and Anton, M. (2007) A weighted total least-squares algorithm for fitting a straight
1318 line. *Meas Sci Technol* 18(11), 3438-3442.

1319 Lee, C., Kim, J., Shin, S.G., and Hwang, S. (2006) Absolute and relative QPCR quantification of
1320 plasmid copy number in *Escherichia coli*. *J Biotechnol* 123(3), 273–280.

1321 Leng, F., Chen, B., and Dunlap, D.D. (2011) Dividing a supercoiled DNA molecule into two
1322 independent topological domains. *Proc Natl Acad Sci U. S. A.* 108(50), 19973–19978.

1323 Lineweaver, H., and Burk, D. (1934) The determination of enzyme dissociation constants. *J Am*
1324 *Chem Soc* 56(3), 658–666.

1325 Liu, L.F., and Wang, J.C. (1987). Supercoiling of the DNA template during transcription. *Proc*
1326 *Natl Acad Sci U. S. A.* 84(20), 7024–7027.

1327 Livak, K.J., and Schmittgen, T.D. (2001) Analysis of relative gene expression data using real-
1328 time quantitative PCR and the 2(-Delta Delta C(T)) method. *Methods* 25(4), 402–408.

1329 Lloyd-Price, J., Gupta, A., and Ribeiro, A.S. (2012) SGNS2: A Compartmentalized Stochastic
1330 Chemical Kinetics Simulator for Dynamic Cell Populations. *Bioinformatics* 28(22), 3004-
1331 3005.

1332 Lloyd-Price, J., Startceva, S., Chandraseelan, J.G., Kandavalli, V., Goncalves, N.S.M., Oliveira,
1333 S.M.D., Häkkinen, A., and Ribeiro, A.S. (2016) Dissecting the stochastic transcription ini-
1334 tiation process in live *Escherichia coli*. *DNA Res* 23(3), 203–214.

1335 López-García, P., and Forterre, P. (2000) DNA topology and the thermal stress response, a tale
1336 from mesophiles and hyperthermophiles. *BioEssays*. 22(8), 738–746.

1337 Los, D.A. (2004) The effect of low-temperature-induced DNA supercoiling on the expression of
1338 the desaturase genes in synechocystis. *Cell Mol Biol* (Noisy-le-grand). 50(5), 605–612.

1339 Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Esch-*
1340 *erichia coli* via the LacR/O, the TetR/O and AraC/I1-I 2 regulatory elements. *Nucleic Ac-*
1341 *ids Res* 25(6), 1203–1210.

1342 Lynch, A.S., and Wang, J.C. (1993) Anchoring of DNA to the bacterial cytoplasmic membrane
1343 through cotranscriptional synthesis of polypeptides encoding membrane proteins or pro-
1344 teins for export: A mechanism of plasmid hypernegative supercoiling in mutants deficient
1345 in DNA topoisomerase I. *J Bacteriol* 175(6), 1645–1655.

1346 Ma, D., Cook, D. N., Alberti, M., Pon, N. G., Nikaido, H., and Hearst, J. E. (1995) Genes *acrA*
1347 and *acrB* encode a stress-induced efflux system of *Escherichia coli*. *Mol Microbiol* 16(1),
1348 45–55.

1349 Mäkelä, J., Kandavalli, V., and Ribeiro, A.S. (2017) Rate-limiting steps in transcription dictate
1350 sensitivity to variability in cellular components. *Sci Rep* 7, 10588.

1351 Mäkelä, J., Lloyd-Price, J., Yli-Harja, O., and Ribeiro, A.S. (2011) Stochastic sequence-level
1352 model of coupled transcription and translation in prokaryotes. *BMC Bioinformatics* 12(1),
1353 121.

1354 Mannerström, H., Yli-Harja, O., and Ribeiro, A.S. (2011) Inference of kinetic parameters of de-
1355 layed stochastic models of gene expression, using a Markov chain approximation. *Eurasip*
1356 *J Bioinforma. Syst Biol* 1, 11–15.

1357 Marr, C., Geertz, M., Hütt, M.-T., and Muskhelishvili, G. (2008) Dissecting the logical types of
1358 network control in gene expression profiles. *BMC Syst Biol* 2:18.

1359 Massey, F.J. (1951) The Kolmogorov-Smirnov Test for Goodness of Fit. *J Am Stat Assoc*
1360 46(253), 68–78.

1361 McClure, W.R. (1980) Rate-limiting steps in RNA chain initiation. *Proc Natl Acad Sci U. S. A.*
1362 77(10), 5634–5638.

1363 McClure, W.R. (1985) Mechanism and control of transcription initiation in prokaryotes. *Annu*
1364 *Rev Biochem* 54(1), 171–204.

1365 Mooney, R.A., Darst, S.A., and Landick, R. (2005) Sigma and RNA polymerase: An on-again,
1366 off-again relationship? *Mol Cell* 20(3), 335–345.

1367 Mora, A.D., Vieira, P.M., Manivannan, A., and Fonseca, J.M. (2011) Automated drusen detec-
1368 tion in retinal images using analytical modelling algorithms. *Biomed Eng Online* 10: 59.

1369 Mori, H., Kondo, A., Ohshima, A., Ogura, T., and Hiraga, S. (1986) Structure and Function of
1370 the F Plasmid Genes Essential for Partitioning. *J Mol Biol* 192(1), 1–15.

1371 Motulsky, H. (1995) Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking.

1372 Moulin, L., Rahmouni, A.R., and Boccard, F. (2005) Topological insulators inhibit diffusion of
1373 transcription-induced positive supercoils in the chromosome of *Escherichia coli*. *Mol Mi-*
1374 *crobiol* 55(2), 601–610.

1375 Muthukrishnan, A.-B., Martikainen, A., Neeli-Venkata, R., and Ribeiro, A.S. (2014) In vivo
1376 Transcription Kinetics of a Synthetic Gene Uninvolved in Stress-Response Pathways in
1377 Stressed *Escherichia coli* cells. *PLoS ONE* 9(9), e109005.

1378 Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., and Amster-Choder, O. (2011) Trans-
1379 lation-independent localization of mRNA in *E. coli*. *Science* 331(6020), 1081–1084.

1380 Ochman, H., Lawrence, J.G., and Groisman, E.A. (2000) Lateral gene transfer and the nature of
1381 bacterial innovation. *Nature* 405(6784), 299–304.

1382 Oehler, S., Eismann, E.R., Krämer, H., and Müller-Hill, B. (1990) The three operators of the lac
1383 operon cooperate in repression. *EMBO J* 9(4), 973–979.

1384 Oliveira, S.M.D., Chandraseelan, J.G., Häkkinen, A., Goncalves, N.S.M., Yli-Harja, O., Startce-
1385 va, S., and Ribeiro, A.S. (2015) Single-cell kinetics of a repressilator when implemented
1386 in a single-copy plasmid. *Mol Biosyst* 11(7), 1939–1945.

1387 Oliveira, S.M.D., Häkkinen, A., Lloyd-Price, J., Tran, H., Kandavalli, V., and Ribeiro, A.S.
1388 (2016b) Temperature-Dependent Model of Multi-step Transcription Initiation in *Esche-*
1389 *richia coli* Based on Live Single-Cell Measurements. *PLoS Comput Biol* 12(10),
1390 e1005174.

1391 Oliveira, S.M.D., Neeli-Venkata, R., Goncalves, N.S.M., Santinha, J.A., Martins, L., Tran, H.,
1392 Mäkelä, J., Gupta, A., Barandas, M., Häkkinen, A., Lloyd-Price, J., Fonseca, J.M., and
1393 Ribeiro, A.S. (2016a) Increased cytoplasm viscosity hampers aggregate polar segregation
1394 in *Escherichia coli*. *Mol Microbiol* 99(4), 686–699.

1395 Parry, B.R., Surovtsev, I.V., Cabeen, M.T., O'Hern, C.S., Dufresne, E.R., and Jacobs-Wagner,
1396 C. (2014) The bacterial cytoplasm has glass-like properties and is fluidized by metabolic
1397 activity. *Cell* 156(1-2), 183–194.

1398 Patel, K., Craig, S.B., McBride, M.G., and Palepu, N.R. (1998) Microbial inhibitory properties
1399 and stability of topotecan hydrochloride injection. *Am J Heal Pharm* 55(15), 1584–1587.

1400 Peabody, D.S. (1993) The RNA binding site of bacteriophage MS2 coat protein. *EMBO J* 12(2),
1401 595–600.

1402 Peabody, D.S. (1997) Role of the coat protein-RNA interaction in the life cycle of bacteriophage
1403 MS2. *Mol Gen Genet* 254(4), 358–364.

1404 Peccoud, J., and Ycart, B. (1995) Markovian modelling of gene product synthesis. *Theor Popul*
1405 *Biol* 48(2), 222–234.

1406 Peter, B.J., Arsuaga, J., Breier, A.M., Khodursky, A.B., Brown, P.O., and Cozzarelli, N.R.
1407 (2004) Genomic transcriptional response to loss of chromosomal supercoiling in *Esche-*
1408 *richia coli*. *Genome Biol* 5(11), R87.

1409 Postow, L., Hardy, C.D., Arsuaga, J., and Cozzarelli, N.R. (2004) Topological domain structure
1410 of the *Escherichia coli* chromosome. *Genes Dev* 18(14), 1766–1779.

1411 Proshkin, S., Rahmouni, A.R., Mironov, A., and Nudler, E. (2010) Cooperation between translat-
1412 ing ribosomes and RNA polymerase in transcription elongation. *Science* 328(5977), 504–
1413 508.

1414 Pruss, G.J., and Drlica, K. (1986) Topoisomerase I mutants: the gene on pBR322 that encodes
1415 resistance to tetracycline affects plasmid DNA supercoiling. *Proc Natl Acad Sci U. S. A.*
1416 83(23), 8952–6.

1417 Pruss, G.J., and Drlica, K. (1989) DNA supercoiling and prokaryotic transcription. *Cell* 56, 521–
1418 523.

1419 Raffaele, M., Kanin, E.I., Vogt, J., Burgess, R.R. and Ansari, A.Z. (2005) Holoenzyme switch-
1420 ing and stochastic release of sigma factors from RNA polymerase in vivo. *Mol Cell* 20(3),
1421 357–366.

1422 Rahmouni, A.R., and Wells, R.D. (1992) Direct Evidence for the Effect of Transcription on
1423 Local DNA Supercoiling *in vivo*. *J Mol Biol* 223(1),: 131–144.

1424 Rankin, D.J., Rocha, E.P.C., and Brown, S.P. (2011) What traits are carried on mobile genetic
1425 elements, and why? *Heredity (Edinb)* 106(1), 1–10.

1426 Rasmussen, R. (2001) Quantification on the LightCycler. In: Meuer S, C Wittwer, K Nakaga-
1427 wara, (Eds.), *Rapid Cycle Real-Time PCR* (pp. 21–34). Berlin: Springer-Verlag.

1428 Roussel, M.R., and Zhu, R. (2006) Validation of an algorithm for delay stochastic simulation of
1429 transcription and translation in prokaryotic gene expression. *Phys Biol* 3(4), 274–284.

1430 Rouvière, P.E., De Las Peñas, A., J. Mecsas, Lu, C.Z., Rudd, K.E., and Gross, C.A. (1995) rpoE,
1431 the gene encoding the second heat-shock sigma factor, sigma E, in *Escherichia coli*.
1432 *EMBO J* 14(5), 1032–1042.

1433 Rovinskiy, N., Agbleke, A.A., Chesnokova, O., Pang, Z., Higgins, N.P (2012) Rates of gyrase
1434 supercoiling and transcription elongation control supercoil density in a bacterial chromo-
1435 some. *PLoS Genet* 8(8), e1002845.

1436 Saecker, R.M., Record, M.T. and deHaseth, P.L. (2011) Mechanism of Bacterial Transcription
1437 Initiation: RNA Polymerase - Promoter Binding, Isomerization to Initiation-Competent
1438 Open Complexes, and Initiation of RNA Synthesis. *J Mol Biol* 412(5), 754–771.

1439 Samul, R., and Leng, F. (2007) Transcription-coupled Hypernegative Supercoiling of Plasmid
1440 DNA by T7 RNA Polymerase in *Escherichia coli* Topoisomerase I-Deficient Strains. *J*
1441 *Mol Biol* 374(4), 925–935.

1442 Santinha, J., Martins, L., Häkkinen, A., Lloyd-Price, J., Oliveira, S.M.D., Gupta, A., Annala, T.,
1443 Mora, A., Ribeiro, A.S., and Fonseca, J.M. (2015). iCellFusion: Tool for Fusion and
1444 Analysis of Live-Cell Images from Time-Lapse Multimodal Microscopy. In: Biomedical
1445 Image Analysis and Mining Techniques for Improved Health Outcomes. IGI Global. pp.
1446 71–99.

1447 Stuger, R., Woldringh, C.L., Van der Weijden, C.C., Vischer, N.O.E., Bakker, B.M., van Span-
1448 ning, R.J.M., Snoep, J.L., and Weterhoff, H. V. (2002) DNA supercoiling by gyrase is
1449 linked to nucleoid compaction. *Mol Biol Rep* 29(1-2), 79–82.

1450 Taniguchi, Y., Choi, P.J., Li, G.W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S.
1451 (2010) Quantifying *E. coli* proteome and transcriptome with single-molecule sensitivity in
1452 single cells. *Science* 329(5991), 533–538.

1453 Tran, H., Oliveira, S.M.D., Goncalves, N.S.M., and Ribeiro, A.S. (2015) Kinetics of the cellular
1454 intake of a gene expression inducer at high concentrations. *Mol Biosyst* 11(9): 2579–2587.

1455 Travers, A., and Muskhelishvili, G. (2005) DNA supercoiling - a global transcriptional regulator
1456 for enterobacterial growth? *Nat Rev Microbiol* 3, 157–169.

1457 Ueguchi, C., and Mizuno, T. (1993) The *Escherichia coli* nucleoid protein H-NS functions di-
1458 rectly as a transcriptional repressor. *EMBO J* 12(3), 1039–1046.

1459 Wang, J.C. (1996) DNA Topoisomerases. *Annu Rev Biochem* 65, 635–692.

1460 Wery, M., Woldringh, C.L., and Rouviere-Yaniv, J. (2001) HU-GFP and DAPI co-localize on
1461 the *Escherichia coli* nucleoid. *Biochimie* 83(2), 193–200.

1462 Whelan, J.A., Russell, N.B., and Whelan, M.A. (2003) A method for the absolute quantification
1463 of cDNA using real-time PCR. *J Immunol Methods* 278(1-2), 261–269.

1464 Wilson, R.H., Morton, S.K., Deiderick, H., Gerth, M.L., Paul, H.A., Gerber, I., Patel, A., Elling-
1465 ton, A.D., Hunnicke-Smith, S.P., and Patrick, W.M. (2013) Engineered DNA ligases with
1466 improved activities in vitro. *Protein Eng Des Sel* 26(7), 471–478.

1467 Yamanaka, K. (1999) Cold Shock Response in *Escherichia coli*. *J Mol Microbiol Biotechnol*
1468 1(2), 193–202.

1469

1470 Tables

1471 **Table 1. Number of cells observed, mean, and squared coefficient of variation (CV²)**
1472 **of the absolute integer-valued RNA numbers per cell for the chromosome-integrated**
1473 **and the plasmid-integrated constructs, when induced by 1 mM IPTG.** Cells are in-
1474 duced and kept at 30 °C, 27 °C, 23 °C, 20 °C, 16 °C and 10 °C for 60 minutes prior to the
1475 acquisition of the results. Results are obtained from 3 biological repeats. Since these exhib-
1476 ited no statistically significant differences, the results presented here are composed of the
1477 data from the 3 biological replicates.

Condition	No. cells	Mean integer-valued RNA no. per cell	CV ²
Chromosome construct			
30 °C	645	2.08	2.60
27 °C	632	2.00	1.93
23 °C	668	1.74	2.30
20 °C	646	0.59	7.17
16 °C	668	0.22	15.81
10 °C	648	0.25	16.12
Plasmid construct			
30 °C	675	2.86	1.06
27 °C	654	2.46	1.42
23 °C	665	1.63	2.73
20 °C	660	1.61	3.08
16 °C	663	1.50	2.99
10 °C	676	1.35	3.46

1478 **Table 2. Relative mean duration of the rate limiting steps in transcription initiation at**
1479 **30 °C and 10 °C.** Shown are the mean durations, relative to the mean time-length of the
1480 intervals between transcription events (Δt), of the rate-limiting steps prior ($\frac{t_{prior}}{\Delta t}$) and after
1481 ($\frac{t_{after}}{\Delta t}$) commitment to open complex formation for the chromosome-integrated and the
1482 plasmid-integrated constructs.

Condition	$\frac{t_{prior}}{\Delta t}$	$\frac{t_{after}}{\Delta t}$
30 °C		
Chromosome construct	0.09	0.91
Plasmid construct	0.08	0.92
10 °C		
Chromosome construct	0.27	0.73
Plasmid construct	0.02	0.98

1483

1484 **Figure Legends**

1485 **Figure 1. Induction curves, measured by microscopy imaging and single RNA tagging**
1486 **by MS2-GFP, of the target promoter P_{LacO3O1} when integrated into the chromosome**
1487 **(light grey) and into a single-copy F-plasmid (dark gray) (*E. coli* strain BW25993).**
1488 Shown are the RNA numbers (relative to the reference case, 1 mM IPTG) in individual
1489 cells of the two constructs, 1 hour after induction at 30 °C. Data presented as relative mean
1490 to the reference case with 90% confidence intervals obtained from a two-tailed Student’s t-
1491 test. Sample size per condition, as IPTG is increased, is (chromosome) 665, 655, 675, 670,
1492 660 and 645 cells, and (plasmid) 670, 670, 665, 655, 655, 675 cells. Also shown is the ratio
1493 between the integer-valued mean RNA numbers per cell between cells with the target gene
1494 chromosome-integrated and on a single-copy plasmid. Absolute integer-valued RNA num-
1495 bers per cell in each condition can be obtained from the absolute integer-valued RNA
1496 numbers per cell for 1mM IPTG shown in Table 1 along with the relative values shown
1497 here. Results are obtained from 3 biological repeats. Since these exhibited no statistically
1498 significant differences, the results presented here are composed of the data from the 3 bio-
1499 logical replicates.

1500 **Figure 2. Mean RNA numbers in individual cells, relative to the last time moment, as**
1501 **a function of temperature, measured by microscopy with single RNA tagging by MS2-**
1502 **GFP, when P_{LacO3O1} is integrated into the chromosome (light grey) and in a single-**
1503 **copy F-plasmid (dark gray). (A) Cells are at 10 °C. (B) Cells are at 30 °C. Data presented**
1504 **as relative mean to the reference case with 90% confidence intervals obtained from a two-**
1505 **tailed Student’s t-test. Sample size per condition, as time progresses is: A) Chromosome at**
1506 **10 °C (610, 611, 615, 610, 609, 602 and 605 cells), and Plasmid at 10 °C (615, 610, 610,**
1507 **612, 608, 606 and 606 cells); B) Chromosome at 30 °C (604, 615, 610, 610, 605, 608 and**
1508 **605 cells), and Plasmid at 30 °C (610, 615, 606, 615, 610, 613 and 609 cells). For each**
1509 **time point, new cells were taken from the original culture. Results are obtained from 3 bio-**
1510 **logical repeats. Since these exhibited no statistically significant differences, the results pre-**
1511 **sented here are composed of the data from the 3 biological replicates. Finally, at t = 0 min,**
1512 **the mean absolute number of RNA molecules per cell was (A) 0.1 for chromosome and 0.9**
1513 **for plasmid, and (B) 0.2 for the chromosome and 0.9 for the plasmid.**

Figure 3. Mean RNA numbers in individual cells, relative to the last time moment, as a function of temperature and gyrases and topoisomerases I inhibitors, measured by microscopy with single RNA tagging by MS2-GFP, when $P_{LacO3O1}$ is integrated into the chromosome (light grey) and in a single-copy F-plasmid (dark gray). (A) Cells at 30 °C and subject to Novobiocin. (B) Cells at 10 °C and subject to Novobiocin. (C) Cells at 30 °C and subject to Topotecan. (D) Cells at 10 °C and subject to Topotecan. Data presented as relative mean to the reference case with 90% confidence intervals obtained from a two-tailed Student's t-test. Sample size per condition, as time progresses is: A) Chromosome, 30 °C, Novobiocin (608, 606, 605, 610, 613, 610 and 615 cells) and Plasmid, 30 °C, Novobiocin (615, 605, 613, 610, 610, 601 and 602 cells); B) Chromosome, 10 °C, Novobiocin (615, 610, 615, 625, 605, 620 and 620), and Plasmid, 10 °C, Novobiocin (615, 620, 615, 610, 620, 620 and 615 cells); C) Chromosome, 30 °C, Topotecan (615, 610, 612, 610, 605, 603 and 610 cells) and Plasmid, 30 °C, Topotecan (662, 623, 626, 606, 643, 659 and 647 cells) and, finally, D) Chromosome, 10 °C, Topotecan (620, 610, 620, 610, 610, 610 and 615 cells) and Plasmid, 10 °C, Topotecan (679, 629, 649, 645, 642, 601 and 632 cells). For each time point, new cells were taken from the original culture. Results are obtained from 3 biological repeats. Since these exhibited no statistically significant differences, the results presented here are composed of the data from the 3 biological replicates. In all cases, Novobiocin or Topotecan was added to the culture at the same time as the inducer of the target gene, IPTG. Finally, at $t = 0$ min, the mean absolute number of RNA molecules per cell was (A) 0.3 for chromosome and 0.9 for plasmid, (B) 0.1 for chromosome and 0.9 for plasmid, (C) 0.1 for chromosome and 0.9 for plasmid, and (D) 0.1 for chromosome and 0.9 for plasmid.

Figure 4. Mean RNA numbers in individual cells at 30 °C subject to DNP treatment relative to the last time moment, measured by microscopy with single RNA tagging by MS2-GFP, when $P_{LacO3O1}$ is integrated into the chromosome (light grey). Data presented as relative mean to the reference case with 90% confidence intervals obtained from a two-tailed Student's t-test. Sample size per condition, as time progresses is 601, 610, 601, 605, 610, 605 and 608 cells. For each time point, new cells were taken from the original culture. Results are obtained from 3 biological repeats. Since these exhibited no statistically significant differences, the results presented here are composed of the data from the 3 biological replicates. Finally, at $t = 0$ min, the mean absolute number of RNA molecules per cell was 0.2.

Figure 5. Mean RNA numbers in individual cells, relative to the last time moment, as a function of temperature shifts, measured by microscopy with single RNA tagging by MS2-GFP, when $P_{LacO3O1}$ is integrated into the chromosome (light grey). In these measurements, first, the cells are kept at 30 °C for 30 minutes. Next, they are kept at 30 °C and measurements are conducted (starting point of the measurements is defined as moment ' $t = 0$ '). 30 minutes after starting the measurements, the temperature is changed to 10 °C and then kept constant until reaching moment 120 min. Then it is altered again to 30 °C and kept constant until the end of the measurements. Data presented as relative mean to the reference case with 90% confidence intervals obtained from a two-tailed Student's t-test. Sample size per condition, as time progresses is 600, 601, 603, 615, 613, 610, 603, 614, 611, 608, 607 cells. For each time point, new cells were taken from the original culture.

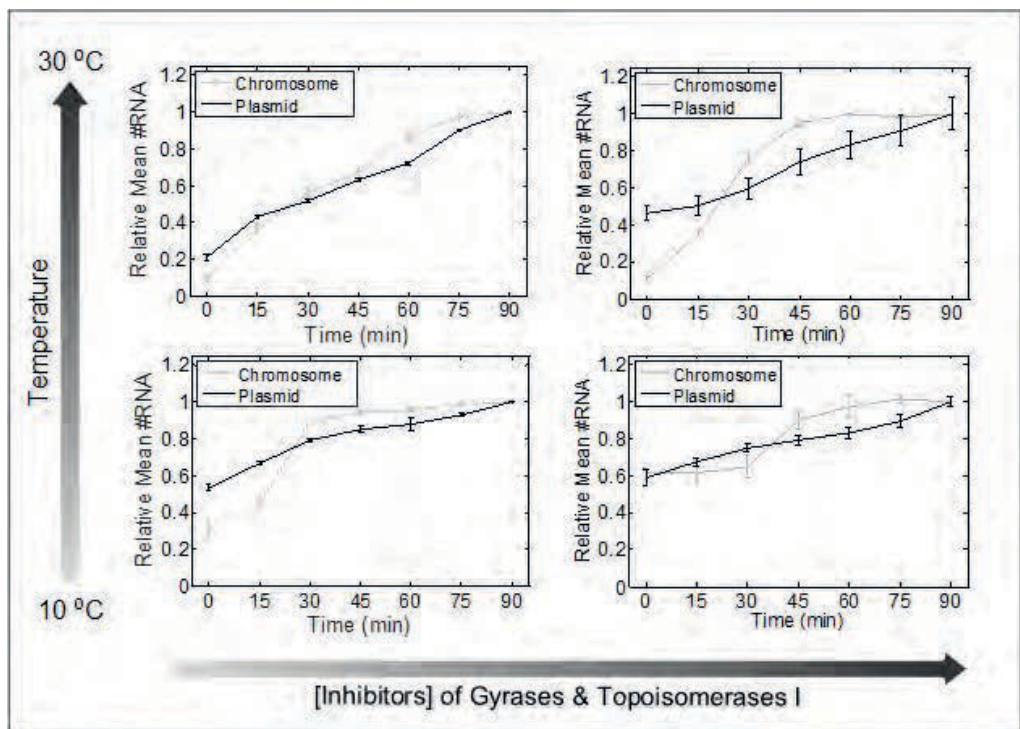
1558 Results are obtained from 3 biological repeats. Since these exhibited no statistically signif-
1559 icant differences, the results presented here are composed of the data from the 3 biological
1560 replicates. Finally, at $t = 0$ min, the mean absolute number of RNA molecules per cell was
1561 1.3.

1562 **Figure 6. Expected mean RNA numbers in individual cells, relative to the last time**
1563 **moment, from simulations of a stochastic transcription model.** The model assumes that
1564 the promoter is integrated into the chromosome for differing values of the rate of promoter
1565 escape from the supercoiled state (solid lines increasingly darker as k_{unlock} increases). Also
1566 shown are the measurements of mean RNA numbers from cells at 10 °C (dashed black
1567 line) and cells at 30 °C (dashed grey line). Data presented as relative mean to the reference
1568 case with 90% confidence intervals obtained from a two-tailed Student's t-test.

Abbreviated Summary

Temperature shifts alter the gene expression profile of *Escherichia coli*. We show that, in P_{LacO301} , these changes in transcription kinetics differ when chromosome-integrated and when on a single-copy plasmid, being weaker and noisier in the former, particularly at critically low temperatures. Measurements at 10 and 30 °C, in the presence/absence of DNP and topoisomerases inhibitors, indicate that these differences at low temperatures are due to longer-lasting super-coiled states of the chromosome-integrated promoter.

Graphical Abstract



Manuscript Figures

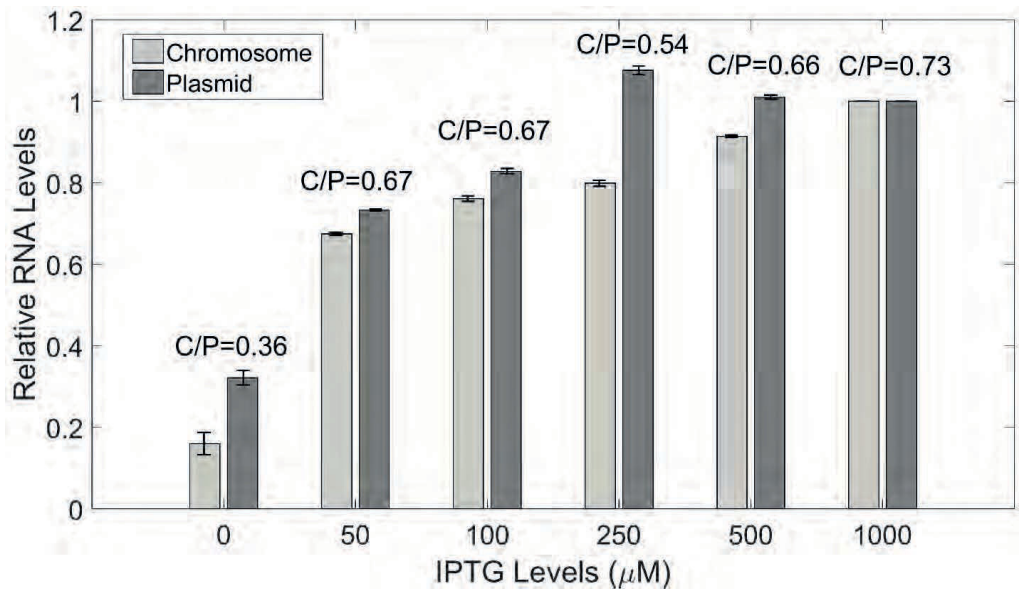


Figure 1.

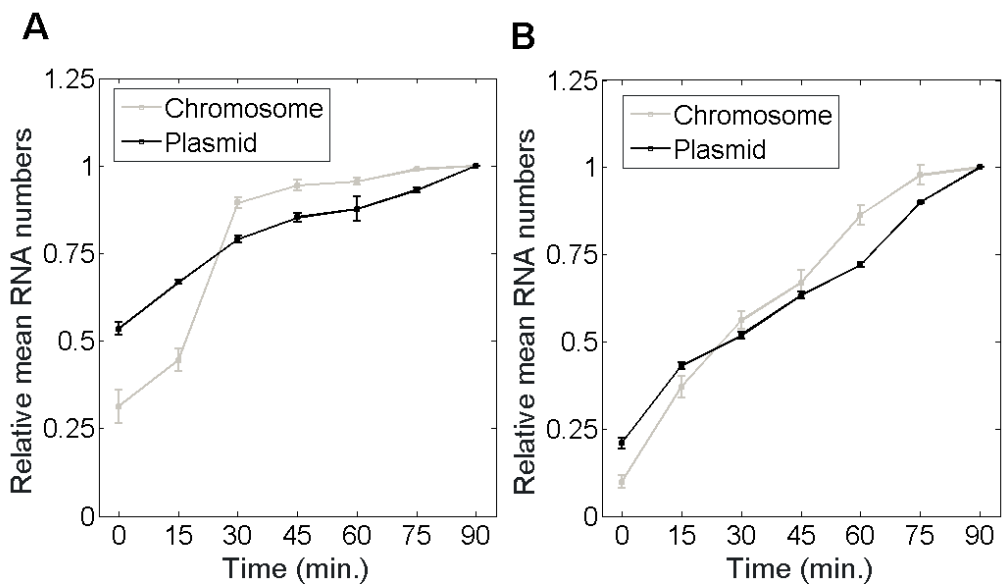


Figure 2.

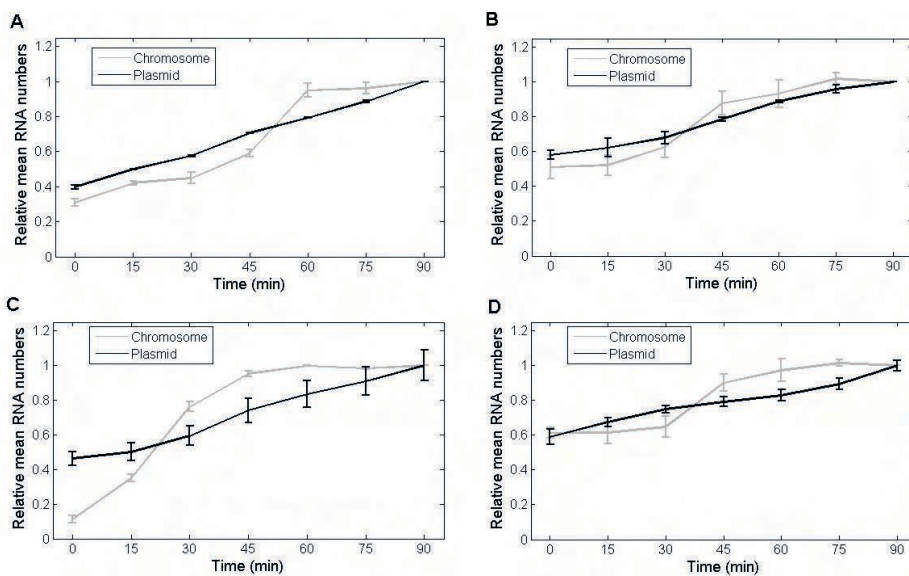


Figure 3.

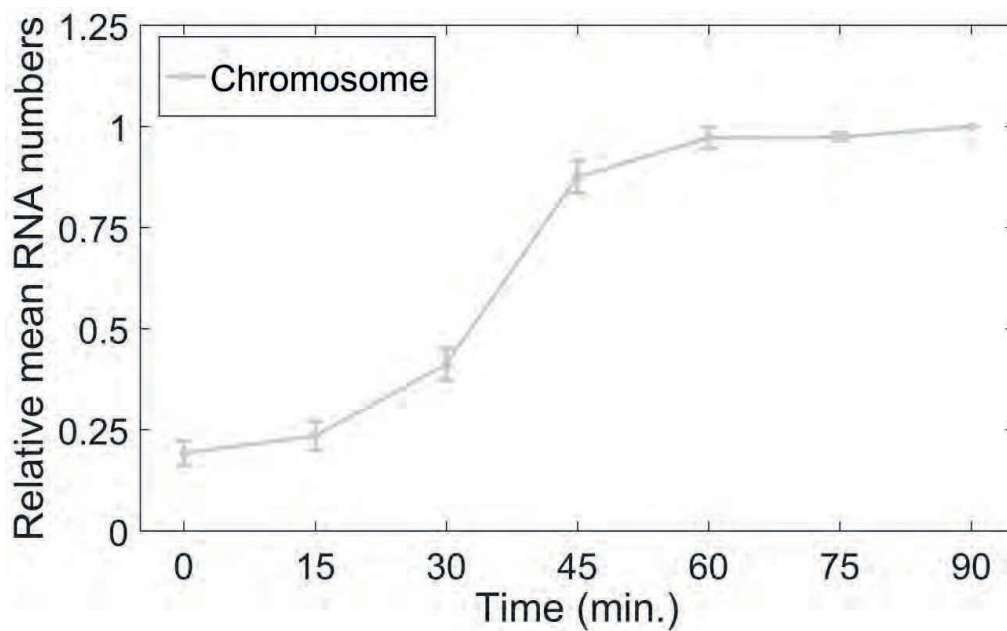


Figure 4.

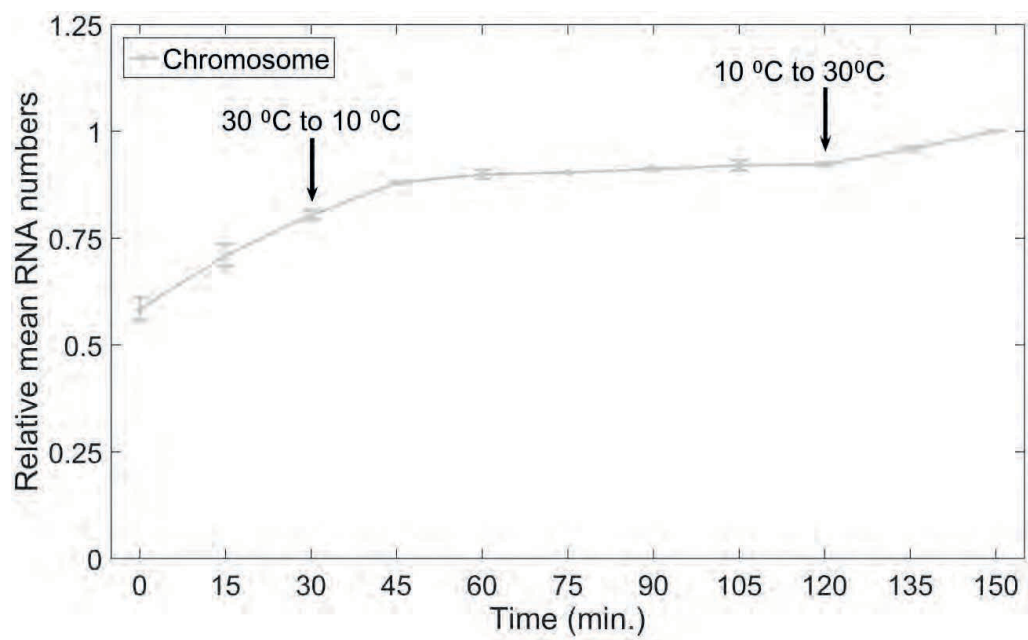


Figure 5.

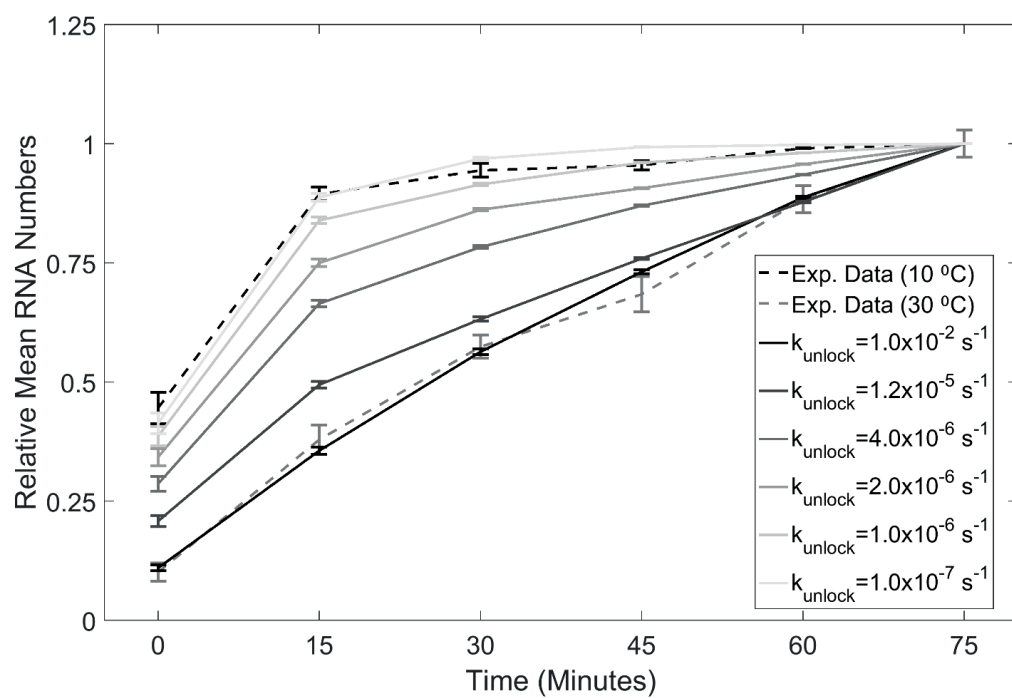


Figure 6.

Supporting Material for ‘Chromosome and plasmid-borne *lacO301* promoters differ in sensitivity to critically low temperatures’

Nadia S.M. Goncalves^{1,*}, Samuel M.D. Oliveira^{1,*}, Leonardo Martins^{1,2}, Ramakanth Neeli-Venkata¹, Jan Reyelt³, Jose M. Fonseca², Jason Lloyd-Price^{4,5}, Harald Kranz³, and Andre S. Ribeiro^{1,2,+}

¹ Laboratory of Biosystem Dynamics, Multi-Scaled Biodata Analysis and Modelling Research Community and BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, PO Box 553, 33101 Tampere, Finland.

² CA3 CTS/UNINOVA. Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal.

³ Gene Bridges, Im Neuenheimer Feld 584, 69120 Heidelberg, Germany.

⁴ Biostatistics Department, Harvard T. H. Chan School of Public Health, Boston, MA 02115, USA

⁵ Infectious Disease and Microbiome Program, Broad Institute, Cambridge, MA 02142, USA

⁺ Corresponding author. Email: andre.ribeiro@tut.fi, Tel: +358408490736, Fax: +358331154989

*Equal contributions

Supporting Figures

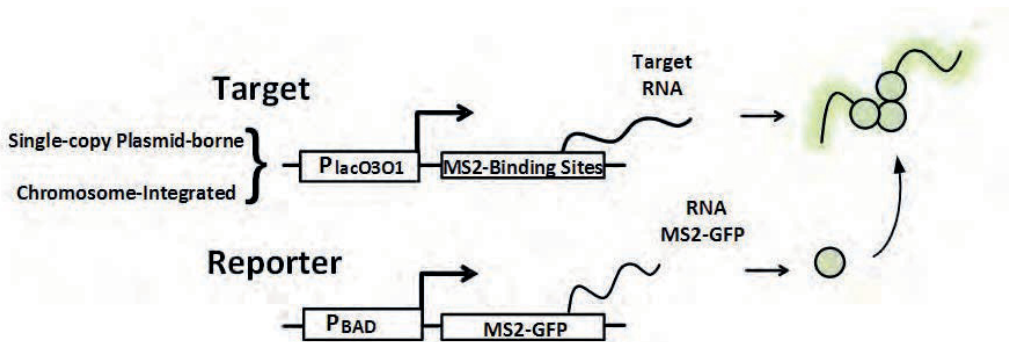


Figure S1. Single-RNA detection system schematic. Single-RNA detection system using MS2-GFP reporter proteins, whose production is controlled by P_{BAD} , when applied to an RNA coding for multiple MS2 binding sites (BS), whose production is controlled by $P_{LacO3O1}$ integrated into the chromosome, and when applied to an RNA with multiple BS for MS2-GFP whose production is controlled by $P_{LacO3O1}$ integrated into a single-copy plasmid. In both systems, when individual target RNA molecules are produced, they are rapidly tagged by multiple MS2-GFP proteins produced by the reporter plasmid, making each RNA target for MS2-GFP visible under the fluorescence microscope as a fluorescent ‘spot’.



Figure S2. Genetic constructs. (A) Single-copy plasmid-borne gene (expected length of 4019 base pairs). MS2-BS cassette (depicted in yellow) under the control of $P_{LacO3O1}$ promoter in a single-copy F-Plasmid in *E. coli* strain BW25993, followed by a Rho-Independent transcription termination site. (B) Chromosome-integrated gene (expected length of 5116 base pairs). MS2-BS cassette (depicted in yellow) under the control of $P_{LacO3O1}$ promoter in the *lac* locus of *E. coli* strain BW25993 ($\Delta lacZ:MS2-BS$), followed by the native *lacY* and *lacA* genes (depicted in purple), and the native Rho-Independent transcription termination site (depicted in white boxes). Constructs were confirmed by sequencing. Expected lengths are obtained from the difference between the position of the RBS (Ribosome Binding Site) and the downstream terminator sequences of the gene being transcribed. As 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 intake of IPTG. (Inset) The inset image in between the images of the two constructs shows in detail the $P_{LacO3O1}$ promoter with functional domains, which is identical in both A and B constructs. It is in this identical region for both constructs that is located the operator site O3 (operator sites depicted in blue), followed by the *lac* promoter's RNAP binding regions (starting from positions -10 and -35), the transcription start site (TSS, at +1 position), and the operator site O1. Note the mCherry sequence in between the promoter and the sequence coding for the MS2-BS in both constructs (depicted in red). Finally, note that in the plasmid construct there is a terminator upstream of the TSS, 27 nucleotides long, located 9 nucleotides downstream of the CmR gene (not represented in the figure), so as to be similar to the chromosome-integrated construct, where there is an upstream transcriptional terminator provided by the *lacI* gene.

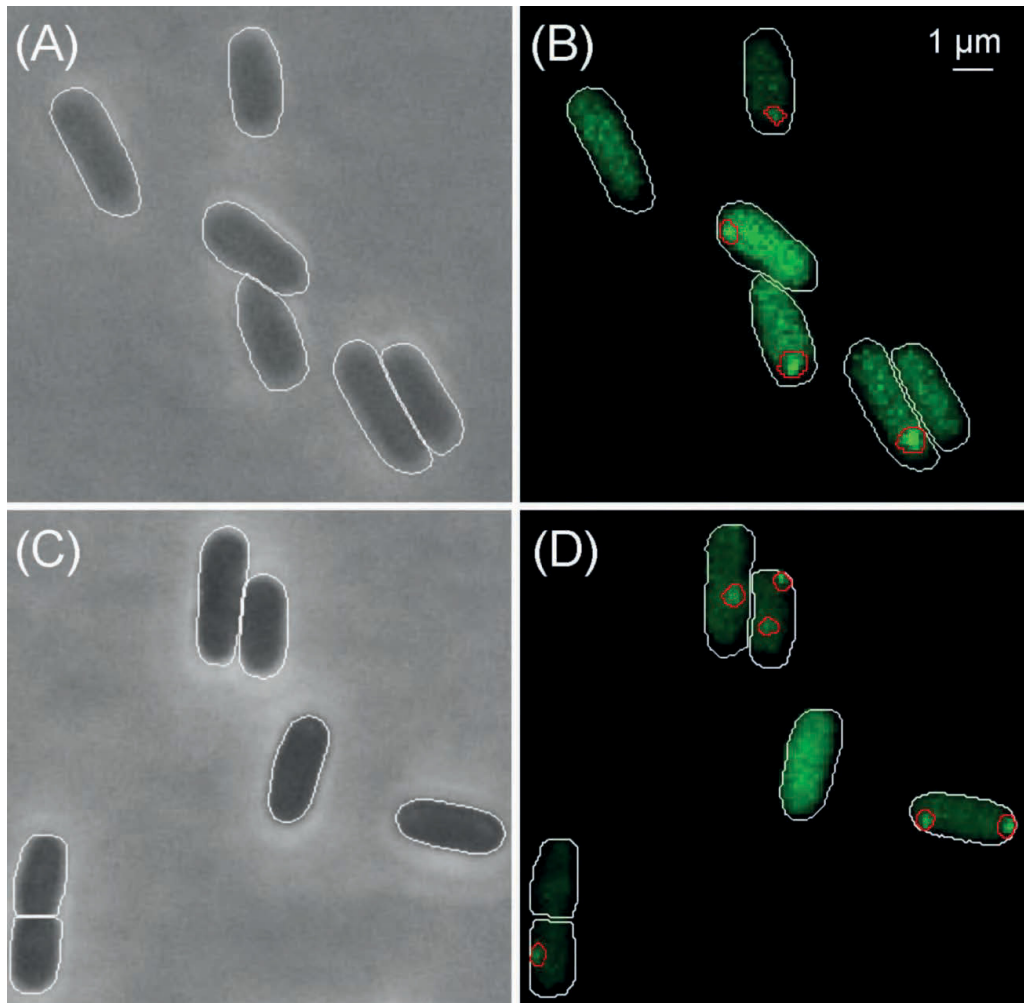


Figure S3. Example Microscopy Images. (A) Phase contrast image of cells at 30 °C along with results of semi-automatic cell segmentation. (B) Confocal microscopy image of the same cells at 30 °C with MS2-GFP tagged RNA spots, detected and segmented by the automatic spot detection method. (C) Phase contrast image of cells at 10 °C along with results of semi-automatic cell segmentation. (D) Confocal microscopy image of the same cells at 10 °C with MS2-GFP tagged RNA spots, detected and segmented by the automatic spot detection method. Note the scale bar in (B), which applies to all figures.

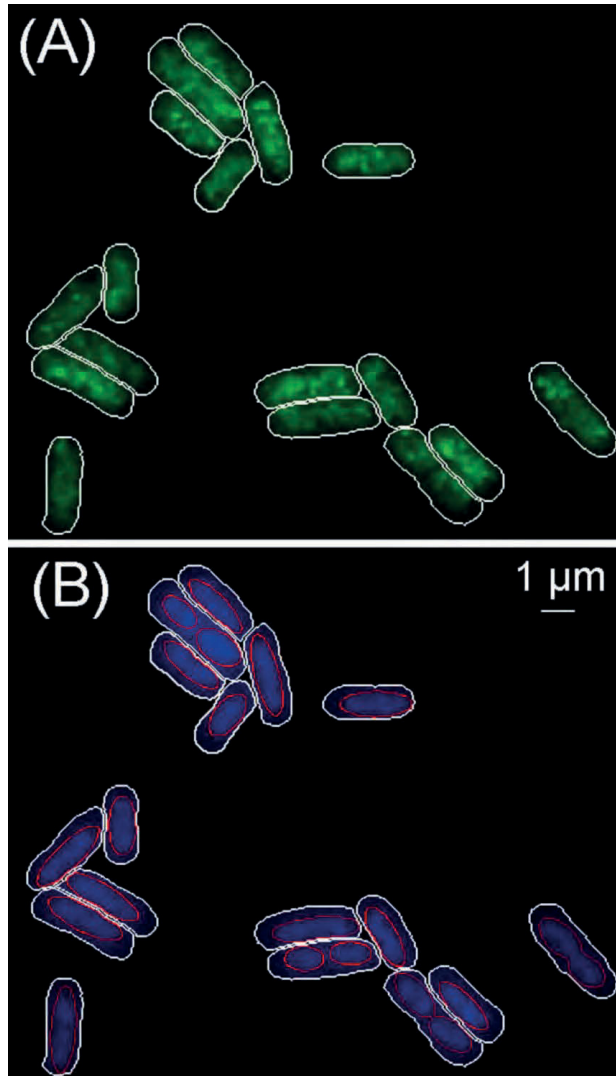


Figure S4. Example images of GFP and DAPI fluorescence distribution within cells. (A) Confocal microscopy image of cells with GFP-tagged RNA Polymerases. (B) Epifluorescence microscopy image of the same cells with DAPI-stained nucleoids. Note the scale bar in (B), which applies to both figures.

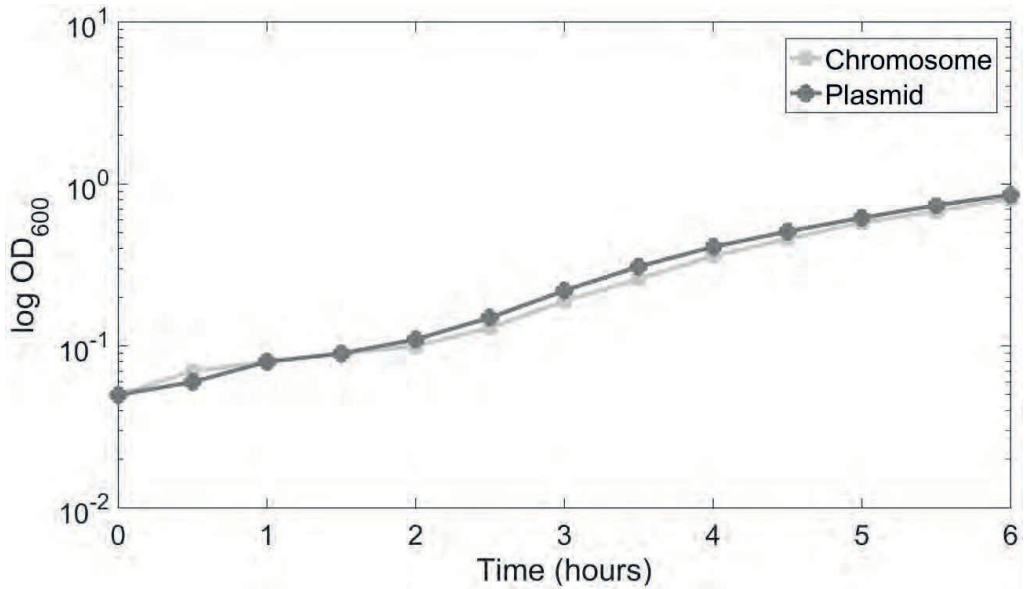


Figure S5. OD curves of cell populations in fresh media at 30 °C. The original *E. coli* strain where both the plasmid and chromosome constructs were inserted is BW25993. From a -80 °C glycerol stock, cells with the target and reporter genes were placed in LB medium agar plates with 34 $\mu\text{g ml}^{-1}$ Chloramphenicol and 35 $\mu\text{g ml}^{-1}$ Kanamycin (Sigma-Aldrich, USA) and incubated overnight at 37 °C. From these plates, a single colony was picked and cultured overnight at 30 °C, with agitation (250 rpm) and aeration, in LB medium supplemented with the appropriate concentration of antibiotics. From the overnight culture, cells were diluted to an initial optical density (OD₆₀₀) of 0.05, in fresh M9 medium supplemented with the appropriate antibiotics and 0.4 % of Glycerol (Sigma-Aldrich, USA), and were incubated at 37 °C until reaching an OD₆₀₀ of 0.3 (for the first 3 hours). They were then placed at 30 °C for 3 hours. The OD₆₀₀ was then measured every 30 minutes for 6 hours.

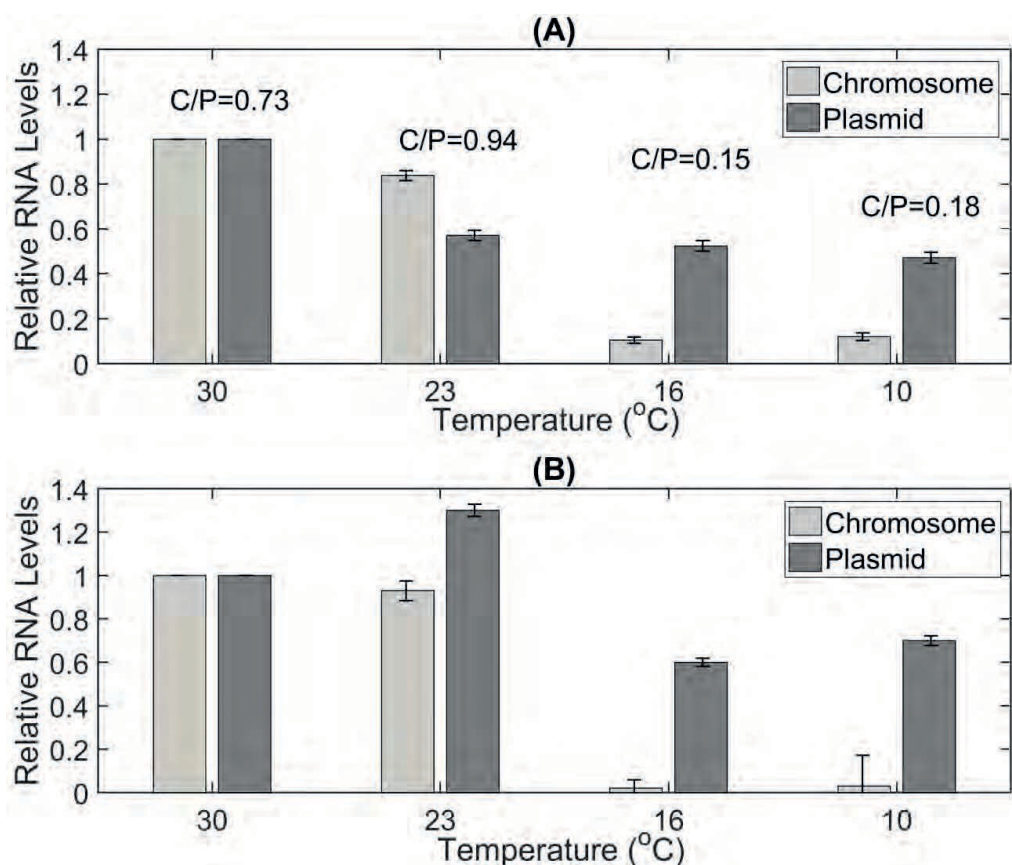


Figure S6. Mean relative RNA numbers in individual cells. Mean relative RNA numbers in individual cells subject to full induction (1 mM IPTG) at different temperatures relative to the control (30°C) as measured by (A) microscopy and by (B) RT-PCR. Cells carrying the chromosome (light gray) and plasmid (dark gray) construct were induced 1 hour prior to the measurements. RT-PCR (3 technical replicates) and microscopy measurements are relative to the 30 °C condition (thus removing the error from that point). Data presented as relative mean to the reference case with 90% confidence intervals obtained from a two-tailed Student's t-test. Also shown is the ratio (C/P) between the integer-valued mean RNA numbers per cell between cells with the target gene chromosomally-integrated (C) and on a single-copy plasmid (P).

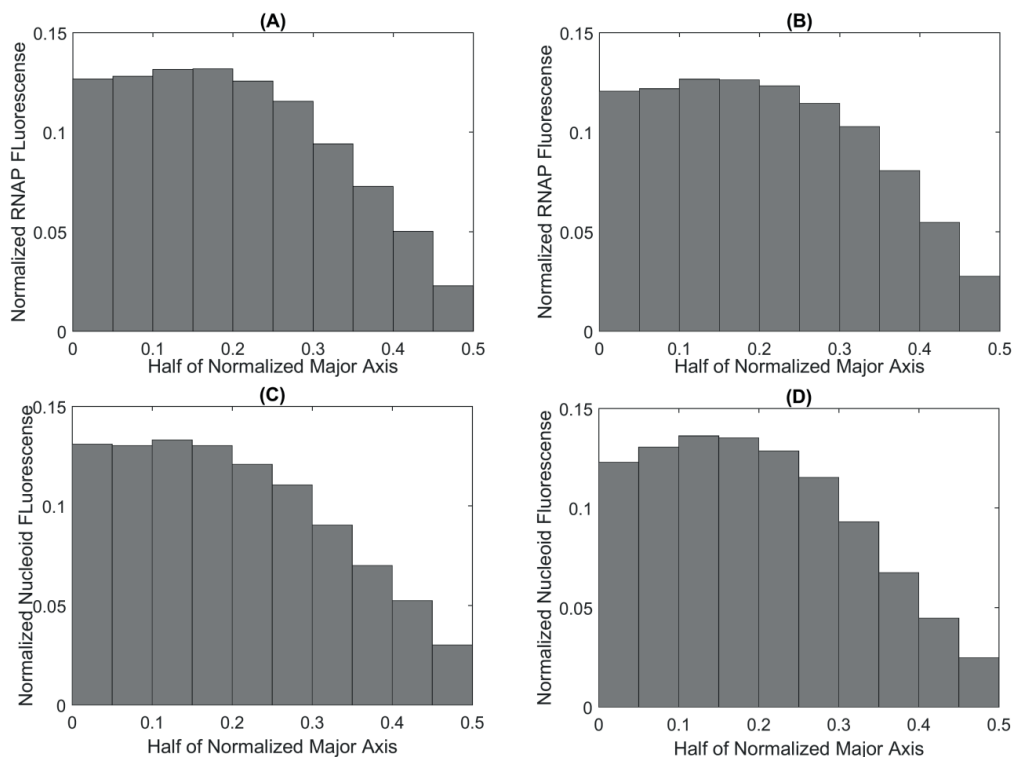


Figure S7. RNAP and nucleoid fluorescence along the major cell axis. (A) RNAP fluorescence along the major cell axis (binned) as normalized by the total mean fluorescence of the cells (RL1314 strain). Measurements at 30 °C from 614 cells. (B) RNAP fluorescence along the major cell axis (binned) as normalized by the total mean fluorescence of the cells (RL1314 strain). Measurements at 10 °C from 613 cells. A two-sample KS test comparing the spatial distributions at 10 °C and 30 °C fails to reject that they are from the same distribution (P value = 0.98). (C) Normalized average nucleoid fluorescence intensity distribution along the normalized major axis of the cells as measured by DAPI staining. Measurements are from 614 cells at 30 °C. (D) Normalized average nucleoid fluorescence intensity distribution along the normalized major axis of the cells (RL1314 strain). Measurements are from 613 cells at 10 °C. A two-sample KS test comparing the spatial distributions at 10 °C and 30 °C fails to reject that they are from the same distribution (P value = 0.99). In all figures, in the x axis, '0' corresponds to the cell center, while '0.5' corresponds to both extremities (cells folded in half, with undefined poles). Cells were fixed with formaldehyde prior to imaging.

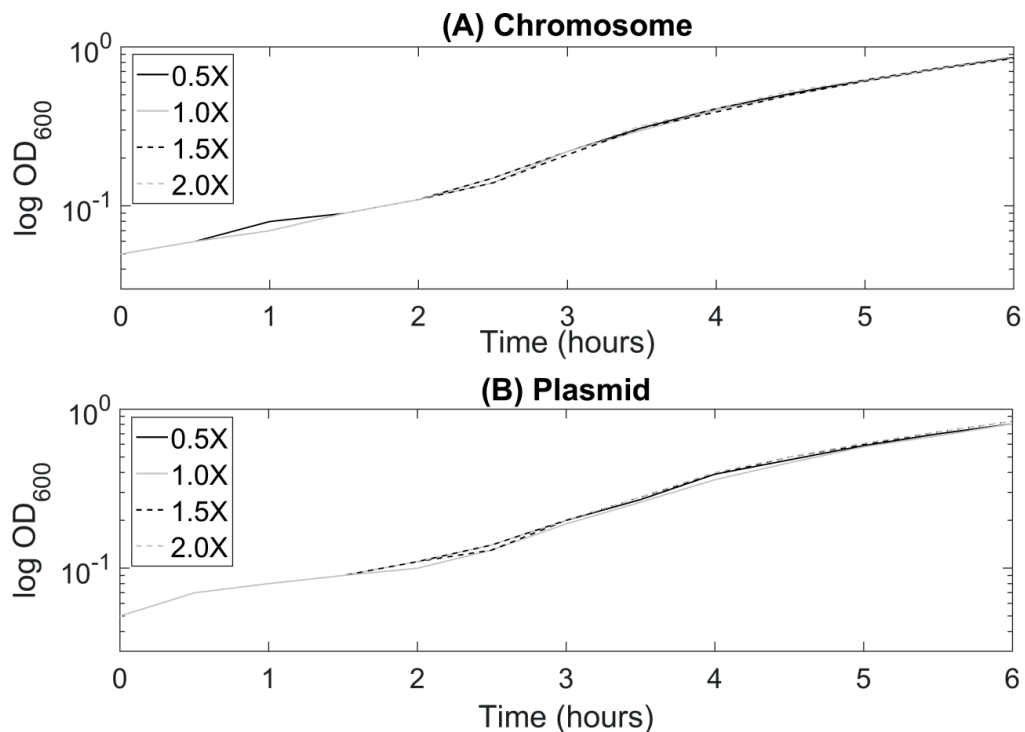


Figure S8. Cell growth curves. Growth curves (OD_{600} , measured by an Ultrospec 10 Cell Density Meter, Amersham Bioscience) of cells carrying the (A) chromosome and (B) plasmid constructs at 30 °C. Cells grown overnight in LB medium at 30 °C with aeration of 250 rpm, and diluted into fresh 0.5X, 1X, 1.5X and 2X medium to an initial OD_{600} of 0.05. Next, they were incubated at 37 °C until reaching the mid-log phase (first 3 hours), and placed at 30 °C for the remaining 3 hours. The OD_{600} was measured every 30 minutes.

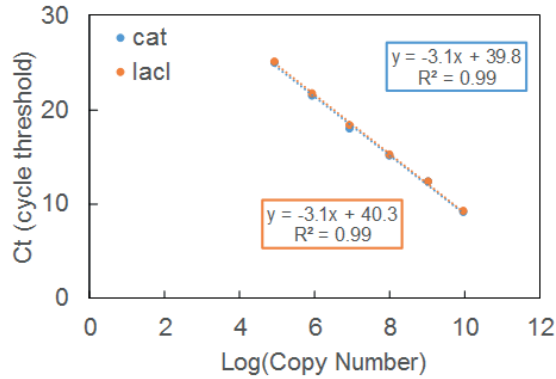


Figure S9. Standard curves for the *cat* and *lacI* genes. The curve was constructed using 10-fold serial dilutions of pCA24N-ligase, (ranging from 10^4 to 10^9 copies μl^{-1}). Each dilution was amplified by RT-qPCR, in triplicates, using primer sets specific for *cat* and *lacI* genes. For each gene, the C_T values were plotted against the logarithm of their known initial copy number, and a standard curve was generated by a linear regression through these values.

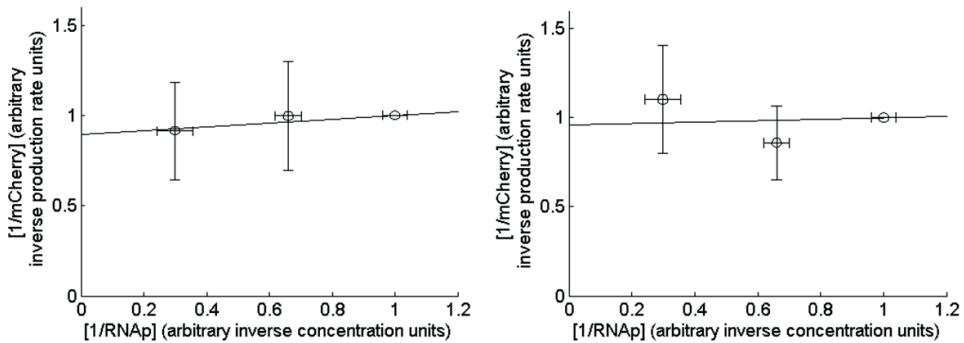


Figure S10. Lineweaver-Burk plots of both constructs at 30°C. (Left) chromosome-integrated construct. (Right) plasmid borne construct. The inverse of the relative production rate of mRNA from the P_{LacO3O1} promoter is plotted against the inverse of relative difference of RNAP concentrations for cells grown in three different media (1.0x, 1.5x, 2.0x M9-Gly). RNAP concentrations are presented relative to the RNAP concentration at 1.0x, from the difference of RNAP concentration at 0.5x (Table S7). Relative production rates were measured by RT-PCR with three technical replicates for each condition relative to 1.0x condition (Table S9). The linear relationship between the points for each relative production rate and its corresponding relative difference of RNAP concentration (circles) is visible in both constructs. A likelihood ratio test

was performed to determine whether the small deviations from linearity are statistically significant. In no construct was linearity rejected ($P > 0.25$ in both cases). Standard uncertainties are also shown (horizontal and vertical error bars in each data point).

Supporting Tables

Table S1. Mean number of tagged RNAs produced under the control of $P_{LacO3O1}$ in cells lacking the target gene for MS2-GFP, and non-induced cells carrying the target gene (chromosome and plasmid constructs). Measurements were obtained 1 hour after the start of incubation of the cells in liquid culture at 30°C. Shown are the number of cells observed and their mean integer-valued RNA numbers per cell. In cells lacking the target gene, there is no coding ability of the RNA target for MS2-GFP and thus, in these, any ‘detected’ RNA is a false positive due to MS2-GFP. Measurements were conducted at 30 °C and at 10 °C.

Condition	No. cells	Absolute mean RNA no. per cell
No target gene (30 °C)	612	0.015
No target gene (10 °C)	630	0.022
Plasmid construct, no induction (30 °C)	670	0.919
Plasmid construct, no induction (10 °C)	605	0.466
Chromosome construct, no induction (30 °C)	665	0.332
Chromosome construct, no induction (10 °C)	615	0.135

Table S2. Absolute mean half-life times of RNA molecules tagged with MS2-GFP in the two strains and temperature conditions, extracted from 1-hour long time series, with images taken every minute. Mean half-lives were obtained by fitting the intensity of each RNA over time with a decaying exponential function and then inferring the degradation rate of the RNA fluorescence intensity. Note that no tagged RNA was observed to ‘disappear’ during any measurement.

Construct	Temperature (°C)	No. RNAs	Mean half-life (min)
Chromosome	30	10	151.3
Plasmid	30	10	148.0
Chromosome	10	10	121.6
Plasmid	10	40	120.2

Table S3. P values of the KS tests comparing pairs of distributions of integer-valued RNA numbers per cell. Each distribution corresponds to a given induction level for the the induction of the target gene of the plasmid construct (top table) and the chromosome construct (bottom table). In these tests, for P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. These tables are related to Figure 1 in main manuscript.

Plasmid					
	50 μ M	100 μ M	250 μ M	500 μ M	1000 μ M
0 μ M	5.6×10^{-27}	1.7×10^{-27}	2.9×10^{-28}	7.8×10^{-33}	5.5×10^{-42}
50 μ M	-	0.2276	8.0×10^{-7}	3.5×10^{-6}	7.9×10^{-7}
100 μ M	-	-	0.001	0.018	3.4×10^{-6}
250 μ M	-	-	--	0.7641	0.015
500 μ M	-	-	-	-	0.206

Chromosome					
	50 μ M	100 μ M	250 μ M	500 μ M	1000 μ M
0 μ M	2.7×10^{-15}	3.1×10^{-13}	2.8×10^{-16}	2.1×10^{-16}	2.7×10^{-13}
50 μ M	-	0.956	0.906	0.040	0.190
100 μ M	-	-	0.893	0.225	0.730
250 μ M	-	-	-	0.190	0.845
500 μ M	-	-	-	-	0.917

Table S4. P values of the KS tests comparing the distributions of integer-valued RNA numbers per cell between temperature conditions for the plasmid construct and for the chromosome construct. In these tests, for P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. This table is related to Table 1 in main manuscript.

Condition	Plasmid	Chromosome
30°C vs 27°C	2×10^{-4}	3.4×10^{-7}
27°C vs 23°C	3.7×10^{-7}	1.7×10^{-4}
23°C vs 20°C	0.138	6.0×10^{-19}
20 °C vs 16 °C	1	0.008
16 °C vs 10°C	0.58	1.000

Table S5. P values of the KS tests comparing the distributions of integer-valued RNA numbers per cell between the two constructs at the various temperatures. In these tests, for P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. This table is related to Table 1 in main manuscript.

Condition	Plasmid vs Chromosome
30 °C	9.2×10^{-24}

27 °C	0.044
23 °C	0.177
20 °C	3.5×10^{-14}
16 °C	1.6×10^{-28}
10 °C	3.9×10^{-12}

Table S6. P values of the KS tests comparing the distributions of the DAPI fluorescence levels in individual cells as a function of the conditions (0.5X, 1.0X, 1.5X and 2.0X), in RL1314 cells at 30 °C and 10 °C. In these tests, for P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. Cells were fixed with formaldehyde prior to imaging.

30 °C			
Media Richness	0.5X	1.0X	1.5X
1.0X	0.655	-	-
1.5X	0.310	0.162	-
2.0X	0.697	0.964	0.182
10 °C			
Media Richness	0.5X	1.0X	1.5X
1.0X	0.029	-	-
1.5X	0.057	0.787	-
2.0X	0.001	0.697	0.511

Table S7. Number of cells observed, mean fluorescence from RNAP-GFP in individual cells (in arbitrary units, A.U.), absolute difference between these values and the 0.5X condition, ratio between these values and the 1X control condition, and inverse of this value. Data from RL1314 cells at 30 °C and at 10 °C. This table is related to Table 2 in main manuscript. Cells were fixed with formaldehyde prior to imaging.

Condition	No. cells	Mean RNAP fluorescence per cell (A.U.)	Absolute difference to 0.5X condition	Relative value to 1X condition	Inverse of the relative value to 1X condition
30 °C					
0.5X media	605	1.50±0.03	0	-	-
1.0X media	614	1.55±0.03	0.05	1	1
1.5X media	614	1.58±0.03	0.07	1.6	0.6
2.0X media	612	1.66±0.05	0.16	3.2	0.3
10 °C					
0.5X media	607	1.47±0.03	0	-	-
1.0X media	613	1.52±0.03	0.05	1	1
1.5X media	616	1.63±0.03	0.16	3.2	0.3
2.0X media	603	1.78±0.03	0.31	6.2	0.2

Table S8. P values of the KS tests comparing the distributions of RNAP fluorescence levels in individual cells as a function of the conditions (0.5X, 1.0X, 1.5X and 2.0X), in RL1314 cells at 30 °C and 10 °C. In these tests, for P values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. This table is related to Table 2 in main manuscript.

30 °C			
Media Richness	0.5X	1.0X	1.5X
1.0X	0.009	-	-
1.5X	6.4×10^{-4}	0.728	-
2.0X	9.0×10^{-5}	0.313	0.253
10 °C			
Media Richness	0.5X	1.0X	1.5X
1.0X	0.168	-	-
1.5X	0.001	0.038	-
2.0X	2.1×10^{-14}	4.4×10^{-10}	6.5×10^{-6}

Table S9. Inverse of the RNA production rates in the 1.5X and 2.0X conditions relative to the 1X condition. Data from the chromosome integrated and the plasmid integrated constructs when cells are at 30 °C and at 10 °C. This table is related to Table 2 in main manuscript.

Condition	1X	1.5X	2.0X
30 °C			
Plasmid construct	1	0.99	0.91
Chromosome construct	1	0.86	1.1
10 °C			
Plasmid construct	1	0.95	1.09
Chromosome construct	1	0.66	1.04

Table S10. Number of cells observed, along with the mean values of the absolute length (in μm) of the major and minor cell axes, and major and minor nucleoid axes at various temperatures. Also shown is the length of the major and minor axes of the nucleoid, relative to the cell major and minor axes lengths. Cells were fixed with formaldehyde prior to imaging.

Measurements	10 °C	16 °C	23 °C	30 °C
No. cells	613	613	615	614
Absolute Major Cell Axis	3.79	3.54	3.14	3.00
Absolute Minor Cell Axis	1.19	1.20	1.13	1.08
Absolute Major Nucleoid Axis	2.47	2.19	1.99	1.77
Absolute Minor Nucleoid Axis	0.84	0.85	0.88	0.78
Relative Major Nucleoid Axis	0.65	0.62	0.64	0.59

Relative Minor Nucleoid Axis	0.71	0.71	0.78	0.73
------------------------------	------	------	------	------

Table S11. Parameter values of the rate constants of transcription in model cells. The value of k_1 accounts for the expected RNA polymerase concentration in the cells. k_3 is set to infinite (∞), as it is much faster than all other rate constants.

Rate Constants	Parameter Value (s ⁻¹)	Reference
k_{ON}	0.011	(1)
k_{rep}	281	(1)
k_1	6469	(1)
k_{-1}	1	(1)
k_2	0.005	(1)
k_3	∞	(1)
k_{unlock}	See Figure 6	(17)

Table S12. Estimated plasmid copy number, using absolute quantification. Shown are the average and the standard deviation of the triplicates results obtained by RT-qPCR.

T (°C)	C _T		Copies (μl ⁻¹)		Absolute Plasmid Copy Number
	<i>cat</i>	<i>lacI</i>	<i>cat</i>	<i>lacI</i>	
10 °C	22.98 ± 0.03	23.34 ± 0.07	5.46 ± 0.02	5.47 ± 0.01	1.00 ± 0.00
30 °C	23.00 ± 0.13	23.68 ± 0.04	5.46 ± 0.01	5.36 ± 0.04	1.02 ± 0.01

Table S13. Bacterial strains used in this study.

Strain	Genotype	Source
<i>E. coli</i> BW25993	F ⁻ , Δ(araD-araB)567, λ ⁻ , rph-1, Δ(rhaD-rhaB)568, lacI ^q , hsdR514	(Datsenko and Wanner, 2000)
<i>E. coli</i> BW25993 ΔlacZ::MS2-BS	F ⁻ , Δ(araD-araB)567, λ ⁻ , rph-1, Δ(rhaD-rhaB)568, lacI ^q , hsdR514, ΔlacZ::MS2-BS	In this study
<i>E. coli</i> RL1314	F ⁻ , λ ⁻ , IN(rrnD-rrnE)1, rph-1, rpoC::GFP-kan	(Bratton et al., 2011)

Table S14. Bacterial plasmids used in this study.

Plasmid	Genotype	Source
pBELOBAC11	Ori2, CmR, lacZα, P _{T7} , P _{SP6}	GenBank Accession #: U51113
pBELOBAC11-PLacO3O1-BS (Target Plasmid)	Ori2, CmR, P _{LacO3O1} -BS	In this study

pZA25-GFP (Reporter Plasmid)	pSC101, KanR, P _{ara} -MS2-GFP	(Nevo-Dinur et al., 2011)
pCA24N-ligase	Ori, CmR, lacI ^q , P _{T5} -lac	(Kitagawa et al., 2005; Wilson et al., 2013)

Table S15. Number of cells observed, mean and squared coefficient of variation (CV²) of the absolute integer-valued RNA numbers per cell, in cells with the chromosome-integrated construct, when grown at 30 °C or 10 °C and subject to Novobiocin or Topotecan for 90 minutes. Also shown are the P values of the KS-test of statistical significance comparing the two pairs of distributions of RNA numbers at the two temperature conditions. For *p* values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected. This table is related to Figure 3 in the main manuscript.

Temperature and Perturbation	No. cells	Mean integer-valued RNA no. per cell	CV ²	<i>P</i> value of the KS-test
Novobiocin, 30 °C	615	0.85	7.62	
Novobiocin, 10 °C	615	0.25	8.7	0.004 (30 °C vs 10 °C)
Topotecan, 30 °C	610	1.23	3.67	
Topotecan, 10 °C	615	0.23	7.95	2.8×10 ⁻¹¹ (30 °C vs 10 °C)

PUBLICATION V

Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*

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, and A.S. Ribeiro

Mol. Microbiol., 99(4):686-699, 2016.
doi: 10.1111/mmi.13257

Publication reprinted with the permission of the copyright holders.

Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*

Samuel M. D. Oliveira,¹ Ramakanth Neeli-Venkata,¹ Nadia S. M. Goncalves,¹ João A. Santinha,² Leonardo Martins,² Huy Tran,¹ Jarno Mäkelä,¹ Abhishekh Gupta,¹ Marilia Barandas,² Antti Häkkinen,¹ Jason Lloyd-Price,¹ José M. Fonseca² and Andre S. Ribeiro^{1*}

¹Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland.

²UNINOVA, Instituto de Desenvolvimento de Novas Tecnologias, Campus FCT-UNL, 2829-516 Caparica, Portugal.

Summary

In *Escherichia coli*, under optimal conditions, protein aggregates associated with cellular aging are excluded from midcell by the nucleoid. We study the functionality of this process under sub-optimal temperatures from population and time lapse images of individual cells and aggregates and nucleoids within. We show that, as temperature decreases, aggregates become homogeneously distributed and uncorrelated with nucleoid size and location. We present evidence that this is due to increased cytoplasm viscosity, which weakens the anisotropy in aggregate displacements at the nucleoid borders that is responsible for their preference for polar localisation. Next, we show that in plasmolysed cells, which have increased cytoplasm viscosity, aggregates are also not preferentially located at the poles. Finally, we show that the inability of cells with increased viscosity to exclude aggregates from midcell results in enhanced aggregate concentration in between the nucleoids in cells close to dividing. This weakens the asymmetries in aggregate numbers between sister cells of subsequent generations required for rejuvenating cell lineages. We conclude that the process of exclusion of protein aggregates from midcell is not immune to stress conditions affecting the cytoplasm viscosity. The findings contribute to our understanding of *E. coli*'s internal organisation and functioning, and its fragility to stressful conditions.

Introduction

Unicellular organisms, such as *Escherichia coli*, particularly when in optimal environments, can continuously divide into genetically identical cells although, similarly to multicellular organisms, they are not free from errors, e.g. in protein production (Miot and Betton, 2004), that result in malfunctioning proteins that can hamper the functioning of cellular processes (Maisonneuve *et al.*, 2008).

Escherichia coli has evolved a complex machinery responsible for ensuring protein functionality that is able to catalyse proper protein folding and assist in the rescue of misfolded ones (Deuerling *et al.*, 1999; Wickner *et al.*, 1999), and can target misfolded proteins for degradation (Viaplana *et al.*, 1997), which allows both error correction and renewal of protein numbers (Willets, 1967; Goldberg, 1972). When this fails, *E. coli* can resort to protein aggregation (Sabate *et al.*, 2010; Tyedmers *et al.*, 2010; Winkler *et al.*, 2010), which likely reduces potentially harmful effects by rendering some of the malfunctioning proteins inert (Bednarska *et al.*, 2013). Recent evidence suggests that the aggregation is not an energy-free process (Govers *et al.*, 2014), which is consistent with being of importance for proper cellular functioning.

Unfavourable growth conditions or continued stress can enhance protein aggregation (Lindner *et al.*, 2008; Maisonneuve *et al.*, 2008; Winkler *et al.*, 2010; Govers *et al.*, 2014). This can lead to excessive aggregate accumulation (Bednarska *et al.*, 2013) that interferes with cellular functioning (Goldberg, 2003; Lindner *et al.*, 2008; Maisonneuve *et al.*, 2008). Recent studies showed that these aggregates are segregated to the cell poles (Winkler *et al.*, 2010; Coquel *et al.*, 2013), due to a volume exclusion effect caused by the presence of the nucleoid at midcell, similar to how plasmids are partitioned (Vecchiarelli *et al.*, 2012; Reyes-Lamothe *et al.*, 2014) and to how other large complexes (Straight *et al.*, 2007) are segregated to the poles. Provided that the segregation process is successful (evidence suggests that it is not entirely successful, even in optimal conditions; Gupta *et al.*, 2014), when the cell divides, it generates an asymmetry, in that both daughter cells will receive one new pole that is free of aggregates (Lindner *et al.*, 2008; Govers *et al.*, 2014). Consequently, as cells continue to divide, this segregation process results in the rejuvenation of several cells of the lineage (freeing them from aggregates), at the

Accepted 25 October, 2015. *For correspondence. E-mail andre.ribeiro@tut.fi; Tel. +358408490736; Fax +358331154989.

cost of a few cells with reduced reproductive vitality, due to inheriting the oldest poles containing several aggregates (Lindner *et al.*, 2008).

Several observations support the conclusion that the exclusion of aggregates from midcell is an energy-free process, caused by nucleoid exclusion (Gupta *et al.*, 2014) (first hypothesised in Winkler *et al.*, 2010). First, the exclusion effect is visible in a strong anisotropy in aggregate kinetics, located at the nucleoid borders, which favours aggregates accumulation at the poles (Gupta *et al.*, 2014). Also, while in cells where the nucleoid is centred the choice of pole is symmetric, in cells with off-centred nucleoids, a higher-than-by-chance fraction of aggregates preferentially locates at the larger pole (Gupta *et al.*, 2014). Finally, the aggregate kinetics, while affected by the nucleoid (Stylianidou *et al.*, 2014), is diffusive-like (Coquel *et al.*, 2013), even when at the pole (in agreement with the absence of transport or anchoring mechanisms).

Consequently, the efficiency with which aggregates are excluded from midcell should depend on factors such as nucleoid size as well as aggregate size and mobility within the cytoplasm, etc (Kuwada *et al.*, 2015). As these properties are likely affected by environmental conditions, e.g. temperature, it is reasonable to hypothesise that this process might lack robustness to some environmental stresses (Jeon *et al.*, 2013; Cherstvy and Metzler, 2015), particularly since, in those conditions, other functions are likely to be more critical (Clegg *et al.*, 2014).

Here, we study the robustness to non-optimal temperatures of the processes of segregation and retention of aggregates at the cell poles in *E. coli*. We address the following questions: To what extent are aggregate intracellular distributions temperature dependent? What are the causes for the temperature dependence? Finally, what are the long-term consequences of sub-optimal temperatures to aggregates numbers in cell lineages? To address these questions, we observed fluorescently tagged natural aggregates as well as synthetic fluorescent aggregates, along with inclusion bodies and nucleoids in individual cells subject to a wide range of temperatures.

Results

To study the temperature dependence of aggregate segregation and polar retention in *E. coli*, we observe aggregates by tracking lbpA-YFP proteins, which are accurate identifiers of the *in vivo* localisation of natural protein aggregates (Lindner *et al.*, 2008; Coquel *et al.*, 2013) (unlike e.g. Clp proteases and other proteins, whose tagging can alter their localisation and, possibly, aggregation; Landgraf *et al.*, 2012). Importantly, these tagged aggregates co-localise with inclusion bodies (Allen *et al.*, 1992). As such, we refer to the aggregates tagged by lbpA-YFP as 'natural' aggregates, even though, in the

measurements here conducted, their emergence in the cells is externally enhanced by the addition of streptomycin to the media (Lindner *et al.*, 2008). We also observe aggregates consisting of RNA sequences bound by multiple MS2-GFP proteins (Golding *et al.*, 2005), as previous studies suggest that they behave similarly to natural aggregates (Gupta *et al.*, 2014), have long lifetimes (Gupta *et al.*, 2014), do not vary significantly in size, do not aggregate and can be tracked individually (Golding *et al.*, 2005; Muthukrishnan *et al.*, 2012; Gupta *et al.*, 2014; Häkkinen *et al.*, 2014). As their composition differs from natural aggregates, we refer to these as 'synthetic' aggregates. Finally, since the segregation of aggregates to the poles is caused by nucleoid exclusion (Gupta *et al.*, 2014), we measure nucleoids' size (measured by the relative length along the major cell axis, given the invariance in width with temperature) and location (position of the nucleoid centre along the major cell axis). For this, we performed 4',6-diamidino-2-phenylindole (DAPI) staining of the nucleoid and, for validation, HupA-mCherry tagging.

We first assessed which range of temperatures is not lethal to the strain used (DH5 α -PRO). Results in Fig. S1 show that between 10°C and 43°C, conditions are not lethal (although at 10°C no growth is visible). Given this, microscopy measurements were performed for temperatures ranging from 10 to 43°C. Prior to image acquisition, cells were kept at the appropriate temperature for 60 minutes (e.g. Fig. S2) since, at this stage, aggregates and nucleoids already exhibit long-term behaviours (see below).

Behaviour of the synthetic MS2-GFP-tagged RNA aggregates

The MS2-GFP-tagged RNA synthetic aggregates are used since previous studies have shown that they behave similarly to the natural aggregates (at least, in optimal growth conditions), such as exhibiting preference for polar localisation (Gupta *et al.*, 2014). Also, they are visible for periods of time significantly longer than cell division (Muthukrishnan *et al.*, 2012). Further, their fluorescence intensity, which is determined by the number of MS2-GFP molecules bound to the target RNA (Golding *et al.*, 2005), has been reported to very stable over time (Muthukrishnan *et al.*, 2012; Gupta *et al.*, 2014; Häkkinen *et al.*, 2014). Finally, these properties do not appear to be significantly affected by temperature (for temperatures ranging from at least 10 to 43°C), suggesting that the number of bound MS2-GFP proteins is also not significantly temperature dependent (provided sufficient number of MS2-GFP proteins in the cytoplasm).

We performed additional tests to verify the robustness in time of their fluorescence intensity, which is a valuable property for purposes of quantification and tracking, and

to verify the similarity in behaviour with natural aggregates. For this, first, we studied the temporal fluorescence intensity of MS2d-GFP-tagged RNA molecules. Namely, we observed the fluorescence intensity of 40 individual, MS2-GFP-tagged RNAs over time (1 min^{-1}) in independent cells at 37°C . By inspection, we verified that each cell contained only one tagged RNA, to facilitate tracking. From the time lapse images, we obtained the fluorescence intensity of each of the 40 individual tagged RNAs for 30 minutes, since first detected. We fitted the intensity of each spot over time with a decaying exponential function and inferred the first-order, degradation rate constant of the spot intensity. We obtained a mean decay rate of $\sim 8.1 \times 10^{-5} \text{ s}^{-1}$, corresponding to a mean half-life of $\sim 144 \text{ min}$, which is longer than our observation window (60 min). As such, we conclude that, during the microscopy measurement period, the fluorescence of synthetic aggregates does not decrease significantly over time (gradually or abruptly), in agreement with previous reports (Golding *et al.*, 2005; Muthukrishnan *et al.*, 2012). These results are expected given previous studies of the coat protein of bacteriophage MS2 that show that MS2 binding sites are constantly occupied by MS2d proteins (Talbot *et al.*, 1999; Fusco *et al.*, 2003), which results in the 'immortalisation' of the target RNA due to isolation from RNA-degrading enzymes.

Next, we verified whether the synthetic aggregates, similar to IbpA-YFP, also co-localise with inclusion bodies. For this, we observed 85 cells at 37°C for 1 hour. At that moment, we counted in each cell the number of inclusion bodies (visible by Phase Contrast) and the number of synthetic aggregates (visible by fluorescence microscopy). Then, we counted how many times an aggregate is co-localised with an inclusion body. We also counted how many inclusion bodies 'contained' at least one aggregate, provided that the cell contained at least one aggregate (Fig. S3). We observed that 83% of the synthetic aggregates were co-localised with an inclusion body and that 91% of the inclusion bodies had a fluorescent synthetic aggregate co-localised with it. We conclude that the synthetic aggregates can be used to accurately inform on the *in vivo* presence and localisation of protein aggregates, similar to IbpA-YFP (Lindner *et al.*, 2008).

Positioning of IbpA-YFP-tagged aggregates as a function of aggregate size

The IbpA-YFP-tagged aggregates (of sufficient size to allow detection) exhibit significant variance in size (as measured by their fluorescence intensity). As such, we investigated whether their size (within the range of detection) significantly affected their behaviour, such as the degree of exclusion from midcell. For this, we investigated the location of IbpA-YFP-tagged aggregates along the

major cell axis as a function of their size, from cells at 37°C . We then extracted the 10% and the 25% smallest aggregates and their location along the major cell axis. From there, we obtained the fraction of aggregates excluded from midcell. Also, we calculated the fraction of aggregates that would be expected to be excluded from midcell if they distributed uniformly along the major cell axis, and accounting for the relative size of the nucleoid. This fraction equals 0.41 (assuming a uniform distribution along the major axis).

We found that 86% of all aggregates were located at the poles. Meanwhile, 85% of the 25% smallest aggregates were found at the poles. Finally, 88% of the 10% smallest aggregates were found at the poles. Thus, the distributions of aggregates along the major cell axis do not differ significantly with aggregate size. We conclude that, for the range of aggregate sizes that we can detect, their size does not influence the degree of exclusion from midcell.

Adaptation time of the relative nucleoid size and aggregate distributions to temperature shifts

We performed multi-modal microscopy at 10, 24, 37 and 43°C of DH5 α -PRO cells expressing MS2-GFP proteins along with the RNA target that form the synthetic aggregates (*Methods*). First, cells were grown at 37°C and then kept at the appropriate temperature for 15, 45, or 60 minutes. At these points in time, we performed DAPI staining and imaged cells once. From the images, we extracted the distributions of location and fluorescence intensity from synthetic aggregates and from stained nucleoids along the major and minor axes of each cell. We then determined whether a cell has one or two nucleoids and the nucleoid(s) borders along the cell axes (*Methods*). We define the region along the major axis containing the nucleoid(s) as 'midcell', while 'poles' are the two regions between these borders and the cell extremities.

Next, we compared the distribution of fluorescence intensity of aggregates along the major cell axis of cells with one nucleoid, when kept at the appropriate temperature for 15, 45 and 60 minutes. Results in Fig. S4 show that, for all temperatures, there are no significant differences between aggregate distributions at 15 and 45 min. We also compared the normalised distances of the nucleoid borders to the cell centre at 15 and 45 min. Again, we found no significant differences (Fig. S4). The same result was obtained when comparing distributions at 45 and 60 min after placing cells at the appropriate temperature.

We conclude that, for both aggregates and nucleoids, the distributions of fluorescence intensity at each temperature beyond 15 min. of adaptation time are representative of the long-term distributions in those conditions. Given this, from here onwards, we analyse

Table 1. Relative concentration of synthetic aggregates at the poles in cells with 1 nucleoid.

<i>T</i> (°C)	No. cells	Mean (standard) relative nucleoid length (μm)	Mean fraction of aggregates at poles	<i>P</i> value of the <i>t</i> -test	Mean relative 3-D concentration of aggregate numbers at poles	<i>P</i> value of the KS test
10	147	0.63 (0.12)	0.44		1.32	
24	604	0.56 (0.11)	0.39	0.21	1.09	0.11 (10 vs 24)
37	300	0.53 (0.11)	0.69	< 0.01	1.86	< 0.01 (24 vs 37)
43	204	0.47 (0.13)	0.81	< 0.01	1.79	0.05 (37 vs 43)

For each temperature condition, it is shown the number of cells analysed along with the mean and standard deviation of the relative nucleoid length. Also shown is the mean fraction of synthetic aggregates' numbers at the poles along with the *p* values of a *t*-test of statistical significance. Next, it is shown the relative 3-D concentration of synthetic aggregates at the poles (as measured by the ratio between the fraction of synthetic aggregate numbers at the poles and the normalised pole volume in individual cells), and the *p* values of a KS-test of statistical significance. In both statistical tests, for *p* values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

data collected from cells kept at the appropriate temperature for 60 min, unless otherwise stated.

Temperature dependence of relative nucleoid lengths and aggregate spatial distributions

From images of cells expressing synthetic aggregates, we obtained the mean relative nucleoid length along the major cell axis by DAPI staining and the mean fraction of synthetic aggregates at the poles, for each condition. Results in Table 1 show that the mean relative nucleoid length decreases slowly with increasing temperature. This decrease is significant according to Kolmogorov–Smirnov (KS) tests between all pairs of conditions (*p* values smaller than 0.01).

Nucleoid size assessment by HupA-mCherry tagging (Table S1) matched the results from DAPI staining for temperatures between 24 and 43°C (showing only slightly larger nucleoids in all cases). We attempted measurements at 10°C, but the HupA-mCherry signal was too weak.

Also from Table 1, in accordance with a *t*-test of statistical significance, the mean fraction of aggregates at the poles increases significantly with temperature, except between 10 and 24°C. This increase appears to be much

larger than what would be expected from the small decrease in relative nucleoid length.

To analyse whether the increase in the mean fraction of aggregates at the poles with temperature can be explained by the decrease in relative nucleoid length, for each condition, we obtained the relative 3-D concentration of aggregate numbers at the poles in each cell, accounting for the nucleoid length and the capped cylindrical shape of the cells (*Methods*). Also, we performed KS tests to compare the distributions of concentrations in individual cells from different conditions. Note that, if the aggregates tend to be excluded from midcell, their relative concentration at the poles will be larger than 1. Else, in the absence of nucleoid exclusion, this concentration should equal 1. In addition, if the degree of exclusion of aggregates from midcell is temperature dependent, we expect their relative concentration at the poles to change with temperature.

Results in Table 1 show that the mean relative concentration of aggregates at the poles is much higher than 1 at 37 and 43°C, but close to 1 at 10 and 24°C. The KS tests confirm that there is a significant change between 24 and 37°C (*p* value < 0.01). This difference in aggregate behavior is also visible when plotting the distances to the cell centre of aggregates and mean nucleoid border (Fig. 1),

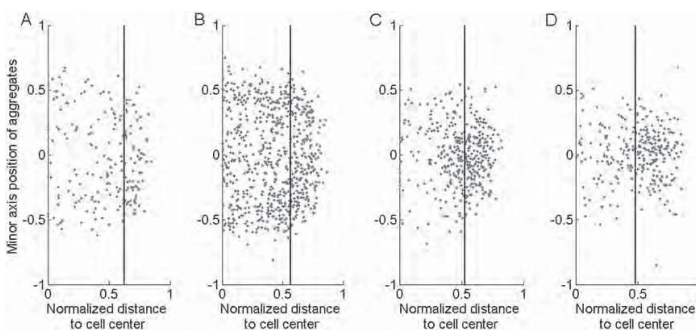


Fig. 1. Relative position versus normalised distance to cell centre of each aggregate, for various temperatures. Images were taken 1 hour after placing the cells at the specific temperature. Also shown, by the black solid line, is the mean relative nucleoid length. Measurements are from (A) 195 aggregates at 10°C (206 cells), (B) 707 aggregates at 24°C (1036 cells), (C) 398 aggregates at 37°C (367 cells) and (D) 288 aggregates at 43°C (306 cells).

Table 2. Correlations between synthetic aggregates positioning and nucleoid size and positioning in cells with one nucleoid.

T (°C)	No. cells	No. aggregates	Correlation between relative distance of aggregates to cell extreme and relative nucleoid length	t-Test	Correlation between the locations of nucleoid centre and aggregates along major cell axis	t-Test	μ_{nucleoid}
10	147	195	-0.01	0.84	-0.03	0.64	0.18
24	604	707	-0.14	< 0.01	-0.09	0.02	0.14
37	300	398	-0.25	< 0.01	-0.2	< 0.01	0.16
43	204	288	-0.25	< 0.01	-0.25	< 0.01	0.19

For each temperature condition, it is shown the number of cells and of synthetic aggregates analysed, along with the correlation between the relative distance of aggregates to the closest cell extreme and the relative nucleoid length and the *P* values of a *t*-test of statistical significance. Also shown is the correlation between the locations (i.e. distance to midcell) of the nucleoid centre and of each synthetic aggregate, followed by a *t*-test of statistical significance. For values < 0.01, it is accepted that the correlation is significant. Finally, the mean distance of the nucleoid centre to the cell centre (μ_{nucleoid}) is presented.

which shows an increase in aggregate density at the poles with increasing temperature. We conclude that the aggregate exclusion from midcell is much weaker at sub-optimal temperatures.

We performed the same measurements in cells expressing lbpA-YFP (Lindner *et al.*, 2008). Results in Table S2 show identical changes with temperature in the relative nucleoid length as well as in the mean fraction of lbpA-YFP aggregates at the poles. This allows concluding that the similarity in behaviour between synthetic and lbpA-YFP-tagged aggregates is maintained in the entire range of temperatures studied here. Further, we conclude that the spatial distributions of natural and synthetic aggregates change with temperature. Namely, the mean fraction of aggregates at the poles is significantly lower at the two lowest temperatures tested, and this cannot be explained by changes in the relative nucleoid length along the major cell axis.

Correlations of aggregate positioning with nucleoid size and positioning in individual cells

Next, we studied how temperature affects the correlation between aggregate positioning and nucleoid size and positioning. We first calculated the correlation between aggregate distance from the closest cell extreme and relative nucleoid length. Also, we performed *t*-tests of statistical significance of the correlation for each condition. Results in Table 2 show a negative correlation in all conditions that weakens with decreasing temperature, becoming not statistically significant at 10°C.

Next, we calculated the correlation between nucleoid centre and aggregate positioning's along the major cell axis (Fig. S5), and performed *t*-tests of statistical significance of the correlation for each condition. Since this correlation depends on the degree of 'off centring' of the nucleoid (Gupta *et al.*, 2014), we also compare the mean distance between nucleoid and cell centre (μ_{nucleoid}) between conditions. From Table 2, first, μ_{nucleoid} shows no

significant temperature dependence. Second, there are statistically significant negative correlations between the positioning of nucleoid centre and aggregates at 37 and 43°C, while at 10 and 24°C, this correlation is weak and not statistically significant.

We conclude that the aggregates positioning becomes less correlated with the nucleoids size and location for decreasing temperature, i.e., there is a reduction in the degree with which nucleoids affect aggregate positioning.

Anisotropies in aggregate dynamics

A previous study (Gupta *et al.*, 2014) showed that the correlations between aggregates and nucleoid, when existing, are generated by multiple encounters over time between them (rather than by a single event, such as a transport process). These encounters generate anisotropies in the aggregate dynamics at the nucleoid borders (Gupta *et al.*, 2014) (Fig. 2). A similar anisotropy, opposite in direction, occurs at the cell extremes, as the aggregates collide with the cell walls (Gupta *et al.*, 2014) (Fig. 2). The combination of anisotropies of opposite directions at the nucleoid borders and at the cell walls explains the long-term spatial distribution of the aggregates, namely, their preference for polar location (Gupta *et al.*, 2014).

To assess how these anisotropies are affected by temperature, we performed time lapse microscopy at 10, 24, 37 and 43°C for 45 minutes long, with images taken every minute, from which we obtained the displacement vectors of individual aggregates between consecutive frames and, from there, the 'anisotropy curve' for each condition (Methods), shown in Fig. 2.

From the distributions in Fig. 2, we quantified the 'degree of anisotropy', for each condition, from the area under the curve in the region of positive anisotropy (responsible for retaining aggregates at the poles; Gupta *et al.*, 2014). The sizes of these areas are shown in the insets in Fig. 2, and inform that, on average, the area has

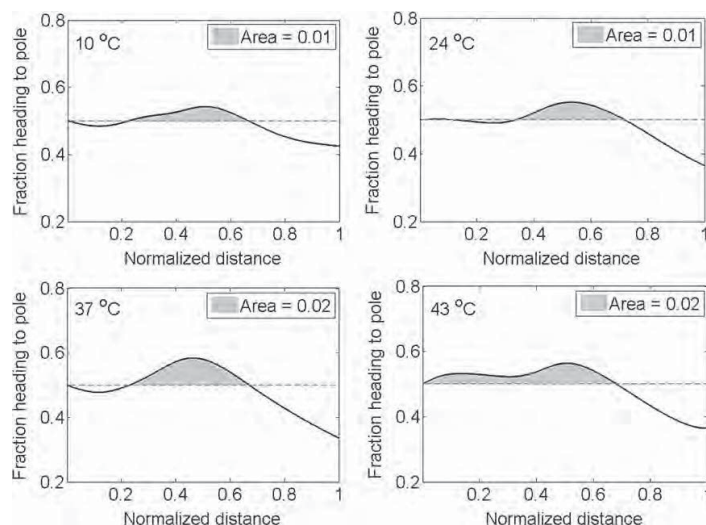


Fig. 2. Anisotropy curves of synthetic aggregates at different temperatures from time lapse images. Kernel Density Estimate (bandwidth of 0.1 normalised cell lengths) of the fraction of displacement vectors oriented towards the cell pole as a function of the distance towards midcell along the major cell axis. Measurements are at 10°C (43 cells), at 24°C (66 cells), at 37°C (184 cells) and at 43°C (41 cells).

half the size in the two lowest temperature conditions, indicating much weaker anisotropy. A similar reduction in the 'negative' areas (at the cell extremes) is also visible in these conditions. These reductions in the degree of anisotropies explain the loss of heterogeneity in the aggregate spatial distribution with decreasing temperature.

Further, we find no changes in nucleoid morphology with temperature changes that could explain this change in aggregate behaviour. Note that the mean positioning of the positive peak of anisotropy along the major cell axis is not significantly affected by temperature. This indicates that the nucleoid relative size only changes mildly with temperature (in accordance with the nucleoid relative size measurements reported in Table 1). In addition, we measured the absolute nucleoid length and width and found no significant changes with temperature (Table S3). From this lack of change in absolute nucleoid size, it is reasonable to assume that the nucleoid density does not change significantly in the range of temperatures tested, and thus is also not likely to be responsible for changes in aggregate spatial distributions with temperature. This is confirmed by inspection of the microscopy images, where it is visible that in no condition do aggregates exhibit a behaviour consistent with, e.g., 'entering' the nucleoid region. Rather, in all conditions, when at midcell, the aggregates locate near the cell inner-membrane.

Aside from this, it is noted that changes in cell morphology also cannot explain the observed changes in aggregate dynamics. First, we found no significant changes in the absolute cell width with temperature (Table S3). Second, while the absolute cell length increases with temperature (Table S3), it cannot explain the changes in

aggregate relative concentrations at the poles given the definition of 'pole region' (*Methods*). Given this, we next investigated the short-term dynamics of the aggregates as a function of temperature.

Spatial dynamics of the aggregates as a function of temperature

We speculated that the reduction in anisotropies with decreasing temperature is caused by a decrease in aggregate mobility. This is supported by the fact that both the area of positive anisotropy at the nucleoid borders and the area of the 'negative anisotropy' at the cell extremes are reduced with decreasing temperature, which is consistent with a general decrease in aggregate mobility throughout the cytoplasm, rather than a change in the properties of nucleoid or cell walls (while the nucleoid's ability to exclude aggregates could be affected by temperature, e.g. due to changes in density, no such changes are expected to occur to the cell walls, for the range of temperatures tested).

We thus performed time-lapsed, multi-modal microscopy at 10, 24, 37 and 43°C to measure the degree of diffusion of synthetic aggregates in each condition (as measured by the Diffusion coefficient, D) at the single cell level (*Methods*). Results in Table 3 show that D changes widely with temperature, being much smaller at lower temperatures.

We next assessed whether these differences in D between conditions could be explained by the differences in temperature alone (i.e. by the differences in free energy). For that, we calculated the relative dynamic vis-

Table 3. Relative dynamic viscosity of the cytoplasm and displacements bias at midcell.

T (°C)	No. cells	D ($\mu\text{m}^2 \text{min}^{-1}$)	Relative dynamic viscosity (relative to 37°C)	Γ ($\mu\text{m}^2 \text{min}^{-1}$)
10	43	0.005	1.67	5.4×10^{-6}
24	110	0.007	1.37	4.6×10^{-5}
37	184	0.01	1	1.6×10^{-4}
43	41	0.015	0.67	2.7×10^{-4}

For each condition, it is shown the number of cells studied along with the synthetic aggregates' diffusion coefficient, D , the relative dynamic viscosity of the cytoplasm relative to 37°C, and the bias in the displacement of aggregates located at midcell, Γ . Cells were kept at 37°C for 1 hour under the microscope and then kept at the appropriate temperature for 1 hour, after which we collected images for 45 minutes, with 1 minute interval.

cosity in each condition (relative to 37°C). This quantity should equal 1, if the differences in D with temperature between conditions are caused solely by differences in free energy.

Results in Table 3 show that the relative dynamic viscosity changes widely with temperature, being much higher at lower temperatures. Thus, we conclude that the differences in D with changing temperature are not caused solely by the differences in free energy, but also by changes in the thermophysical properties of the cytoplasm.

Finally, we verified that D , and thus the relative dynamic viscosity, is not biased by the cell growth rates (which differ between conditions). For that, we calculated the bias in the displacements of aggregates at midcell, Γ (Table 3). Since the values of this quantity are much smaller than the values of D in all conditions, we conclude that this bias is not significant.

Given this, and considering also the results on the correlations between the aggregates positioning and the nucleoids size and location (Table 2), as well as the measurements of local anisotropies along the major cell axis, we conclude that, at lower temperatures, the aggregates and nucleoid interact much less frequently during the measurement period. This, along with the stochasticity in diffusion, explains the observed near-uniform distribution of aggregates along the major cell axis at the two lower temperature conditions.

Finally, we considered another possibility, namely, that the lower fraction of aggregates at the poles at the lower temperature conditions could be explained by the fact that the aggregates do not have sufficient time to reach the poles prior to image acquisition. However, if this, rather than the reduced interactions between nucleoid and aggregates, was the cause, one would not observe the decrease with decreasing temperature in the area of the regions of positive anisotropy that is visible in the plots of the fraction of aggregates heading towards the poles along the major cell axis (Fig. 2), since this quantity is independent of the number of aggregates studied.

Spatial distribution and dynamics of aggregates following osmotic stress

Given the above, it is reasonable to expect that different means to increase the cytoplasm viscosity will cause similar changes on the short- and long-term aggregate behaviours.

It is known that, under osmotic stress, cells (when plasmolysed) exhibit enhanced cytoplasmic viscosity (van den Bogaart *et al.*, 2007; Konopka *et al.*, 2009; Mika *et al.*, 2010; Jin *et al.*, 2013). We placed cells under osmotic stress, and then assessed the consequences to the short-term dynamics and long-term spatial distribution of the synthetic aggregates within. We studied plasmolysed and adapted cells, which we compared with control cells (in optimal growth conditions).

We first assessed the spatial distributions of aggregates in control, plasmolysed and adapted cells for population images obtained by DAPI staining, following the application of osmotic stress (*Methods*). Results in Table S4 show that the relative 3-D concentration of aggregate numbers at the poles is much lower in plasmolysed cells than in the control (P value much smaller than 1). Further, it is close to 1, similar to low temperature conditions (Table 1). We conclude that the phenomenon of aggregate exclusion from midcell is absent in plasmolysed cells. Further, from Table 1, in adapted cells, this quantity is not statistically different from the control, from which we conclude that these cells recovered the ability to exclude aggregates from midcell.

To determine if the change in aggregates spatial distribution in plasmolysed cells has the same cause as in cells in low temperatures, we conducted 1-hour long, time-lapsed microscopy measurements with the perfusion of osmotic stress-inducing agent, to measure the diffusion coefficient of the aggregates and, thus, the cytoplasm's relative viscosity. From Table S5, the relative dynamic viscosity of plasmolysed cells is much higher than of control or adapted cells. Also, the bias in the displacements of aggregates at midcell, Γ , is not significant. We

Table 4. Mean relative concentration of synthetic aggregate numbers in between nucleoids in cells with two nucleoids.

T (°C)	No. cells	Mean relative length of midcell (two nucleoids and space in between) (μm)	Mean relative 1-D concentration of aggregate numbers in the gap	P value of permutation test
10	38	0.75	0.85	
24	154	0.68	0.78	< 0.01 (10°C vs 24°C)
37	46	0.72	0.69	< 0.01 (24°C vs 37°C)
43	93	0.7	0.68	< 0.01 (37°C vs 43°C)

For each temperature condition, it is shown the number of cells studied, the mean relative length along the major cell axis of the midcell region (which includes the two nucleoids and the space in between), the relative 1-D concentration of synthetic aggregate numbers in the space in between nucleoids (gap) (as measured by dividing the fraction of aggregate numbers in the gap by the distance between the inner borders of the nucleoids in individual cells) and the results of a test of statistical significance between differences in concentration (permutation test between pairs of conditions). For *P* values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

conclude that the changes in aggregate spatial distribution in plasmolysed cells and in cells subject to low temperatures are both due to increased cytoplasm viscosity.

Long-term consequences of the effects of temperature changes on the degree of exclusion of aggregates from midcell

Finally, we assessed whether the loss of effectiveness in excluding aggregates from midcell at lower temperatures causes tangible, long-term effects in cells of subsequent generations. These are expected to emerge, provided that the fraction of aggregates in between nucleoids in cells close to division increases significantly (see for example Fig. S6) (Stewart *et al.*, 2005; Lindner *et al.*, 2008; Govers *et al.*, 2014; Gupta *et al.*, 2014).

We thus investigated the distribution of synthetic aggregates along the major cell axis in cells with two nucleoids as a function of temperature. From each cell, we extracted the mean relative length along the major axis of each nucleoid (in general, the two nucleoids are of nearly identical size) and of the space in between the inner borders of the nucleoids (named here as the 'gap'). Also, we

obtained the mean relative 1-D concentration of aggregate numbers in the gap for each cell and performed a permutation test to compare the mean concentration between pairs of conditions.

Results in Table 4 first show that the mean relative length of the midcell region does not exhibit a consistent, significant change with temperature. Meanwhile, the mean relative 1-D concentration of synthetic aggregates in the gap region increases significantly with decreasing temperature. We conclude that, at lower temperatures, a larger fraction of aggregates will be randomly partitioned in division and then located at the new pole of the daughter cells, thus hampering the generation of asymmetries in aggregate numbers between the cells of a lineage.

Finally, we performed the same measurements in cells expressing IbpA-YFP. Results in Table 5 show no change in the mean relative length of the midcell region but a very significant increase in mean relative concentration of IbpA-YFP aggregate numbers in the gap. Thus, we conclude that, for both the natural and synthetic aggregates studied here, the relative concentration of aggregates in between nucleoids in cells near division is significantly

Table 5. Mean relative concentration of natural aggregate numbers in between nucleoids in cells with two nucleoids.

T (°C)	No. cells	Mean relative length of midcell (two nucleoids and space in between) (μm)	Mean relative 1-D concentration of aggregate numbers in the gap	P value of permutation test
10	195	0.72	0.53	
24	346	0.75	0.36	< 0.01 (10°C vs 24°C)
37	66	0.76	0.23	< 0.01 (24°C vs 37°C)
43	209	0.71	0.13	< 0.01 (37°C vs 43°C)

For each temperature condition, it is shown the number of cells studied, the mean relative length along the major cell axis of the midcell region (which includes the two nucleoids and the space in between), the relative 1-D concentration of natural IbpA-YFP aggregate numbers in the space in between nucleoids (gap) (as measured by dividing the fraction of aggregate numbers in the gap by the distance between the inner borders of the nucleoids in individual cells) and the results of a test of statistical significance between differences in concentration (permutation test between pairs of conditions). For *P* values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

higher at lower temperatures and that this is not due to changes in the relative nucleoid length along the major cell axis but rather due to the homogenous distribution of the aggregates in the cytoplasm.

Discussion

In optimal conditions, *E. coli* cells segregate large protein aggregates to the poles via nucleoid exclusion (Gupta *et al.*, 2014). Following cell divisions, this will result in the renewal of some cells of a lineage, which will be void of aggregates, at the expense of others that will contain several aggregates and exhibit accelerated aging (Lindner *et al.*, 2008). We observed live cells within the range of temperatures where they exhibit replication and found that, at low temperatures, the aggregate segregation and retention processes become non-functional in that, on average, aggregates no longer preferentially locate at the poles. Also, at the single cell level, their positioning no longer correlates with nucleoid size or positioning.

This non-functionality is shown to be due to an alteration in the aggregates' short-term behaviour. Namely, their displacements distribution no longer exhibits strong anisotropies at the nucleoid borders and cell extremes, due to a much enhanced cytoplasm viscosity that renders the interactions between nucleoid and aggregates too infrequent and weak to generate significant heterogeneities in the aggregate spatial distribution. To validate these findings, we subjected cells to osmotic stress. We observed that plasmolysed cells were also unable to segregate aggregates to the poles, due to their much enhanced cytoplasmic viscosity (Konopka *et al.*, 2009) (here verified).

The similarity in aggregate behaviour at low temperatures and under osmotic stress suggests that, in any conditions where cytoplasm viscosity is increased, one should expect loss of aggregate preference for polar localisation. Relevantly, increases in cytoplasm viscosity are known to occur under common stresses, such as carbon starvation and energy depletion, as well as during the stationary growth phase (Parry *et al.*, 2014). Further, we expect that, aside from large aggregates, other large cellular components such as plasmids, enzyme complexes, micro-compartments (Kerfeld *et al.*, 2010) and other macromolecules will be subject to nucleoid exclusion under optimal conditions. As such, we expect their spatial localisation in the cytoplasm to be similarly affected by increases in cytoplasm viscosity.

What is the origin of the increase in relative viscosity with decreasing temperature? Given the size of the aggregates studied here, in accordance to (Parry *et al.*, 2014), this is likely caused by the decrease in cellular metabolism rates responsible for 'fluidising' the cytoplasm (which

is in a 'near-glass' transition state when in optimal temperature; Parry *et al.*, 2014). Meanwhile, in the case of plasmolysed cells, it may be that the increased viscosity results from increased macromolecular crowding rather than reduced metabolism.

Our findings complement recent findings on how the bacterial cytoplasm functions. In Parry *et al.* (2014), the cytoplasm viscosity was shown to be 'metabolism dependent'. Conditions imposing poor metabolic rates cause increased cytoplasmic viscosity, which decreases the diffusion rate of macromolecules and other large cell components. We found that this has long-term consequences, namely, it renders nucleoid exclusion of protein aggregates far less effective, which perturbs the internal organisation of these components in the cell.

The negative effects of hampered aggregate preference for polar localisation should increase with prolonged exposure to stressful conditions. However, the consequences of failures in segregation and polar retention are likely to be rapidly dealt with (i.e. in a few generations), once conditions return to optimal, particularly in *E. coli*, which is capable of rapid division rates. This may explain its lack of energy-dependent 'repair' mechanisms (Clegg *et al.*, 2014). It might be that other bacteria, with much lower division rates (e.g. extremophiles), cannot employ the same strategy. In that scenario, compartmentalisation (Kerfeld *et al.*, 2010; Cornejo *et al.*, 2014) or transport mechanisms (in the case of eukaryotes) might be the adopted solutions.

Experimental procedures

Briefly, we used *E. coli* strain DH5 α -PRO, generously provided by I. Golding (Baylor College of Medicine, Houston, TX) to study synthetic aggregates, and *E. coli* strain MGAY (kind gift from Ariel Lindner, Paris Descartes University, France) to study tagged natural aggregates. Bacterial cell cultures were grown in lysogeny broth (LB). Synthetic aggregates were induced with 100 ng ml⁻¹ of anhydrotetracycline (aTc) and 0.1% L-arabinose for 50 min. After, 1 mM IPTG is added for 10 min. Natural aggregate production is induced by adding streptomycin (10 μ g ml⁻¹) for 30 min. Nucleoids were visualised by either DAPI staining or HupA-mCherry tagging. Live single-cell, single-molecule experiments were performed using Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with C2+ (Nikon) confocal laser-scanning system, and a thermal imaging chamber (CFCS2, Biopetechs, USA). Also, a peristaltic pump was used to provide continuous flow of fresh media to the cells. Example movies of time-lapse microscopy at 10°C (Movie S1) and 43°C (Movie S2) are provided. Cells were segmented from phase contrast images using software 'MAMLE' (Chowdhury *et al.*, 2013). Fluorescent aggregates were segmented as in Gupta *et al.* (2014) and Häkkinen *et al.* (2014). Nucleoids were detected and quantified as in Mora *et al.* (2011). Lineages were constructed by the software 'CellAging' (Häkkinen *et al.*, 2013). For additional information, see below.

Cells and plasmids

Experiments using synthetic aggregates were conducted in *E. coli* strain DH5 α -PRO, generously provided by I. Golding (Baylor College of Medicine). The strain information is: *deoR*, *endA1*, *gyrA96*, *hsdR17*($r_K^- m_K^+$), *recA1*, *relA1*, *supE44*, *thi-1*, $\Delta(lacZYA-argF)U169$, $\Phi80\delta lacZ\Delta M15$, *F⁻*, λ^- , *P_{N25}/tetR*, *P_{lacI}^{9/}/lacI* and *SpR*. This strain contains two constructs: (i) PROTET-K133 carrying *P_{LtetO-1}*-MS2d-GFP and (ii) a pIG-BAC vector carrying *P_{lac/ara-1}*-mRFP1-MS2-96bs (MS2-96bs stands for 96 MS2 binding site array) (Golding *et al.*, 2005). Dimeric-fused proteins MS2d-GFP are produced from the medium-copy vector, controlled by *P_{LtetO-1}*, regulated by tetracycline repressor and aTc inducer. RNA targets for multiple MS2d-GFP are produced from a single-copy F-plasmid, controlled by *P_{lac/ara-1}* (Lutz and Bujard, 1997), regulated by LacI and AraC repressors and IPTG and L-arabinose inducers. Further, to validate the results from DAPI measurements of nucleoid size and location, we inserted the plasmid pAB332 carrying *hupA-mcherry* (Fisher *et al.*, 2013). Expression of HupA-mCherry is controlled by a constitutive promoter (*hupA*).

Experiments to study natural aggregates were conducted using the *E. coli* MG1655 (MGAY) strain carrying the *ibpA-yfp* sequence in the chromosome under the control of the endogenous chromosomal *ibpA* promoter (kind gift from Ariel Lindner, Paris Descartes University, France).

Media and chemicals

Bacterial cell cultures were grown in LB media. The chemical components of LB (Tryptone, Yeast extract and NaCl) were purchased from LabM (Topley House, Bury, Lancashire, UK) and the antibiotics from Sigma-Aldrich (St. Louis, MO). Iso-propyl *b*-D-1-thiogalactopyranoside (IPTG) and aTc used for induction of the target genes are from Sigma-Aldrich. Agarose (Sigma-Aldrich) was used for microscope slide gel preparation. Finally, DAPI from Sigma-Aldrich was used to stain nucleoids.

Induction of production of fluorescent synthetic and natural aggregates

Pre-cultures were diluted from the overnight culture to OD₆₀₀ of 0.1 in fresh LB media, supplemented with appropriate antibiotics and kept at 37°C at 250 r.p.m. in a shaker until reaching OD₆₀₀ \approx 0.3.

Next, to produce synthetic aggregates, we proceeded as follows. After the DH5 α -PRO cells reached an OD₆₀₀ \approx 0.3, they were induced with 100 ng ml⁻¹ of aTc and 0.1% L-arabinose for 50 minutes until OD₆₀₀ \approx 0.5. At that stage, cells contain sufficient MS2d-GFP to detect target RNAs, and induction of *P_{lac/ara-1}* was completed by adding 1 mM IPTG. After 10 more minutes, cells were placed the appropriate temperature (10, 24, 37 or 43°C) for 1 hour.

To induce the production of natural aggregates, first, after MGAY cells reached an OD₆₀₀ \approx 0.3, they were placed at the appropriate temperature (10, 24, 37 or 43°C) for 1 hour. Then, they were incubated with streptomycin (10 μ g ml⁻¹) for 30 min.

Nucleoid visualisation by DAPI nucleoid staining

For nucleoid staining, cells were kept at a specific temperature for 60 min, and then fixed with 3.7% formaldehyde for 30 min. Next, cells were re-suspended in PBS, and DAPI (2 mg ml⁻¹) was added and cells were incubated for 20 min at room temperature. Finally, cells were washed twice with PBS (to remove DAPI in excess), and placed on a 1% agarose gel pad prepared with the appropriate media for microscopy (Chazotte, 2011).

Nucleoid visualisation by hupA-mCherry nucleoid tagging

The dimeric histone-like protein HU is one of the most abundant nucleoid-associated proteins that participates in the DNA structural organisation (Claret and Rouviere-Yaniv, 1997; Azam *et al.*, 1999). A version of this protein (HupA) has been tagged with the red fluorescent protein (mCherry) to study nucleoids in live *E. coli* cells (Maisonneuve *et al.*, 2008). This study showed that hupA-mCherry allows a proper assessment of the location and size of nucleoids *in vivo*. Expression of this synthetic protein was placed under the control of a constitutive promoter (*hupA*).

Osmotic stress

In van den Bogaart *et al.* (2007), it was reported that increasing sodium chloride (NaCl) concentration in the media results in a rapid osmotic upshift (from 0.15 to 0.6 Osm). This causes cytoplasm plasmolysis of *E. coli* cells (the water in the cytoplasm is expelled to the environment in a few seconds). Subsequently, cells undergo an adaptation process that allows recovering the ability to divide. As the adaptation time differs from cell to cell, when observing a population shortly after imposing osmotic stress conditions, one usually finds two distinct populations: adapted and non-adapted (Jin *et al.*, 2013), which differ in cell and nucleoid morphology, as well as in division rate (Jin *et al.*, 2013). Namely, plasmolysed (non-adapted) cells exhibit longer length and elliptic shape, contain only one condensed nucleoid and do not divide (Konopka *et al.*, 2009; Mika *et al.*, 2010; Jin *et al.*, 2013). Relevantly, in these cells, the high osmolality (> 0.15 Osm) causes the Diffusion coefficient of GFP to be heavily reduced (van den Bogaart *et al.*, 2007; Konopka *et al.*, 2009).

To expose cells to osmotic stress during time-lapse microscopy, 300 mM of NaCl was added to the growth media and pumped into the thermal chamber (set to 37°C) for 1 hour. For population microscopy imaging, the cells were kept under osmotic stress for 30 minutes (osmotic stress-inducing media with 300 mM of NaCl). In both cases, approximately \sim 0.68 Osm was reached (Konopka *et al.*, 2009).

Cells were considered to be plasmolysed when exhibiting filamentous and elliptical morphology (Konopka *et al.*, 2009), not dividing during the measurement period (1 hour), and if containing only one, condensed nucleoid (Mika *et al.*, 2010; Jin *et al.*, 2013).

Microscopy

Cells were visualised using a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with a 100 \times Apo TIRF (1.49

NA, oil) objective. The software for image acquisition was NIS-Elements (Nikon). Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system. The pinhole size was set to 1.2 AU. For confocal images, the size of a pixel corresponds to 0.124 μm using a scan area resolution of 1024 \times 1024 pixels. To visualise MS2-GFP-RNA 'spots', we used a 488 nm laser (Melles-Griot) and an emission filter (HQ514/30, Nikon). To visualise HupA-mCherry-tagged nucleoids, we used a 543 nm HeNe laser (Melles-Griot) and an emission filter (HQ585/65, Nikon). Phase contrast images of cells were captured using an external setup using a CCD camera (DS-Fi2, Nikon). Size of the images was 2560 \times 1920 pixels, in which a pixel corresponds to 0.049 μm . Epifluorescence images, for visualisation of DAPI-stained nucleoids, were taken by a mercury lamp excitation and a DAPI filter cube (EX 340-380, DM 400, BA 435-485, Nikon).

For fixed and live cell measurements, we placed 5 μl of culture on, respectively, 1% and 2% agarose gel pads of LB media between a microscope slide and a cover slip. Fixed cells were imaged once, while for time lapse microscopy the fluorescence images were taken once per minute for 45 minutes. In the latter, the desired temperature was kept from start to end of the session by a cooling/heating microfluidic system that provides continuous flow of deionised water at stable temperature (which does not enter in contact with the cells) into a thermal imaging chamber (CFCS2, Biopetech, USA). Meanwhile, a peristaltic pump provided continuous flow of fresh media to the cells, at the rate of 0.3 ml min⁻¹, through the thermal chamber. In the case of cells with synthetic aggregates, we added to the media the inducers of fluorescent synthetic aggregate production in the appropriate concentrations.

Spot detection

Fluorescent 'spots' are automatically segmented inside each cell using the kernel density estimation method for detecting fluorescently labelled subcellular objects in microscope images. This method measures the local smoothness of the image and determines spot locations by designating areas with low smoothness as a fluorescent spot. The spot intensity is then corrected by subtracting the mean cell background intensity multiplied by the area of the spot from the total fluorescence intensity of the spot (Gupta *et al.*, 2014).

Spot tracking

Spot tracking was performed using a semi-automatic method. First, we perform spot segmentation in each frame using the method above. An ID number is provided to the spot (automatically and then manually adjusted if needed) to identify it in each frame. Next, we manually correct for possible errors in the detection of the location of the spot. Afterwards, a displacement vector is automatically inserted, based on the shortest distance between the locations of the spot in consecutive frames. When the cell contains more than one spot, spots locations are determined as before but making use of the ID numbers so as to not misidentify spots between consecutive frames. Displacement vectors are then placed as before, based on the ID numbers of the spots. If, at any time point of the measurements, there are any doubts on the ID of the spots in a cell, that cell is discarded from the analysis.

Nucleoid detection and segmentation

Nucleoid detection and segmentation was done in each cell (in time series it was done at each frame). The nucleoid detection is performed using the Gradient Path Labelling algorithm (Mora *et al.*, 2011). This method starts by labelling each pixel based on its gradient azimuth and propagating these labels according to its gradient paths. The reduction of labels is obtained by applying equivalences (two labels are tagged as equivalents when both belong to the same maximum). Afterwards, a segmented image is obtained with the number of labels equaling the number of nucleoids. The Levenberg-Marquardt Least-Squares optimisation algorithm (More, 1978) is then used to obtain the parameters of 3D modified Gaussian functions that, in the case of two nucleoids, is described by $F(x, y) = G_1(x, y) + G_2(x, y) + z_0$, which fits each of the detected maximums. If only one nucleoid is present, G_2 is set to zero. In general:

$$G_i(x, y) = A_i \cdot \exp\left(-\left(a_i(x - x_{0i})^2 + 2b_i(x - x_{0i})(y - y_{0i}) + c(y - y_{0i})^2\right)^{(2/d)}\right)$$

where:

$$a = \cos^2 \theta / 2\sigma_x^2 + \sin^2 \theta / 2\sigma_y^2$$

$$b = -\sin \theta / 4\sigma_x^2 + \sin \theta / 4\sigma_y^2$$

$$c = \sin^2 \theta / 2\sigma_x^2 + \cos^2 \theta / 2\sigma_y^2$$

These functions allow translation in the xyz axes (x_0 , y_0 , z_0), amplitude scaling (A), rotation (θ), width in x-plane (σ_x), width in y-plane (σ_y) and amplitude profiling between square shape, bell shape and thin shape (d). The nucleoid fitting is done using a predefined value for d of 10, which was empirically selected to allow using the value z_0 as a threshold, in order to obtain the segmented nucleoid masks.

After nucleoid detection and segmentation, principal component analysis was used to obtain the position, dimension and orientation of the nucleoid in each cell. The polar region of a cell was defined as the area between the nucleoid and the major axis extremities (Fig. S7).

Example microscope images of cells along with the results of the segmentation process are shown in Fig. S2.

Estimation of the 3-D concentration of aggregates at the cell poles

The estimation of the concentration of aggregates at the poles accounts for the measured nucleoid size and the capped cylindrical shape of the cells. Fig. S7 shows a 2-D representation of a cell with a nucleoid within.

Let x be the absolute length of the nucleoid, w be the width of the cell along the minor axis and l be the length of the cell along the major axis. Then, the volume of the bacteria equals, approximately:

$$\text{Volume} = \pi \left(\frac{w}{2}\right)^2 (1-w) + \frac{4}{3} \pi \left(\frac{w}{2}\right)^3$$

To attain the volume of the midcell region, for simplicity, we assume that: $x < 1 - w$. Given this:

$$\text{Midcell volume} = \pi \left(\frac{w}{2} \right)^2 x$$

From this, one can derive the normalised fraction of midcell and poles volumes. These equal respectively:

$$\text{Normalized Fraction of midcell volume} = \frac{\frac{x}{1}}{1 - \frac{w}{3 \times 1}}$$

$$\text{Normalized Fraction of poles volume} = 1 - \frac{\frac{x}{1}}{1 - \frac{w}{3 \times 1}}$$

Note that the latter quantity is also the expected fraction of aggregates at the poles, assuming uniform intracellular distribution and accounting for the capped cylindrical shape of the cells.

To obtain the measured mean 3-D concentration of aggregate numbers at poles, we divide the mean fraction of aggregate numbers at the poles (Table 1) by the normalised volume of that region:

$$\begin{aligned} \text{Mean 3D concentration of aggregate numbers at poles} \\ \equiv \frac{\text{Mean fraction of aggregate numbers at poles}}{1 - \frac{\frac{x}{1}}{1 - \frac{w}{3 \times 1}}} \end{aligned}$$

Results of this estimation are presented in Table 1, for each condition. The values for x , l and w are shown in Table S3.

The above definitions and formulas are applied also to cells with two nucleoids, where the midcell region is defined as the region between the outer borders of the two nucleoids. Note that these formulas apply to all temperature conditions tested, even though as temperature increases the cells become, on average, longer.

Mean square displacement of the aggregates and relative viscosity of the medium

To measure the diffusion coefficient of the aggregates, we use the Mean Squared Displacement (MSD) of the aggregates after a time lag τ :

$$MSD(\tau) = E \|\mathbf{p}(t) - \mathbf{p}(t - \tau)\|^2 + \xi^2 = 2D\tau + \xi^2$$

where $\mathbf{p}(t)$ is the position of a spot at time t , E is the expectation over all spots and over all t , and ξ^2 is the measurement noise. To extract D discounting ξ^2 , we use the slope of the line taken from the first two points, i.e. $D = (MSD(2) - MSD(1))/2$. As seen in Fig. S8, the MSD is approximately linear for the first few τ . That is, for all measurements, $MSD(3)$ lies immediately beneath the line going through $MSD(1)$ and $MSD(2)$, justifying the assumption of approximately diffusive motion at this timescale.

From D , and assuming that the aggregates are spherical, the dynamic viscosity η of the medium in which the diffusive particle is moving is (the Stokes-Einstein equation):

$$\eta = \frac{k_B T}{6\pi r D}$$

where k_B is Boltzmann's constant and r is the particle radius. If the changes in temperature alone suffice to explain the changes in D , the relative dynamic viscosity between conditions should be approximately 1. The relative dynamic viscosity η_1/η_2 between two temperatures T_1 and T_2 with diffusion coefficients D_1 and D_2 can be quantified as:

$$\frac{\eta_1}{\eta_2} = \frac{T_2 D_2}{T_1 D_1}$$

Anisotropy distributions

From the time lapse images, we obtained their displacement vectors along the major cell axis between consecutive frames. These inform on the directionality of an aggregate between consecutive images (assessed by the sign of the displacement vector). Also, they inform on the degree of diffusion of the aggregates along the major cell axis during the intervals between consecutive images (assessed by the magnitude of the displacement vector).

As in Gupta *et al.* (2014), we extracted the displacement vectors going toward a pole and towards the cell centre, as a function of their point of origin. Next, we defined a sliding window with a width of 0.1 cell lengths and determined which displacement vectors originated within that window and their direction. We then analysed the directionality of the displacement vectors by counting the number of displacement vectors originated in the window, which were directed towards the midcell and towards the poles. Finally, we calculated the fraction of synthetic aggregates moving towards the poles in each window, as a function of the normalised distance to the cell centre. Cell growth between consecutive frames was accounted for by projecting the origin of each displacement vector into the cell space in the following frame, before calculating the magnitude and direction. The 'anisotropy curves' obtained for each condition are shown in Fig. 2.

Possible biases in the mean square displacement of aggregates due to cell growth

Escherichia coli cells grow by increasing the walls' length via incorporating new components at the midcell region (Laloux and Jacobs-Wagner, 2014). This is likely accompanied by the absorption of materials from the environment, which maintains the cytoplasm density nearly constant (90% of the cells in exponentially growing populations exhibit densities differing less than 0.75% from the mean) (Martinez-Sala *et al.*, 1981). Consequently, the process of cell walls growth is likely heterogeneous. However, this does not affect our calculations of MSD, since, during growth, the increase in cytoplasm volume is approximately homogenous along the major cell axis, as extracellular materials (e.g. water) do not enter the cell through a particular cell region. Given this and that the position of aggregates is not determined by the cell walls, as they float in the cytoplasm, our calculation of aggregate displacement vectors between consecutive frames does not need to be compensated for the heterogeneity of the cell walls growth process.

In any case, we can assume the extreme scenario of new materials entering the cell solely at the midcell region (which can bias our estimations of displacement vectors in cells with fast growth relative to aggregate movement), and estimate the upper bound of this bias. Let $x(t)$ be the un-normalised position of an aggregate along the major cell axis at time t and let the cell length at time t be $l(t)$. In the extreme case, the position of the aggregate at time $t + \tau$ is:

$$x(t + \tau) = x(t) + N(0, D) + \frac{l_{t+\tau} - l_t}{2}$$

where N is a Gaussian distribution with a mean of 0 and a variance of D . The displacement vectors are calculated as:

$$x(t + \tau) - \frac{l_{t+\tau}}{l_t} x(t) = (l_{t+\tau} - l_t) \left(\frac{1}{2} - \frac{x(t)}{l_t} \right) + N(0, D)$$

Given the above formula, and assuming the 'worst case' scenario of all aggregates being located at midcell, the bias in the measurement of D equals:

$$\Gamma = \left(\frac{1}{2} (l_{t+\tau} - l_t) \right)^2$$

We obtained this quantity for each temperature condition and compared with the measured diffusion coefficient, D . Results in Table 3 show that even in this extreme case, the values of D are much larger than those of Γ , which can thus be ignored.

Acknowledgements

Work supported by Academy of Finland (257603, ASR), Portuguese Foundation for Science and Technology (PTDC/BBB-MET/1084/2012, JMF and ASR), Jenny and Antti Wihuri Foundation (AH), TUT President's grant (JLP and JM) and Finnish Cultural Foundation (AG). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Conflict of interests

The authors declare that they have no competing interests.

References

- Allen, S.P., Polazzi, J.O., Gierse, J.K., and Easton, A.M. (1992) Two novel heat shock genes encoding proteins produced in response to heterologous protein expression in *Escherichia coli*. *J Bacteriol* **174**: 6938–6947.
- Azam, T.A., Iwata, A., Nishimura, A., and Ueda, S. (1999) Growth phase-dependent variation in protein composition of the *Escherichia coli* nucleoid. *J Bacteriol* **181**: 6361–6370.
- Bednarska, N.G., Schymkowitz, J., Rousseau, F., and Van Eldere, J. (2013) Protein aggregation in bacteria: the thin boundary between functionality and toxicity. *Microbiology* **159**: 1795–1806.
- van den Bogaart, G., Hermans, N., Krasnikov, V., and Poolman, B. (2007) Protein mobility and diffusive barriers in *Escherichia coli*: consequences of osmotic stress. *Mol Microbiol* **64**: 858–871.
- Chazotte, B. (2011) Labeling nuclear DNA using DAPI. *Cold Spring Harb Protoc* **1**: pdb prot5556.
- Cherstvy, A., and Metzler, R. (2015) Ergodicity breaking and particle spreading in noisy heterogeneous diffusion processes. *J Chem Phys* **142**: 144105.
- Chowdhury, S., Kandhavelu, M., Yli-harja, O., and Ribeiro, A.S. (2013) Cell segmentation by multi-resolution analysis and maximum likelihood estimation (MAMLE). *BMC Bioinformatics* **14**: 1–13.
- Claret, L., and Rouviere-Yaniv, J. (1997) Variation in HU composition during growth of *Escherichia coli*: the heterodimer is required for long term survival. *J Mol Biol* **273**: 93–104.
- Clegg, R.J., Dyson, R.J., and Kreft, J.-U. (2014) Repair rather than segregation of damage is the optimal unicellular aging strategy. *BMC Biol* **12**: 52.
- Coquel, A.-S., Jacob, J.-P., Primet, M., Demarez, A., Dimiccoli, M., Julou, T., et al. (2013) Localization of protein aggregation in *Escherichia coli* is governed by diffusion and nucleoid macromolecular crowding effect. *PLoS Comput Biol* **9**: e1003038.
- Cornejo, E., Abreu, N., and Komeili, A. (2014) Compartmentalization and organelle formation in bacteria. *Curr Opin Cell Biol* **26**: 132–138.
- Deuerling, E., Schulze-Specking, A., Tomoyasu, T., Mogk, A., and Bukau, B. (1999) Trigger factor and DnaK cooperate in folding of newly synthesized proteins. *Nature* **400**: 693–696.
- Fisher, J.K., Bourniquel, A., Witz, G., Weiner, B., Prentiss, M., and Kleckner, N. (2013) Four-dimensional imaging of *E. coli* nucleoid organization and dynamics in living cells. *Cell* **153**: 882–895.
- Fusco, D., Accornero, N., Lavoie, B., Shenoy, S.M., Blanchard, J.M., Singer, R.H., and Bertrand, E. (2003) Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol* **13**: 161–167.
- Goldberg, A. (1972) Degradation of abnormal proteins in *Escherichia coli*. *Proc Natl Acad Sci USA* **69**: 422–426.
- Goldberg, A.L. (2003) Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**: 895–899.
- Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005) Real-time kinetics of gene activity in individual bacteria. *Cell* **123**: 1025–1036.
- Govers, S.K., Dutré, P., and Aertsen, A. (2014) *In vivo* disassembly and reassembly of protein aggregates in *Escherichia coli*. *J Bacteriol* **196**: 2325–2332.
- Gupta, A., Lloyd-Price, J., and Ribeiro, A.S. (2014) *In silico* analysis of division times of *Escherichia coli* populations as a function of the partitioning scheme of non-functional proteins. *In Silico Biol* **12**: 9–21.
- Gupta, A., Lloyd-Price, J., Neeli-Venkata, R., Oliveira, S.M.D., and Ribeiro, A.S. (2014) *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli*. *Biophys J* **106**: 1928–1937.
- Gupta, A., Lloyd-Price, J., Oliveira, S.M.D., Yli-Harja, O., Muthukrishnan, A.-B., and Ribeiro, A.S. (2014) Robustness of the division symmetry in *Escherichia coli* and functional consequences of symmetry breaking. *Phys Biol* **11**: 066005.

- 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**: 1708–1709.
- Häkkinen, A., Kandhavelu, M., Garasto, S., and Ribeiro, A.S. (2014) Estimation of fluorescence-tagged RNA numbers from spot intensities. *Bioinformatics* **30**: 1146–1153.
- Jeon, J.-H., Barkai, E., and Metzler, R. (2013) Noisy continuous time random walks. *J Chem Phys* **139**: 121916.
- Jin, D.J., Cagliero, C., and Zhou, Y.N. (2013) Role of RNA polymerase in the organization of the bacterial nucleoid. *Chem Rev* **113**: 8662–8682.
- Kerfeld, C.A., Heinhorst, S., and Cannon, G.C. (2010) Bacterial microcompartments. *Annu Rev Microbiol* **64**: 391–408.
- Konopka, M.C., Sochacki, K.A., Bratton, B.P., Shkel, I.A., Record, M.T., and Weisshaar, J.C. (2009) Cytoplasmic protein mobility in Osmotically stressed *Escherichia coli*. *J Bacteriol* **01**: 231–237.
- Kuwada, N., Traxler, B., and Wiggins, P.A. (2015) Genome-scale quantitative characterization of bacterial protein localization dynamics throughout the cell cycle. *Mol Microbiol* **95**: 64–79.
- Laloux, G., and Jacobs-Wagner, C. (2014) How do bacteria localize proteins to the cell pole? *J Cell Sci* **127**: 11–19.
- Landgraf, D., Okumus, B., Chien, P., Baker, T.A., and Paulsson, J. (2012) Segregation of molecules at cell division reveals native protein localization. *Nat Methods* **9**: 480–482.
- Lindner, A.B., Madden, R., Demarez, A., Stewart, E.J., and Taddei, F. (2008) Asymmetric segregation of protein aggregates is associated with cellular aging and rejuvenation. *Proc Natl Acad Sci USA* **105**: 3076–3081.
- Lutz, R., and Bujard, H. (1997) Independent and tight regulation of transcriptional units in *Escherichia coli* via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res* **25**: 1203–1210.
- Maisonneuve, E., Ezraty, B., and Dukan, S. (2008) Protein Aggregates: an Aging factor involved in cell death. *J Bacteriol* **190**: 6070–6075.
- Martinez-Sala, E., Martin, J.A., and Vicente, M. (1981) Relationship of *Escherichia coli* density to growth rate and cell age. *J Bacteriol* **147**: 97–100.
- Mika, J.T., van den Bogaart, G., Veenhoff, L., Krasnikov, V., and Poolman, B. (2010) Molecular sieving properties of the cytoplasm of *Escherichia coli* and consequences of osmotic stress. *Mol Microbiol* **77**: 200–207.
- Miot, M., and Betton, J.M. (2004) Protein quality control in the bacterial periplasm. *Microb Cell Fact* **3**: 4.
- Mora, A.D., Vieira, P.M., Manivannan, A., and Fonseca, J.M. (2011) Automated drusen detection in retinal images using analytical modelling algorithms. *Biomed Eng Online* **10**: 1–15.
- More, J.J. (1978) The Levenberg-Marquardt algorithm: implementation and theory. In *Numerical Analysis*. Watson, G.A. (ed.). Heidelberg: Springer-Verlag, pp. 105–116.
- Muthukrishnan, A.-B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O., and Ribeiro, A.S. (2012) Dynamics of transcription driven by the tetA promoter, one event at a time, in live *Escherichia coli* cells. *Nucleic Acids Res* **40**: 8472–8483.
- 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.
- Reyes-Lamothe, R., Tran, T., Meas, D., Lee, L., Li, A.M., Sherratt, D.J., and Tolmasky, M.E. (2014) High-copy bacterial plasmids diffuse in the nucleoid-free space, replicate stochastically and are randomly partitioned at cell division. *Nucleic Acids Res* **42**: 1042–1051.
- Sabate, R., de Groot, N.S., and Ventura, S. (2010) Protein folding and aggregation in bacteria. *Cell Mol Life Sci* **67**: 2695–2715.
- Stewart, E.J., Madden, R., Paul, G., and Taddei, F. (2005) Aging and death in an organism that reproduces by morphologically symmetric division. *PLoS Biol* **3**: e45.
- Straight, P.D., Fischbach, M.A., Walsh, C.T., Rudner, D.Z., and Kolter, R. (2007) A singular enzymatic megacomplex from *Bacillus subtilis*. *Proc Natl Acad Sci USA* **104**: 305–310.
- Stylianidou, S., Kuwada, N.J., and Wiggins, P.A. (2014) Cytoplasmic dynamics reveals two modes of nucleoid-dependent mobility. *Biophys J* **107**: 2684–2692.
- Talbot, S., Goodman, S., Bates, S., and Fishwick, C. (1999) Stockley P. Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex. *Nucleic Acids Res* **18**: 3521–3528.
- Tyedmers, J., Mogk, A., and Bukau, B. (2010) Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* **11**: 777–788.
- Vecchiarelli, A.G., Mizuuchi, K., and Funnell, B.E. (2012) Surfing biological surfaces: exploiting the nucleoid for partition and transport in bacteria. *Mol Microbiol* **86**: 513–523.
- Viaplana, E., Feliu, J.X., Corchero, J., and Villaverde, A. (1997) Reversible activation of a cryptic cleavage site within *E. coli* β -galactosidase in β -galactosidase fusion proteins. *Biochim Biophys Acta* **1343**: 221–226.
- Wickner, S., Maurizi, M.R., and Gottesman, S. (1999) Post-translational quality control: folding, refolding, and degrading proteins. *Science* **286**: 1888–1893.
- Willems, N.S. (1967) Intracellular protein breakdown in non-growing cells of *Escherichia coli*. *Biochem J* **103**: 453–461.
- Winkler, J., Seybert, A., König, L., Pruggnaller, S., Hasselmann, U., Sourjik, V., et al. (2010) Quantitative and spatio-temporal features of protein aggregation in *Escherichia coli* and consequences on protein quality control and cellular aging. *EMBO J* **29**: 910–923.

Supporting information

Additional supporting information may be found in the online version of this article at the publisher's web-site.

Supporting Information for: “Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli*”

Samuel M.D. Oliveira¹, Ramakanth Neeli-Venkata¹, Nadia S.M. Goncalves¹, João A. Santinha², Leonardo Martins², Huy Tran¹, Jarno Mäkelä¹, Abhishekh Gupta¹, Marilia Barandas², Antti Häkkinen¹, Jason Lloyd-Price¹, José M. Fonseca², and Andre S. Ribeiro^{1,*}

¹ Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, 33101 Tampere, Finland

² UNINOVA, Instituto de Desenvolvimento de Novas Tecnologias, Campus FCT-UNL, 2829-516 Caparica, Portugal.

* Corresponding author: Tel: +358408490736, Fax: +358331154989, Email: andre.ribeiro@tut.fi
Office TC336, Department of Signal Processing, Tampere University of Technology
P.O Box 553, 33101 Tampere, Finland

Supplementary Figures

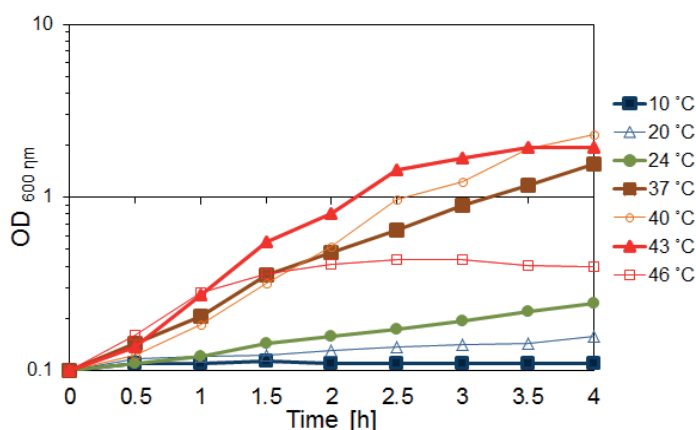


Fig. S1. Cell growth rate analysis. OD curves at 10°C, 20°C, 24°C, 37°C, 40°C, 43°C and 46°C. DH5α-PRO cells were grown in liquid LB media and the culture absorbance (OD at 600

nm) was measured every 30 minutes for 4 hours at 10 °C, 20 °C, 24 °C, 37 °C, 40 °C, 43 °C and 46 °C. The Y-axis is presented in the log scale.

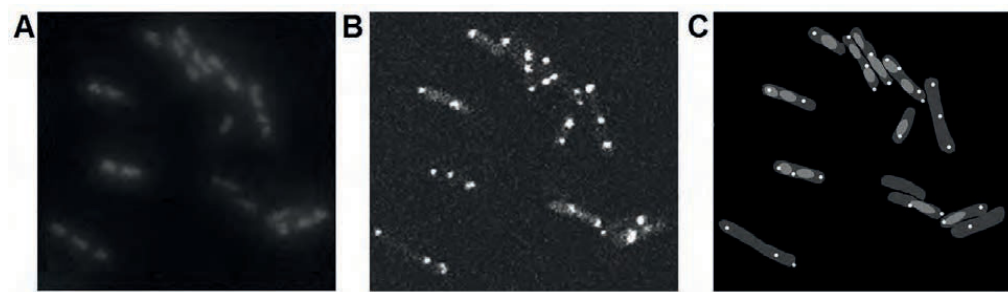


Fig. S2. Example microscopy images prior and after segmentation. (A) DAPI-stained nucleoids in cells, (B) cells with visible cytoplasm (filled with MS2-GFP proteins) along with MS2-GFP tagged RNA molecules (synthetic aggregates), visible as bright white “spots”, and (C) segmentation of the images in (A) and (B) merged into one image. Dark grey areas show segmented cells while segmented nucleoids are shown in lighter grey and synthetic aggregates are shown as small white spots.

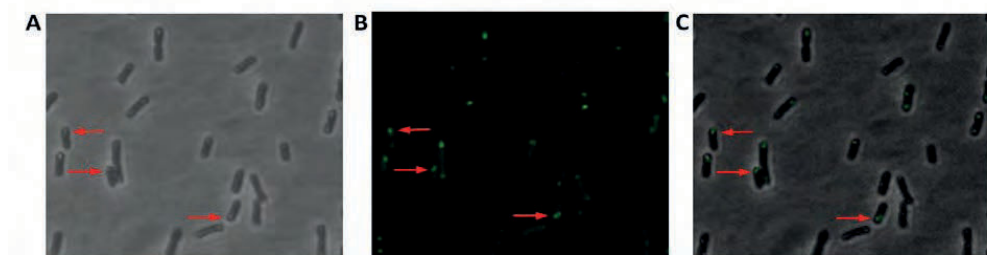


Fig. S3. Example images of cells visualized by Phase-Contrast and confocal microscopy along with merged image. (A) Example image of cells visualized by Phase-Contrast microscopy. The red arrows indicate example inclusion bodies. (B) Image by confocal microscopy of the cells with visible cytoplasm (filled with MS2-GFP proteins) along with MS2-GFP tagged RNA molecules (synthetic aggregates), visible as bright “spots”. The red arrows indicate example synthetic aggregates. (C) Images in (A) and (B) merged into one image. The red arrows indicate examples of co-localization between synthetic aggregates and inclusion bodies.

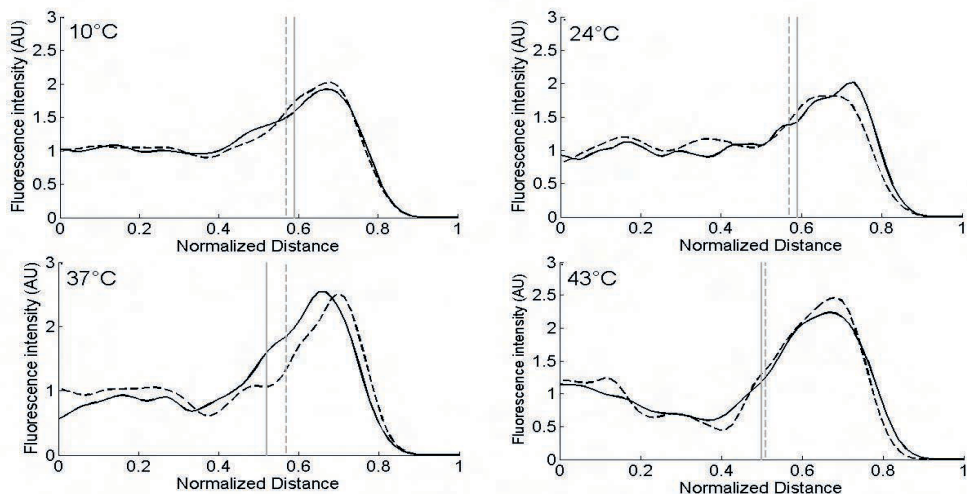


Fig. S4. Kernel density estimation of distribution of fluorescence intensity of aggregates and mean nucleoid border positioning. Kernel density estimation (KDE) of the spatial distribution of fluorescence intensity (in arbitrary units) of aggregates (black lines) and mean positioning of the nucleoid borders (vertical lines) relative to the cell center, 15 minutes (dashed lines) and 45 minutes (solid lines) after maintaining the cells at the appropriate temperature. Distances are normalized by the length of the major cell axis. All cells contain only 1 nucleoid. Measurements are from more than 300 cells per condition.

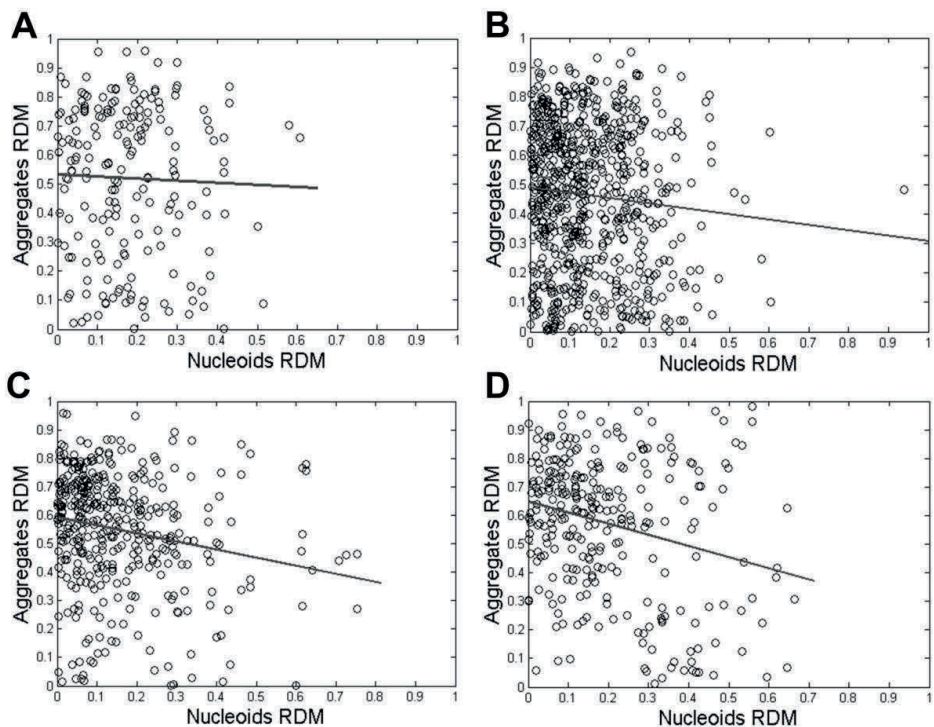


Fig. S5. Relative distance to midcell (RDM) of aggregates versus RDM of nucleoid center. RDM of individual aggregates versus RDM of the nucleoid center (along the major cell axis) measured from: (A) 195 aggregates at 10 °C, (B) 707 aggregates at 24 °C, (C) 398 aggregates at 37 °C, and (D) 288 aggregates at 43 °C. All cells contain only 1 nucleoid. The black solid line is the linear fit to the aggregates' RDM along the major cell axis as a function of the RDM of the nucleoid center. The negative inclination of the lines shows that, on average, if the nucleoid is off-centre, the aggregates will be located on the opposite side of the cell and closer to the cell center.

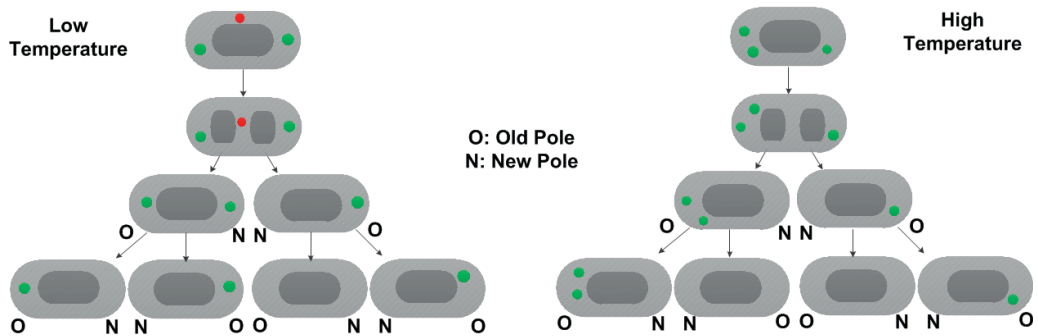


Fig. S6. Schematic representation of long-term effects of aggregates in between nucleoids prior to cell division. Cells are represented in light grey while nucleoids are represented in dark grey. (Left) Aggregates present at midcell (red balls), unlike segregated ones (green balls), will likely be located at the new poles of the cells of the next generation (with the selection of which cell following a random unbiased partitioning scheme). Consequently, only one cell of the last generation is free of aggregates. (Right) When the segregation and retention of aggregates at the poles is efficient, in the next generation all new poles of the cells will be free of aggregates, and when these cells divide, each will produce one daughter cell free from aggregates (unless new aggregates are produced in that time period). The letters 'O' and 'N' near the cells indicate whether a pole is old or new, respectively.

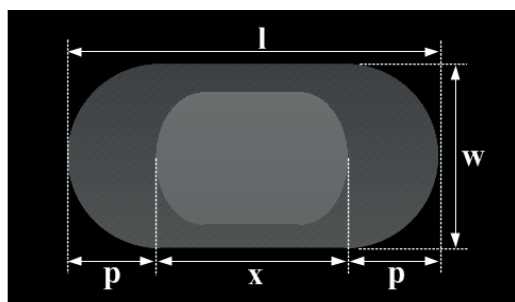


Fig. S7. Scheme of a cell with stained nucleoid. Also shown are l , the length of the cell along the major axis, w , the width of the cell along the minor axis, x , the length of the nucleoid region along the major cell axis, and p , the length of a 'polar region' of the cell.

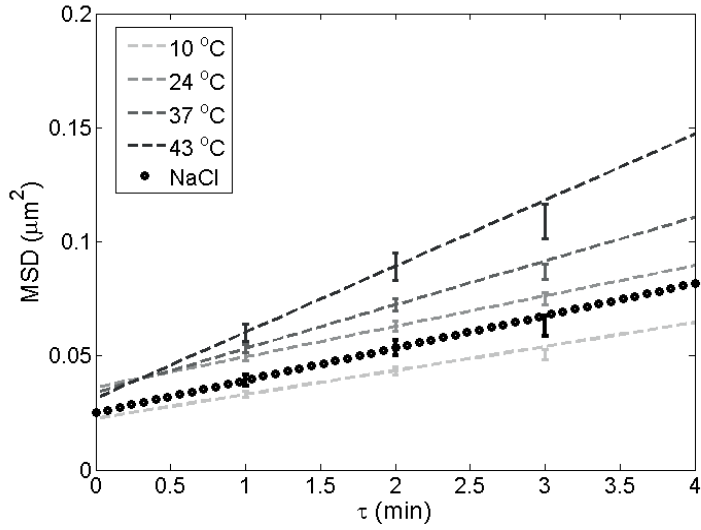


Fig. S8. Mean squared displacement of aggregates against the time lag τ . Error bars denote one standard error of the mean. The slopes of the lines represent the diffusion coefficients corresponding to the measurements presented in Table 3 in the main manuscript and in Table S5. Measurements are from 43 cells (at 10 °C), 66 cells (at 24 °C), 184 cells (at 37 °C), 41 cells (at 43 °C), and 43 plasmolyzed cells (at 37 °C and 300mM NaCl).

Supplementary Tables

Table S1. Relative nucleoid length along the major cell axis versus temperature as measured by HupA-mCherry tagging. Both mean and standard deviation (in parentheses) of each quantity are shown. Also shown is the number of cells analysed in each condition. For each condition, the mean relative length of nucleoids is shown. Visibly, as temperature is increased, this quantity decreases.

T (°C)	No. Cells	Relative Nucleoid Length (Mean and Std) in μm
24	78	0.60 (0.16)
37	53	0.57 (0.07)
43	60	0.51 (0.16)

Table S2. Relative nucleoid length along the major cell axis (mean and standard deviation), along with the fraction of IbpA-YFP aggregates at the cell poles at various temperatures, in cells with 1 nucleoid. For each condition, the mean relative length of nucleoids and the fraction of IbpA-YFP tagged aggregates at the poles are shown. Visibly, as temperature is increased, the former quantity decreases while the latter increases.

T (°C)	No. Cells	Mean (std) Relative Nucleoid Length (μm)	Mean Fraction of IbpA-YFP Aggregates at Poles
10	166	0.65 (0.07)	0.56
24	122	0.62 (0.12)	0.64
37	306	0.59 (0.10)	0.73
43	409	0.54 (0.09)	0.78

Table S3. Changes in absolute cell and nucleoid length and width along the major and minor cell axes with temperature as measured by DAPI staining. Both mean and standard deviation (in parentheses) of each quantity are shown. For each condition, width and length of cells and nucleoids within are shown. Note that, of these quantities, only the absolute cell length differs significantly with temperature.

T (°C)	Absolute Cell Length (Mean and Std) in μm	Absolute Cell Width (Mean and Std) in μm	Absolute Nucleoid Length (Mean and Std) in μm	Absolute Nucleoid Width (Mean and Std) in μm
--------	--	---	--	---

10	2.69 (0.86)	1.05 (0.12)	1.66 (0.56)	0.66 (0.11)
24	2.40 (0.67)	1.06 (0.11)	1.33 (0.39)	0.74 (0.11)
37	2.80 (0.67)	1.06 (0.12)	1.46 (0.43)	0.67 (0.11)
43	3.77 (1.18)	1.07 (0.12)	1.76 (0.57)	0.74 (0.11)

Table S4. Mean relative 3-D concentration of synthetic aggregate numbers at the poles, in cells under osmotic stress, containing 1 nucleoid. For each condition, it is shown the number of cells studied in the microscopy measurements, along with the relative 3-D concentration of synthetic aggregate numbers at the poles at 37°C in control (LB media), plasmolyzed and adapted cells. Cells were subject to osmotic stress (300 mM NaCl) for 30 min. prior to imaging. A permutation test was applied to test for statistical differences between the concentrations between the stress conditions and the control. For p-values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

Condition	Cells Analyzed (No. Cells)	Mean Relative 3-D Concentration of Aggregate Numbers at Poles	P-value of a permutation test (vs. Control)
Control	300	1.86	
Plasmolyzed (NaCl)	19 (from 222)	1.07	< 0.01
Adapted (NaCl)	203 (from 222)	1.55	0.06

Table S5. *In vivo* diffusion coefficient, D , relative dynamic viscosity (relative to Control), η , and bias in the displacement of aggregates located at midcell, Γ , for cells under osmotic stress (plasmolyzed and adapted) along with control cells, containing 1 nucleoid. For each condition, it is shown the number of cells studied in the time-lapsed microscopy measurements, the diffusion coefficient, D , the relative dynamic viscosity at 37°C in control (LB media), and the bias in the displacement of aggregates located at midcell, Γ , in plasmolyzed and adapted cells as well as in control cells. Cells were subject to osmotic stress (300mM NaCl) during the 1-hour imaging procedure. A permutation test was applied to test for statistical differences between the concentrations between the stress conditions and the control. For p-values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution is rejected.

Condition	No. Cells	D ($\mu\text{m}^2 \text{ min}^{-1}$)	Relative Dynamic Viscosity (η , relative to Control)	Γ ($\mu\text{m}^2 \text{ min}^{-1}$)
Control	184	0.01	1	1.6×10^{-4}

Plasmolyzed (NaCl)	43	0.007	1.35	1.5×10^{-4}
Adapted (NaCl)	61	0.015	0.62	0.8×10^{-4}
