TAMPEREEN TEKNILLINEN YLIOPISTO
TAMPERE UNIVERSITY OF TECHNOLOGY

Nadia Moreira Goncalves

Temperature Dependence of the Transcription Dynamics of Synthetic Genes in *Escherichia coli*

Julkaisu 1567 • Publication 1567

Tampere 2018

Tampereen teknillinen yliopisto. Julkaisu 1567
Tampere University of Technology. Publication 1567

Nadia Moreira Goncalves

Temperature Dependence of the Transcription Dynamics of Synthetic Genes in *Escherichia coli*

Thesis for the degree of Doctor of Philosophy to be presented with due permission for public examination and criticism in Sähkötalo Building, Auditorium SA203, at Tampere University of Technology, on the 21st of September 2018, at 12 noon.

Tampereen teknillinen yliopisto - Tampere University of Technology
Tampere 2018

Doctoral candidate: Nadia Moreira Goncalves
Laboratory of Biosystem Dynamics
Faculty of Biomedical Sciences and Engineering
Tampere University of Technology
Finland

Supervisor: Andre S. Ribeiro, Professor
Laboratory of Biosystem Dynamics
Faculty of Biomedical Sciences and Engineering
Tampere University of Technology
Finland

Pre-examiners: Supreet Saini, Associate Professor
Department of Chemical Engineering
Indian Institute of Technology Bombay
India

Petri-Jaan Lahtvee, PhD, Group leader
Institute of Technology
University of Tartu
Estonia

Opponent: Marjan De Mey, Associate Professor
Faculty of Bioscience Engineering
Ghent University
Belgium

ISBN 978-952-15-4186-5 (printed)
ISBN 978-952-15-4191-9 (PDF)
ISSN 1459-2045

Abstract

One of the major current goals in synthetic biology is the design of genetic components with more predictable functions. This predictability, however, does not depend solely on these components, but also on the environment where they will be inserted in.

Escherichia coli is one of the most studied microorganisms in Microbiology, and it is commonly used in Synthetic Biology as a host strain to test the functioning of the components of genetic systems. These components are typically well characterized under controlled laboratory conditions. However, it is unclear how unfavorable environmental conditions, such as temperature fluctuations, can affect their functionality and robustness.

In this thesis, we investigated how temperature affects the kinetics of transcription activation and subsequent dynamics of RNA production of synthetic genes in *E. coli*. For this, we made use of state-of-the-art *in vivo* single RNA-detection techniques and image analysis tools, to dissect, at the single-cell and single-RNA level, the kinetics of the rate-limiting steps in transcription, as well as the intake kinetics of inducer molecules. In addition, we analyzed how the temperature dependency is affected by the promoter structure.

Specifically, first, we characterized the intake kinetics of inducer molecules, from the media to the cell periplasm and then cytoplasm, at optimal and suboptimal temperatures. We found that, for a wide range of extracellular inducer concentrations, and in the absence of a transporter protein, the intake process is diffusive-like. The results also show that, the mean intake time increases non-linearly with decreasing temperature, likely due to the emergence of additional rate-limiting steps at low temperatures. Finally, our results indicate that the dynamics of this intake process affects significantly the expected RNA numbers in individual cells for a significant amount of time following induction and, thus, the overall distribution of RNA numbers of the cell population.

Next, we studied the temperature dependence of the dynamics of transcription initiation of a synthetic gene, engineered from a viral promoter. This dependency is shown to occur at the level of the underlying kinetics of the rate limiting steps in initiation. From the analysis of the empirical data, we found that, first, similarly to *E. coli* promoters, the T7 phage Phi 10 promoter exhibits more than one rate-limiting step during initiation. Also, the mean time-length of these steps is temperature dependent. However, contrary to *E. coli* promoters, the noise in RNA production increases with increasing temperature within the range of temperatures tested.

Finally, we investigated a key mechanism of transcription, namely, the robustness of a transcription repression mechanism by analyzing the rate of 'leaky' transcription events, i.e., RNA production events when under full repression. Using the $LacO_3O_1$ as a model promoter, from the analysis of the empirical data on single RNA production kinetics, we found that this promoter exhibits a leakiness rate that is higher at low temperatures, suggesting that its repression mechanism is less efficient under these conditions.

We believe that the studies presented here contribute to a better understanding of how temperature affects the transcription dynamics of synthetic genes in environments where temperature fluctuations occur. Since the acquired knowledge is of use to better understand the behavior of synthetic promoters, we expect our main contribution to be in the area of Synthetic Biology, namely, to be of value in predicting the robustness of future synthetic genetic circuits to temperature shifts. In particular, our results show that, in the genes studied, the repression mechanism is the most affected by temperature. This strong temperature dependence translates into the hindering of the promoter responsiveness to induction at sub-optimal temperature conditions. Additionally, our results suggest that this temperature-dependence of the robustness and responsiveness can be tuned, which indicates that it is possible to engineer synthetic promoters of higher response accuracy for a wider range of environmental conditions than those studied here. This knowledge can be used in the construction of synthetic genetic circuits with a more predictable, robust behavior.

Preface

This study was carried out at the BioMediTech Institute (BMT) and the Department of Signal Processing (DSP), of the Faculty of Biomedical Sciences and Bioengineering, and the Faculty of Computing and Electrical Engineering, respectively, at the Tampere University of Technology (TUT), under the supervision of Professor Andre S. Ribeiro.

First, I would like to express my sincere gratitude to my supervisor, Professor Andre S. Ribeiro for his persistent guidance, support and motivation during my doctoral studies. I am truly grateful for the opportunity to work in such a motivated and multi-disciplinary environment and it has been an honor to be a member of this group.

I am also deeply grateful to Professor Jose M. Fonseca, Universidade Nova de Lisboa, Portugal, for all his support in the collaborative research work, throughout the years of my doctoral studies.

I would also like to thank all the co-authors and the members of the Laboratory of Biosystem Dynamics (LBD), including alumni. I am deeply grateful for all the valuable discussions, useful comments and suggestions that made this thesis possible. Especially to Ramakanth Neeli-Venkata, Samuel Oliveira and Vinodh Kandavalli for all the scientific discussions on microscopy and experimental techniques, and for their support. I would also like to thank Huy Tran, Sofia Startceva, Leonardo Martins, Mohammed Nasuredeen and Cristina Palma for the useful discussions and contributions on the theoretical approaches that made this thesis possible. I am also grateful to the LBD alumni, namely Antti Häkkinen, Jason Lloyd-Price, Jarno Mäkelä, Anantha-Barathi Muthukrishnan and Jerome Chandraseelan for all their work, which helped in the completion of this thesis.

I am very grateful to all the personnel of the BMT, the DSP, the International Student Office and the Academic Coordinators for their excellent help during my studies at TUT. I would also like to thank the Supplementary Funding for Completing Dissertations and the Preparation Funding to Finalize PhD Thesis, for their financial support in concluding this work.

I would like also the thank Associate Professor Supreet Saini and Dr. Petri-Jaan Lahtvee for reviewing the manuscript of this thesis and for their insightful comments.

Finally, I am deeply grateful to my family and friends for all their support over the years, especially to my mother and brother, for their love and encouragement, which made this work possible.

Tampere, July 2018
Nadia Moreira Goncalves

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 3

1.3 Thesis Outline 4

2. BIOLOGICAL BACKGROUND 7

2.1 Gene Expression in *Escherichia coli* 7

2.2 Mechanism of Transcription 9

2.2.1 Transcription Initiation and its Rate-limiting Steps 10

2.3 Regulation of Transcription 14

2.3.1 Transcription Factors and Inducers 14

2.3.2 The Lac Operon 18

2.3.3 The Arabinose Operon 22

2.4 Gene Expression in T7 Phage 24

2.4.1 Transcription Initiation in T7 phage 25

2.4.2 T7 RNAp Promoters 26

2.5 Temperature Shifts and the Transcription of Synthetic Genes in *E. coli*..... 27

3. EXPERIMENTAL AND THEORETICAL METHODS FOR STUDYING THE DYNAMICS OF TRANSCRIPTION INITIATION 31

3.1 Fluorescent Proteins..... 31

3.2 Single-Molecules Approaches for Studying Transcription..... 33

3.2.1 The MS2-GFP Tagging Method..... 34

3.2.2 Constructing a Single-Copy Plasmid Carrying the Binding Sites for the MS2-GFP Reporter Proteins 35

3.2.1 Live Cell Imaging and Time-Lapse Microscopy..... 38

3.3 Methods for Validating Live Cell Imaging Measurements..... 40

3.4 Modelling Transcription Initiation in *E. coli*..... 41

3.4.1 Model of Inducer Intake and Active Transcription 45

4. COMPUTATIONAL TOOLS FOR IMAGE AND DATA ANALYSIS 47

4.1 Image Analysis and RNA Quantification..... 47

4.2 Measurement of Time Intervals and Estimation of Inducer Intake Times..... 48

4.3 Tau (τ) Plots 52

5. CONCLUSIONS AND DISCUSSION..... 55

BIBLIOGRAPHY 61

List of Figures

Figure 2.1: The central dogma of molecular biology 7
Figure 2.2: Electron microscope images of the coupling between transcription 8
Figure 2.3: In prokaryotes, genes are organized as operon.. 9
Figure 2.4: A schematic representation of the *E. coli* RNAp holoenzyme 10
Figure 2.5: The transcription process in *E. coli*. 10
Figure 2.6: Summary of the intermediate steps involved in isomerization 11
Figure 2.7: A τ-plot depicting the effect of temperature 13
Figure 2.8: Examples of repression and activation at promoters 16
Figure 2.9: An illustration of *E. coli*'s membrane. 17
Figure 2.10: Diagram of the operon model. 18
Figure 2.11: Illustration of the lactose operon, from *E. coli*. 19
Figure 2.12: Illustration of the operators of the lac operon. 20
Figure 2.13: A representation of the *lac* operon repression, by LacI.. 21
Figure 2.14: The L-Arabinose system of *E. coli*. 22
Figure 2.15: The L-Arabinose system in *E. coli* 23
Figure 2.16: Illustration of the regulatory regions of the *araCBAD* genes. 24
Figure 2.17: The genetic map of the T7 phage. 24
Figure 3.1: *In vivo* detection of RNA in *E. coli* cells. 34
Figure 3.2: An illustration representing the one-step isothermal reaction 36
Figure 3.3: An illustration showing the steps. 37
Figure 3.4: Schematic illustration of the CFCS2 microfluidics 39
Figure 3.5: An example of qPCR measurements 41
Figure 4.1: Example of phase contrast images. 48
Figure 4.2: Detection of RNA production events from individual cells.. 50
Figure 4.3: Tau plots for the T7 Φ 10 promoter activity. 54

List of Abbreviations

ATP	Adenosine triphosphate
bp	base pairs
BS	Binding sites
cAMP	Cyclic Adenosine Monophosphate
CAP	Catabolite activator protein
CRP	cAMP receptor protein
DNA	Deoxyribonucleic Acid
FISH	Fluorescence *in situ* hybridization
GFP	Green fluorescent protein
HSPs	Heat-shock proteins
IPTG	Isopropyl-β-D-1-Thiogalactopyranoside
KDE	Kernel density estimation
mRNA	messenger RNA
qPCR	Quantitative polymerase chain reaction
RNA	Ribonucleic acid
RNAp	RNA polymerase
TSS	Transcription start site
UTP	Uridine triphosphate
YFP	Yellow fluorescent protein

List of Publications

This thesis is a compilation of the following publications. In the text, these are referred to as **Publication I, II, III** and **IV**. The publications are reproduced with the permission from the publishers.

I. H Tran, SMD Oliveira, **NSM Goncalves**, and AS Ribeiro (2015) Kinetics of the cellular intake of a gene expression inducer at high concentrations. *Molecular Biosystems*, 11:2579-2587. DOI: 10.1039/C5MB00244C.

II. **NSM Goncalves**, L Martins, H Tran, SMD Oliveira, R Neeli-Venkata, JM Fonseca and AS Ribeiro (2016) In vivo single-molecule dynamics of transcription of the viral T7 Phi 10 promoter in *Escherichia coli*. *Proceedings of the 8th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies* (BIOTECHNO 2016), June 26-30, Lisbon, Portugal. ISBN: 978-1-61208-488-6 (pp. 9-15).

III. **NSM Goncalves**, SMD Oliveira, VK Kandavalli, JM Fonseca and AS Ribeiro (2016) Temperature Dependence of Leakiness of Transcription Repression Mechanisms of *Escherichia coli*. *Proceedings of the Computational Methods in Systems Biology* (CMSB 2016), September 21-23, Cambridge, United Kingdom.

IV. **NSM Goncalves**, S Startceva, CSD Palma, MNM Bahrudeen, SMD Oliveira and AS Ribeiro (2018) Temperature-dependence of the single-cell kinetics of transcription activation in *Escherichia coli*. *Physical Biology*, 15(2). DOI: 10.1088/1478-3975/aa9ddf

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

In **Publication I**, the intake dynamics of an inducer from the media was characterized at the single cell level. A novel stochastic model by which the inducers cross the bilayer membrane of the cells was proposed, based on measurements of the activation of the target gene. Importantly, the mean passage time in between the outer and inner membranes was obtained. Further, it was also established that it is the inner membrane that most retains inducers from entering the cytoplasm. The methodology developed in this work can be used in future studies of the cellular intake mechanisms, at the single cell level.
The author of this thesis contributed to Publication I as follows. The author was the main responsible for the measurements, ensuring that the target and reporter genes were working properly in all conditions used. Further, the author was responsible for planning and ensuring the media and

growth conditions and the cells fitness at the moment of the measurements. Finally, the author carried out the microscope measurements, ensuring identical microscope setting in all measurements and proper time intervals between measurements and proper total measurement times, based on cell health, etc. In the end, the author contributed to the writing of the manuscript, particularly in the sections describing the measurements and their outcomes, for which the author was the main responsible for its correctness.

Publication II presents the characterization of the *in vivo* dynamics of RNA production by the T7 RNA polymerase, as a function of temperature, at the single RNA level. Using single RNA detection techniques and qPCR for validation, the transcription kinetics of the T7 Phi 10 promoter was assessed at different temperatures, at the single cell, single RNA level. The results showed that for this promoter, as temperature is decreased, the mean rate of RNA production increases and, surprisingly, the noise associated with this process decreases.
The contribution of the author to this publication was as follows. First, the author proposed the study and planned it along with the corresponding author. Second, the author was responsible for planning and carrying out all experimental measurements. These included (i) the construction of the single copy plasmid carrying the target gene and the selection and transformation of the appropriate medium copy reporter plasmid expressing MS2-GFP; (ii) planning and conducting all necessary microscope measurements; (iii) and the supporting qPCR measurements. Third, the author assisted in the image analysis of the microscopy data. Finally, the author participated in the discussions throughout the project execution and was the main contributing author in the writing of the manuscript.

Publication III consists of an innovative study that characterizes the dynamics of leaky RNA production of the $LacO_3O_1$ promoter (i.e. RNA production kinetics in the absence of inducers and presence of repressors), at optimal and sub-optimal temperatures. By analyzing the empirical data with the support of a stochastic model of transcription, we determined which of the parameters of the model most contributed to the observed increase in leakiness with decreasing temperatures. These results showed that the parameters associated to the repression mechanism, rather than the transcription-associated parameters, were the most temperature-dependent, and thus were the ones that most contributed to the observed increase in leakiness at lower temperatures.
The author of this thesis contributed to this work as follows. First, the author, along with ASR conceived the study. Second, the author proposed, planned, and conducted all measurements, assisted by SMDO in the microscopy measurements and by VK in the qPCR measurements. Finally, the author assisted in image analysis, participated in the discussions during the execution of the project, and played a major role in the writing of the manuscript.

Publication IV presents the results of a study of the robustness of the intake kinetics of a transcription inducer, following temperature shifts of various degrees, using single-RNA *in vivo* measurements techniques. A new methodology for extracting the mean intake time from the single cell, single RNA measurements was proposed and validated. Next, measurements were conducted that, following image and data processing, showed that, as the degree of temperature downshift is increased, the intake time of the inducer (IPTG) from the media increases. Importantly, from the dissection of the parameter values associated to transcription and inducer intake kinetics, this study provides evidence that this phenomenon is likely due to the increased significance of a rate-limiting step in the intake process as temperatures are decreased.
The author of this thesis contributed to this study, as follows. First, the author conceived the study with ASR. Second, the author designed and conducted all experiments. Third, the author actively participated in the discussions during the execution of the project and in the writing of the manuscript.

Publication I have been used by Huy Tran in his PhD dissertation.

1 Introduction

1.1 Background and Motivation

Organisms have developed sophisticated mechanisms to survive in a wide variety of environmental conditions, suggesting the existence of mechanisms coordinating the timing and the type of responses that need to be activated when facing these conditions (Yamanaka 1999; Arsène *et al.* 2000; Kannan *et al.* 2008; Stoebel *et al.* 2009). One of the most difficult changes to adapt to is temperature fluctuations because it affects the physical properties of the cell and its components (e.g. proteins structure and kinetics of processes) and it affects the chemical components (due to altering balances between external and internal components). The strategies of microorganisms to cope with these changes usually involve reducing the degree of changes caused by the temperature shift. E.g., it tries to reduce changes in gene expression dynamics, and also the effects of physiological changes (Yamanaka 1999; Arsène *et al.* 2000).

The former strategy involves the cell triggering the expression of specific sets of proteins. In particular, when temperature raises, heat shock proteins are expressed to help preventing protein aggregation and assist the refolding of misfolded proteins (Arsène *et al.* 2000). When temperature drops, the cell activates the cold shock response (Yamanaka 1999) by expressing another specific set of proteins. Namely, upon cold shocks, there is a reduction in the efficiency of transcription and translation (Yamanaka 1999; López-García and Forterre 2000; Phadtare 2004). Cold shock proteins act at this level, helping to overcome this reduction by, e.g., facilitating translation at low temperatures by preventing the formation of secondary structures in RNAs (Yamanaka 1999; Phadtare 2004).

Regarding physiological changes, such as changes in the cell's membrane fluidity, processes are activated to counteract these effects. E.g. the homeoviscous adaptation consists of readjusting the membrane's lipid content, in order to reduce the changes in the fluidity (Sinensky 1974; Vigh *et al.* 1998). This adaptation is particularly important since the cell membrane plays a crucial role in the proper cell functioning. It acts as a physical barrier between the cell and the environment, allowing the cell to regulate its components, e.g. by regulating the processes of transmembrane transportation of molecules (Hazel 1995). In addition, evidence suggests that the fluidic properties of the membrane can act as a temperature sensor, allowing the cells to couple the detection of temperature fluctuations with the induction of expression of specific genes (Vigh *et al.* 1998; Los and Murata 2004). For example, such a link between the membrane's physical properties and gene expression has been reported in the case of genes involved in osmoregulation (Inoue *et al.* 1997) and heat shock response (Horvath *et al.* 1998).

Several studies have contributed to a better understanding of cellular adaptation to environmental changes. In the case of *Escherichia coli* and similar organisms, most studies indicate that the cells achieve this adaptability by tuning their transcription level (Tao *et al.* 1999; Yamanaka 1999; Arséne *et al.* 2000; Gunasekera *et al.* 2008). Particularly, it has been reported that the key regulatory steps in gene expression occur at the transcription initiation level (McClure 1980; Browning and Busby 2004).

Most of our present knowledge on transcription initiation resulted from *in vitro* studies (McClure 1980; Bertrand-Burggraf *et al.* 1984; Buc and McClure 1985; Lutz *et al.* 2001). However, recent improvements in live cell imaging techniques have made possible the *in vivo* characterization of the transcription process with great detail. Namely, the expression of fluorescent proteins fused with the desired gene of interest has allowed the *in vivo* study of gene expression (Elowitz *et al.* 2002; Yu *et al.* 2006).

One more recent method that is becoming popular for conducting such studies, at the RNA level, is the MS2-GFP tagging of RNA sequences, which allows the real time detection of individual RNA molecules with high spatial and temporal resolution (Fusco *et al.* 2003). From these measurements, it is possible to detect changes in the numbers of RNA molecules in single cells with single-RNA resolution over time, from which the information regarding the processes responsible for the observed changes can be extracted (Golding *et al.* 2005; Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016; Mäkelä *et al.* 2017).

The MS2-GFP tagging method has been particularly useful in understanding how temperature changes affect the dynamics of the underlying steps of transcription initiation. An example is a

study showing that for the TetA promoter, a third rate-limiting step appears at sub-optimal temperatures (Muthukrishnan 2014). Contrary to the behavior observed for TetA, a recent study has shown that for the Lac/ara-1 promoter, temperature changes affect the closed complex formation, suggesting that the interaction between the RNAp and the promoter region is more sensitive to temperature fluctuations (Oliveira *et al.* 2016a). These studies suggest that the dynamics of the rate-limiting steps in initiation of these promoters cause them to respond differently, in their RNA production kinetics, to temperature changes.

It has been further shown that temperature changes also affect the dynamics of synthetic genetic circuits (Stricker *et al.* 2008; Oliveira *et al.* 2015). These studies have demonstrated that when these synthetic circuits are subject to changing temperatures conditions one of the components loses functionality, thus hampering the stability of the circuit. Other studies have also reported this effect but used the temperature sensitivity of the components of the circuit as a means to control the activity of the circuit (Isaacs *et al.* 2003), or to engineer circuits exhibiting temperature compensation (Hussain *et al.* 2014). Such studies show the importance of knowing how the components of genetic circuits respond to environmental fluctuations. Further, given that these circuits are to be inserted into host cells, it is also important to assess their behavior in the cellular environment. For instance, several cellular processes depend either directly or indirectly on the host cell and, as such, they are also subject to these environmental conditions (Cardinale and Arkin 2012). Thus, it is safe to assume that synthetic genes or gene networks will also be affected by these conditions, which can compromise their activity and predictability. This can be a major issue when working with such systems, and to better assess this, it is necessary not only to characterize the transcription kinetics of its constituent genes, in case of a gene network, but also to evaluate the impact that the host factors have on its kinetics. Understanding how synthetic genes respond to environmental fluctuations can help in the engineering of synthetic genes or gene networks with a more predictable behavior.

1.2 Thesis Objectives

The aim of this thesis is to characterize the effects of temperature on the dynamics of transcription initiation of synthetic genes in *E. coli*. First, to investigate how the intake kinetics of inducer molecules affects cell-to-cell variability in the expression dynamics of its target gene, we studied the kinetics of this process and its dependency on temperature fluctuations. For this, we characterized the intake kinetics of IPTG molecules, at optimal and sub-optimal temperatures, using *in vivo* measurements at a single RNA level (Golding *et al.* 2005) for the Lac/ara-1 promoter (Lutz and Bujard 1997).

Second, we studied how the temperature dependency differs between promoter structures. Here, we first measured the transcription rate of the T7 phage constitutive Phi (Φ) 10 promoter (Dunn and Studier 1983) at various temperatures, using single RNA detection methods (Golding *et al.* 2005). To better assess the observed changes in the transcription rate of this promoter, we conducted additional measurements that allowed us to identify which step in the transcription initiation regulatory process is most affected by temperature.

Finally, we made use of the $LacO_3O_1$ promoter to evaluate the temperature dependency of its repression mechanism. For this, we first measured its active transcription rate (when fully active) and its leakiness rate (when repressed) under different temperature conditions. Next, after establishing that its leakiness rate is temperature dependent, we used a standard model of transcription and additional measurements to identify which parameters contributed the most to the observed temperature dependency.

In summary, this thesis' objectives can be listed as follows:

I. Characterization of the intake kinetics of IPTG inducer molecules at optimal and sub optimal temperatures.

II. Investigate the temperature dependency of the transcription initiation dynamics of a synthetic gene as a function of:

a. Promoter structure. Specifically, we addressed the following question: Does a viral, constitutive promoter exhibit similar behavioral changes as an *E. coli* promoter, to changing temperatures?

b. Promoter repression mechanisms. How efficient is the repression mechanism of the $LacO_3O_1$ promoter at sub-optimal temperatures?

Objective I was achieved in **Publication I** and **Publication IV**. Objective IIa was achieved in **Publication II**. Finally, Objective IIb was achieved in **Publication III**.

1.3 Thesis Outline

This thesis is organized as follows. Chapter 2 introduces the biological background, presenting the current knowledge on gene expression and its regulatory mechanisms. It also describes the current knowledge, preceding our results, on the effects of temperature shifts on the dynamics of transcription initiation. Chapter 3 provides an overview on the experimental and theoretical methods

used in gene expression studies, with those used in this thesis being described in detail. Also, it introduces the models and modeling strategies used in this thesis. Chapter 4 presents the computational tools used for image analysis and data extraction, with emphasis on cell segmentation, quantification of single RNA molecules, and measurements of time intervals between consecutive RNA productions in individual cells. Finally, the conclusions and discussion are presented in Chapter 5.

2 Biological Background

This chapter gives an overview on the biological processes studied in this thesis. It describes the gene expression process in *E. coli* and presents a brief summary of this process in T7 phage. This chapter also presents the mechanisms of transcription and its regulation. Finally, the effects of temperature changes in the transcription dynamics of synthetic genes is discussed.

2.1 Gene Expression in *Escherichia coli*

Gene expression can be defined as the process by which the genetic information, stored in a gene or genome is expressed into functional units. There are two main steps by which this process occurs, namely transcription and translation, forming the central dogma of molecular biology (Crick 1970) (Figure 2.1). During transcription, the information present in the DNA is transferred into the complementary messenger RNA (mRNA), which is then used to synthesize proteins by a process called translation (McClure 1985; Ramakrishnan 2002).

Figure 2.1: The central dogma of molecular biology (Crick 1970). The arrows indicate the direction of the flow of information, with the information stored in the DNA transferred to RNA, by transcription. Translation converts the information stored in the RNA into proteins. The information stored in the DNA can be replicated in the DNA replication process. Additionally, in some special cases, the information stored in the RNA can be transferred to the DNA, by a process called Reverse Transcription (dashed arrow).

Unlike eukaryotes, *E. coli* does not have a nucleus, and therefore there is no physical separation between transcription and translation, allowing these two process to be physically coupled as they occur in the cell cytoplasm (Miller *et al.* 1970) (Figure 2.2).

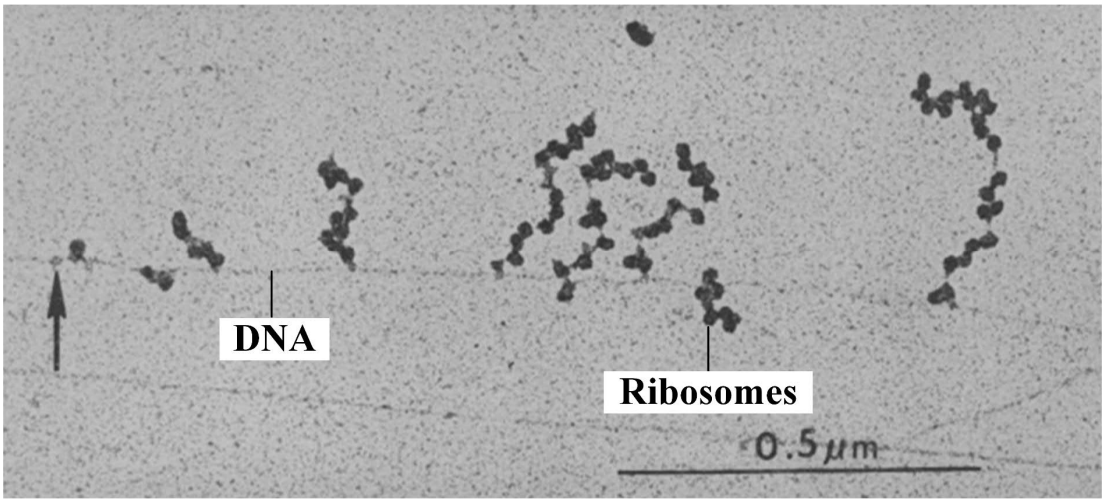


Figure 2.2: Electron microscope images of the coupling between transcription and translation, in *E. coli*. The arrow indicates putative RNA polymerase molecules, presumably near the transcription start site. Additionally, multiple ribosomes are also visible translating the mRNA as it is being transcribed. Adapted from (Miller *et al.* 1970), reprinted with permission from AAAS.

E. coli is one of the most studied microorganisms in biology, and has been widely used in the identification and characterization of several molecular processes (Blattner *et al.* 1997; Lee and Lee 2003; Faith *et al.* 2007). The genome of *E. coli*, comprising approximately 4000 genes coding for structural and regulatory proteins (Blattner *et al.* 1997), consists of a single circular double-stranded chromosomal DNA kept in a highly supercoiled and compact structure known as the nucleoid. In addition to its chromosomal genes, *E. coli* can also carry extra-chromosomal DNA, known as plasmids, which contain genes that, for example, can provide these cells with antibiotic resistance (Eliasson *et al.* 1992).

In prokaryotes such as *E. coli*, most genes are comprised of three components: a promoter, operator sites and structural genes (Alberts *et al.* 2002). The promoter is a specific DNA region that is recognized by the RNA polymerase (RNAp) as a location to start transcription. The operator sites are small sections in the DNA that are recognized by regulatory molecules that modulate the transcription of a gene.

Unlike regulatory genes, which appear in isolated form, structural genes are usually organized in a single transcriptional unit, called operon, where a group of genes are under the control of a single promoter (Lewis 2005; Schleif 2010) (Figure 2.3). As a result, the operon is transcribed into a

single mRNA molecule carrying the code for multiple proteins which will perform correlated functions (Jacob *et al.* 1961).

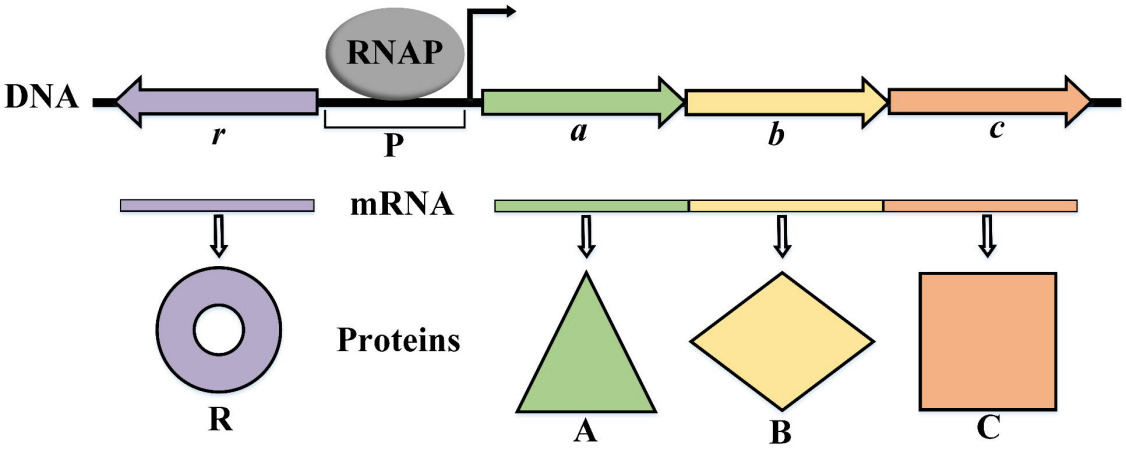


Figure 2.3: In prokaryotes, genes are organized as operon. The promoter (P) located in an intergenic region initiates the synthesis of mRNA by recruiting RNA polymerases and facilitating the formation of a transcription elongation complex, which produces mRNAs that terminate at an intrinsic terminator. These genes are transcribed as a single mRNA, which codes for the proteins A, B and C. Adapted from Nature Education (Adapted from Pierce, Benjamin. Genetics: A Conceptual Approach, 2nd ed. All rights reserved).

2.2 Mechanism of Transcription

The enzyme responsible for the transcription process is the core RNAp, which consists of five core subunits (ββ', α_2, and ω). The core subunits form a stable enzyme, and are necessary for the RNA synthesis, but are not able to start transcription from a promoter (Young *et al.* 2002). For it to be able to initiate transcription, the core RNAp enzyme must be bound by one of the several sigma (σ) subunits of *E. coli* (Murakami *et al.* 2002). This results in a RNAp in the holoenzyme form, which contains one copy of a σ subunit, with a high affinity for specific *E. coli* promoters (Wösten 1998) (Figure 2.4).

The bacterial transcription process occurs in three sequential steps: initiation, elongation and termination (Figure 2.5). Initiation consists of the RNAp holoenzyme binding to the promoter region, unwinding the DNA sequence and escape from the promoter region, so as to initiate elongation. During elongation, the RNAp moves along the DNA, in the 3' to 5' direction, while synthesizing the mRNA. When reaching a specific termination signal the elongation is halted, and the newly synthesized mRNA and the RNAp are released, in a step known as termination (Nudler and Gottesman 2002).

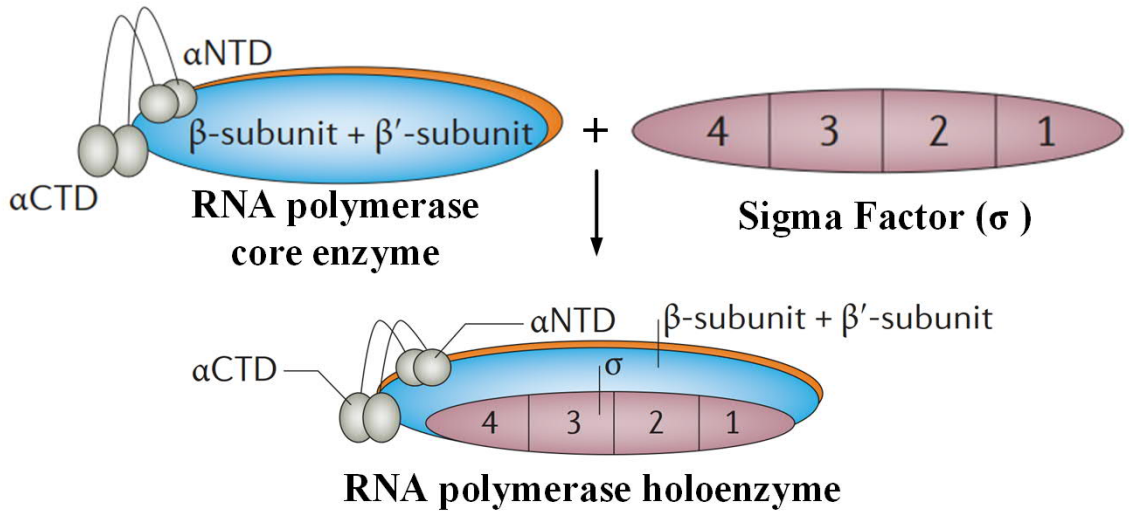


Figure 2.4: A schematic representation of the *E. coli* RNAp holoenzyme, comprising the RNAp core enzyme and a σ factor. Adapted from [Nature Reviews Microbiology] (Browning and Busby 2016), copyright (2016).

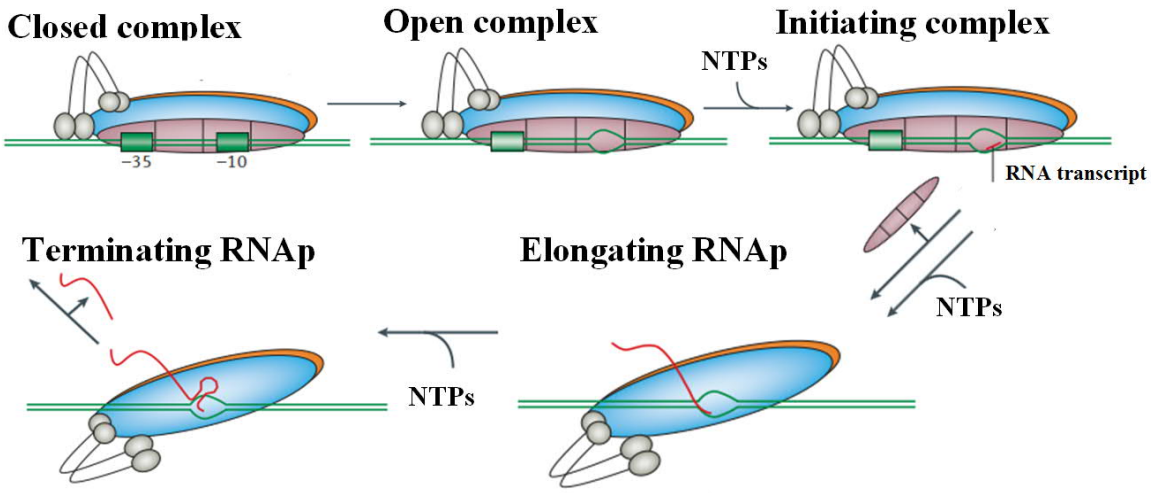


Figure 2.5: The transcription process in *E. coli*. The RNAp holoenzyme, binds to the promoter forming the closed complex, followed by the formation of the open complex where the double strand DNA is unwound. Next, the synthesis of RNA transcripts starts, allowing the transition to the initiating complex. The RNAp then goes into the elongation phase, leading to the release of the σ factor and elongation of the RNA transcript. Transcription then proceeds until the RNAp finds a terminator, after which the RNA transcript is released and the RNAp dissociates from the template. Adapted from [Nature Reviews Microbiology] (Browning and Busby 2016), copyright (2016).

2.2.1 Transcription Initiation and its Rate-limiting Steps

Transcription initiation in *E. coli* has been described as a complex multi-stepped process, which includes three main steps: binding, isomerization and promoter clearance (McClure 1985; Saecker *et al.* 2011).

In *E. coli*, the promoter region is defined by the presence of the consensus sequence at -10 (TATAAT) and -35 (TTGACA) regions, upstream of the transcription start site (TSS) (Harley and Reynolds 1987; Cho *et al.* 2009). This highly conserved sequence is necessary for the RNAp holoenzyme to recognize and bind the promoter region (von Hippel *et al.* 1984; Wang *et al.* 2011).

Following the binding, the holoenzyme makes the gene segment accessible for reading, by unwinding the DNA's double strands and exposing a small region in each of the strands. The consequent isomerization into the open complex has been shown to be a sequential process, with at least three intermediate steps: DNA loading, DNA opening and assembly of the polymerase clamp (Figure 2.6) (Saecker *et al.* 2011).

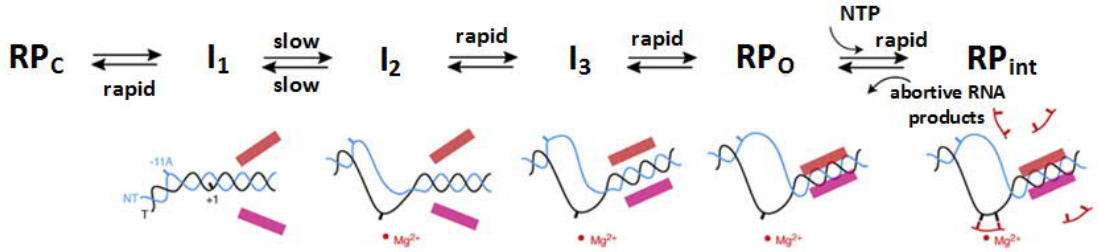


Figure 2.6: Summary of the intermediate steps involved in isomerization that form the initiating complex (RP_{init}), following the closed complex formation (RP_C). This closed complex formation triggers several conformational changes (I_1, I_2, I_3), eventually leading to the open complex formation (RP_O). Reprinted from (Saecker *et al.* 2011), with permission from Elsevier.

Before committing to elongation, the holoenzyme goes through a series of abortive initiation events (Hsu 2002). As a result, small transcripts of approximately 10 nucleotides are synthesized. After the initial RNA synthesis and successful escape from the promoter, the holoenzyme enters the elongation phase.

It has been suggested that the transcription dynamics of several *E. coli* genes is mainly controlled at the initiation stage (Browning and Busby 2004; Saecker *et al.* 2011). The main reasons for this is that most of the regulatory molecules of transcription act at this stage and this is the step that appears to be longer-lasting (McClure 1985; Browning and Busby 2004).

Most studies on the dissection of the regulating steps in transcription initiation have been conducted *in vitro* using two methods: the abortive initiation assay and the *in vitro* transcription assay (McClure *et al.* 1978; McClure 1980; Buc and McClure 1985; Lutz *et al.* 2001). The abortive initiation assay is based on the binding of the first two triphosphates, Adenosine Triphosphate (ATP) and Uridine Triphosphate (UTP), in an RNA sequence, in a condition where saturating quantities of RNAp and promoter DNA are used. The first nucleotide is always ATP, followed by UTP. The resulting products of the phosphodiester bond formation, pppApU and PPi, dissociate

due to the absence of other triphosphates required for the synthesis of RNA, thus aborting the initiation process. The resulting steady state conversion of ATP and UTP into pppApU is the basis for the abortive initiation assay (McClure *et al.* 1978).

From measurements of the abortive initiation assay, the kinetics of the intermediate steps can be derived. For instance, the rate of the open complex formation can be determined by the delay to reach the steady-state production of the abortive product (McClure *et al.* 1978; McClure 1980). During such experiments, a promoter-specific lag time was observed, before reaching the steady state. Given its sequence-specificity, this lag time was assumed to be the time taken for the RNAp to bind to the promoter and form the closed complex. In addition, the closed complex formation varied with different concentrations of RNAp. This dependence of the closed complex formation on the RNAp concentration, allows it to be distinguishable from the open complex formation (Buc and McClure 1985).

Importantly, also based on this dependence, from *in vitro* measurements with differing RNAp concentrations in the reaction vessel, it is possible to draw a 'tau (τ)-plot', which depicts a direct relationship between RNA production lag times and the inverse of the RNAp concentration (McClure 1980). From this plot, the slope equals the mean time for the closed complex formation, while the intercept with the y axis is the mean time for the open complex formation.

When compared with the time required for elementary steps in enzyme catalyzed reactions, the observed time lags were found to be significantly longer, lasting up to several minutes. Thus, these steps were considered to be the rate-limiting ones for transcription initiation (McClure 1980; Buc and McClure 1985).

While at optimal temperatures only two rate-limiting steps were found, in (Buc and McClure 1985) it was reported that, for the LacUV5 promoter, a third rate-limiting step emerges at temperatures lower than 20^0C (Figure 2.7). This third step, which is very fast at 37^0C, was hypothesized to be the unstacking of the DNA immediately before the completion of the open complex formation, that is, the isomerization step (Buc and McClure 1985).

Subsequently to these early studies, a detailed study on the rate-limiting steps in transcription initiation showed the intrinsic regulation of these steps by the promoter sequence. To show this, Lutz and colleagues (Lutz *et al.* 2001) engineered several different promoters derived from *E. coli*'s Lac promoter, by modifying the -10 and -35 regions.

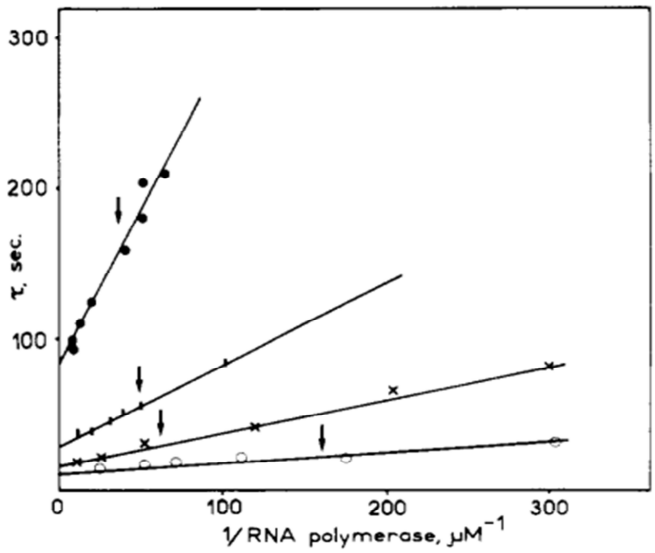


Figure 2.7: A τ-plot depicting the effect of temperature on the kinetics of the LacUV5 promoter. The time for the open complex formation is plotted against the inverse of the RNAp concentration, for all the temperatures tested: 19^0C (●), 25^0C (I), 30^0C (**x**) and 37^0C (○). Reprinted with permission from (Buc and McClure 1985). Copyright (1985) American Chemical Society.

The results showed that, in the presence of LacI, the closed complex formation was the most rate-limiting step, while AraC affected the three rate-limiting steps. For some promoters it accelerated the formation of a stable closed complex, while for other promoters its main impact was on the conversion of the closed to the open complex (Lutz *et al.* 2001).

When compared with *in vitro* studies, the *in vivo* measurement of the rates of transcription initiation are more difficult to perform because there is a limit in how much the *in vivo* intracellular concentration of RNAp can be changed, without affecting the cell functionality.

To cope with this, initially, a different approach was taken. Namely, the rate constants of the underlying steps of transcription were derived by statistical analysis from distributions of time intervals between consecutive RNA productions in individual cells, which were obtained by *in vivo* single-cell, single-RNA levels measurements for several promoters (Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016).

Similar to *in vitro* studies, these studies proposed that the distribution of time intervals could not be explained by the existence of a single step. They also showed that, in general, the models that best fit the data contain two main rate-limiting steps (associated with the closed and open complex formation), and that other steps become rate-limiting for temperatures as low as 24^0C (Muthukrishnan 2014). Based on this, it was suggested that the dynamics could be explained by the multiple rate-limiting steps model in transcription, in line with past results in the studies using *in vitro* techniques (McClure 1985; Lutz *et al.* 2001).

More recently, techniques were developed that allow altering sufficiently the RNAp concentration in live cells so as to 'recreate' *in vivo* (Lloyd-Price *et al.* 2016) the early *in vitro* measurements (McClure *et al.* 1978; McClure 1980; Buc and McClure 1985).

Overall, the above studies suggest that the mean rate of transcription of a gene is determined by the promoter sequence and the regulatory molecules which accelerate or hamper the steps involved in transcription initiation. Additionally, the kinetics of the steps involved in transcription initiation are not immune to environmental factors such as temperature.

2.3 Regulation of Transcription

Under normal conditions, the number of proteins and RNA molecules inside a cell are tightly controlled, and this appears to depend on the cell's ability to control the frequency by which a gene is expressed, rather than its degradation, which is kept at a near constant rate to allow for renewal and error correction (Willetts 1967; Goldberg 1972; Bernstein *et al.* 2002). In *E. coli* the critical regulatory steps modulating gene expression occur at the transcription level (Young *et al.* 2002; Browning and Busby 2004, 2016). Several mechanisms responsible for this regulation have been identified, with the majority involving a direct interaction with the promoter (von Hippel *et al.* 1984; Reznikoff *et al.* 1985; Browning and Busby 2004, 2016).

Given that the promoter sequence can determine the RNAp binding affinity, this can affect the rate of closed complex formation of a promoter, as shown in previous studies (McClure 1980; Buc and McClure 1985). This shows that the promoter sequence also plays a critical role in the regulation of transcription (von Hippel *et al.* 1984; Reznikoff *et al.* 1985).

Although this sequence-based type of regulation is relevant for the overall modulation of transcription, this only provides a "fixed" type of regulation, as it cannot be tuned according to environmental conditions. Thus, most regulation based on environmental changes involves components responsive to these changes and that act accordingly (von Hippel *et al.* 1984; Browning and Busby 2004; 2016).

2.3.1 Transcription Factors and Inducers

Transcription factors are one of the most important components in transcription regulation, given their ability to sense environmental fluctuations and modulate gene expression accordingly (Babu and Techmann 2003; Browning and Busby 2016). These proteins usually have two domains, where one receives internal/external signal, while the other interacts with the DNA (Babu and Teichmann 2003).

Upon binding to a promoter, a transcription factor may limit the interaction of the RNAp with the promoter, thus hampering transcription. These are known as repressors. Several mechanisms have been identified by which they are able to halt transcription (Figure 2.8A). For some promoters this occurs by steric hindrance, where the operator site overlaps with the -10 and -35 elements recognized by the RNAp, thus the binding of the repressor to the operator site blocks the RNAp access to the promoter region (Browning and Busby 2004). Repression can also be achieved by DNA looping, where repressor molecules bind simultaneously to operator sites located upstream and downstream of the promoter region inducing the formation of a DNA loop, also preventing the RNAp to access the promoter region (Schleif 2010; Browning and Busby 2016). For other promoters, the repressor can act as an "anti-activator", thus preventing activator molecules from "activating transcription" (Browning and Busby 2016).

Additionally, transcription factors can also actively recruit RNAp molecules to the promoter region, thus activating transcription. Here, they are called activators and there are several mechanisms by which they can activate the transcription process (Lee *et al.* 2012) (Figure 2.8B). In Class I activation, the activator binds to an operator located upstream of the promoter region, and recruits the RNAp by interacting with the C-terminal domain of its α-subunit (Busby and Ebright 1999; Lee *et al.* 2012). In Class II activation, the operator region where the activator binds overlaps with the -35 element, thus the activator recruits the RNAp by interacting with the domain 4 of the sigma factor (Lee *et al.* 2012; Browning and Busby 2016). Transcription activation can also occur through conformational changes on the promoter DNA. Here, the activator binds to the operator site located at or near the -35 and -10 elements, and rearrange them, so that they are better positioned for the binding of the RNAp (Lee *et al.* 2012; Browning and Busby 2016).

Given that the interaction between transcription factors and promoter is dependent on the promoter sequence and the regulatory protein structure (Babu and Teichman 2003; Browning and Busby 2004, 2016), this type of mechanisms allows the cells to diversify their gene expression profile.

Some transcription factors can be affected by specific molecules called inducers. For instance, these molecules can inactivate repressors, thus promoting transcription. One of the best known examples is the repressor of the *lac* operon, LacI, which is inactivated by its inducer, lactose (Jacob and Monod 1961). Other inducers can promote gene expression by increasing the functionality of its activators or by inverting the repressor function, turning it into an activator, as in the case of AraC, the repressor of the Arabinose operon (Englesberg *et al.* 1965; Schleif 2010).

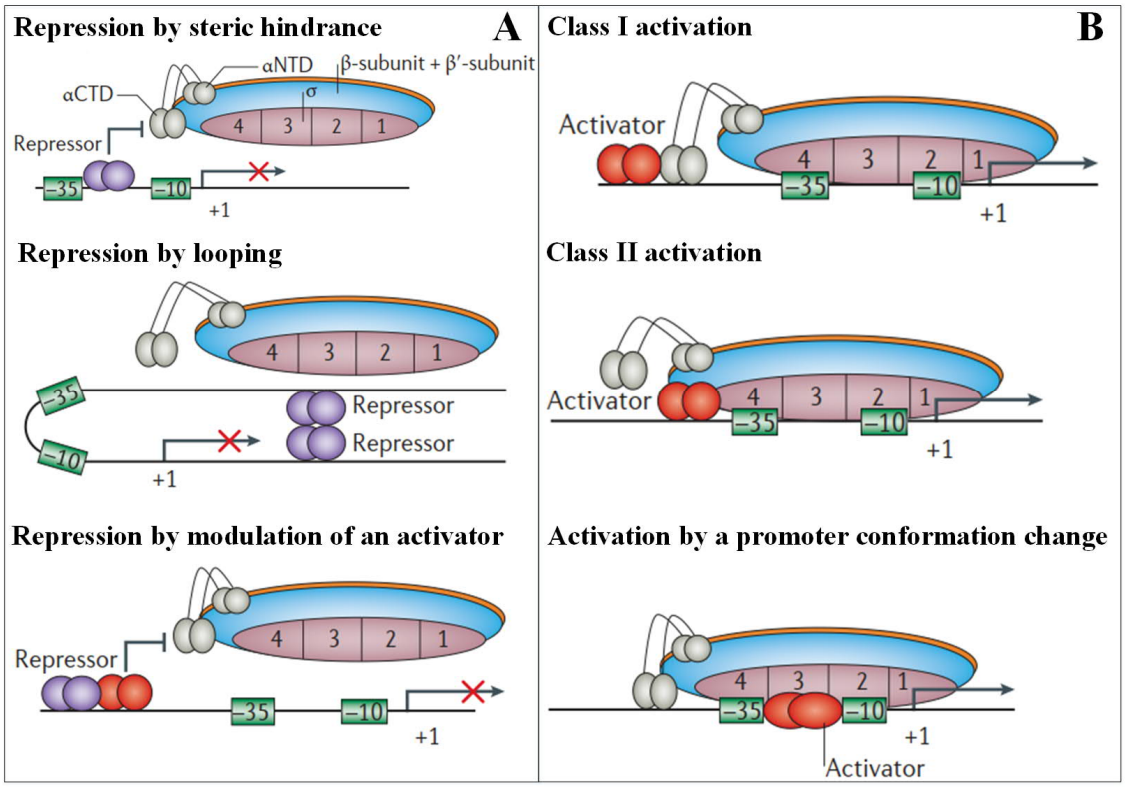


Figure 2.8: Examples of repression and activation at promoters by transcription factors. Represented are several mechanisms by which these transcription factors are able to repress (A) and activate (B) transcription. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Microbiology] (Browning and Busby 2016), copyright (2016).

Several gene expression inducers are not synthesized by the cell, but rather enter the cell from the environment. For this to occur, these molecules must travel through the cell membrane. The *E. coli*'s membrane (Figure 2.9) consists of two layers, the outer and the inner membrane, separated by the periplasmic space (Zimmermann and Rosselet 1977; Alberts *et al.* 2002). The outer membrane is semi-permeable, allowing the crossing of lipophilic, and some small uncharged molecules (Finkelstein 1976; Willey *et al.* 2008), while preventing larger and ionic molecules from crossing (Decad and Nikaido 1976). On the outer membrane, transmembrane proteins, known as porins, act as channels for the entrance/exit of some specific molecules (Alberts *et al.* 2002; Nikaido 2003; Willey *et al.* 2008). Along with the lipid bilayers, these proteins are important for the dynamics of intracellular level of inducer molecules, and consequently the gene expression response to their presence in the environment.

The lactose intake system is one of the most studied mechanism for inducer intake (Jensen *et al.* 1993; Marbach and Bettenbrock 2012). The majority of these studies were conducted in a regime of low concentrations of lactose, where the rate of the target gene production presents an almost linear dependence on the intracellular level of inducer (Jensen *et al.* 1993).

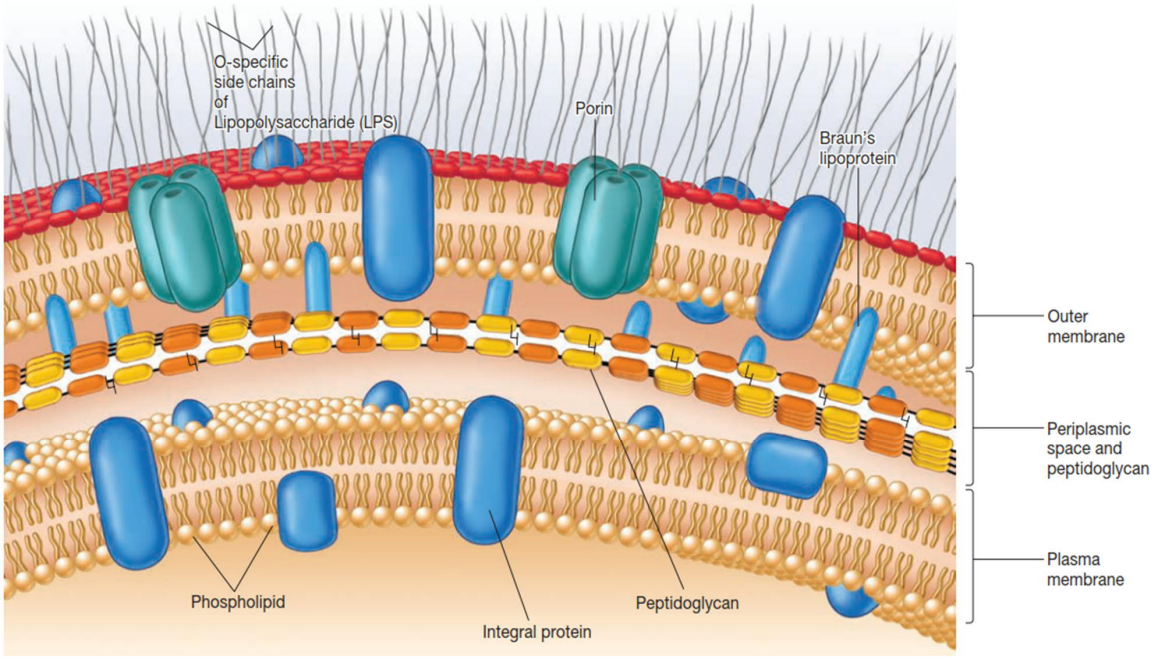


Figure 2.9: An illustration of *E. coli*'s membrane. The membrane consists of two lipid bilayers, the outer and the inner membrane, which are separated by the periplasmic space and peptidoglycan. Bound to the membrane are porins and other transmembrane proteins. Reproduced from (Willey *et al.* 2008), with permission from McGraw-Hill Education, copyright (2008).

At these concentrations (less than 0.25 mM), it was found that the cellular intake of lactose had a positive feedback, due to the activity of the lactose permease, LacY (Jensen *et al.* 1993; Ozbudak *et al.* 2004; Marbach and Bettenbrock 2012). This is a proton symporter protein, that uses the transmembrane proton gradient to simultaneously transport protons and lactose (Ramos and Kaback 1977; Kaback 1983). Upon the entrance of lactose molecules in the cell, the lactose operon is activated (see section 2.3.2 for details), producing LacY proteins, which in turn lead to an increase in the intake of lactose molecules (Jensen *et al.* 1993; Ozbudak *et al.* 2004). At higher concentrations of lactose, the role of LacY is no longer significant, with the lactose molecules entering the cell through alternative symporters and potentially through passive diffusion (Decad and Nikaido 1976; Jensen *et al.* 1993).

At high concentrations, the kinetics of inducer intake is less studied, mainly because under this condition, the activity of the target gene is close to full induction, which no longer reflects the intracellular changes in the inducers level. In **Publication I**, to address this issue we implemented a method to characterize the intake of IPTG in cells lacking the LacY protein, using *in vivo* single RNA measurements, at the single RNA and single cell level (Golding *et al.* 2005).

Another factor that can influence the intake of inducer molecules is temperature fluctuations, which is known to affect several cellular processes in *E. coli* (Yamanaka 1999; Arsène *et al.* 2000).

For instance, it can alter the functionality of proteins (Arsène *et al.* 2000) and also change the physical properties of the cell membrane and cytoplasm (Sinensky 1974; Yamanaka 1999; Oliveira *et al.* 2016b). Changes in these properties are expected to have severe effects on the intake kinetics of inducers. To address this, in **Publication IV** we assessed the temperature dependence of the intake process of the inducer IPTG at different temperatures using *in vivo* single RNA measurements techniques.

2.3.2 The Lac Operon

The lactose operon was the first gene regulatory network to be described (Jacob and Monod 1961). The concept of operon was developed by Jacob and Monod (1961), based on observations of bacterial cultures containing different types of sugar, where they found that under these conditions, *E. coli* preferably uses glucose as its primary source of carbon. However, when glucose is depleted from the media, *E. coli* would use the other sugars available. An operon was then described as a unit of gene expression, where structural genes are coordinately regulated, allowing for that choice to be made.

The structural genes of the operon code for proteins responsible for carrying out specific tasks and their regulation depends on the metabolic needs of the cell. The repressor molecule, which is also produced by a regulatory gene (R), regulates the production of the structural genes (A, B) by interacting with a regulatory element, known as the operator (O) (Figure 2.10).

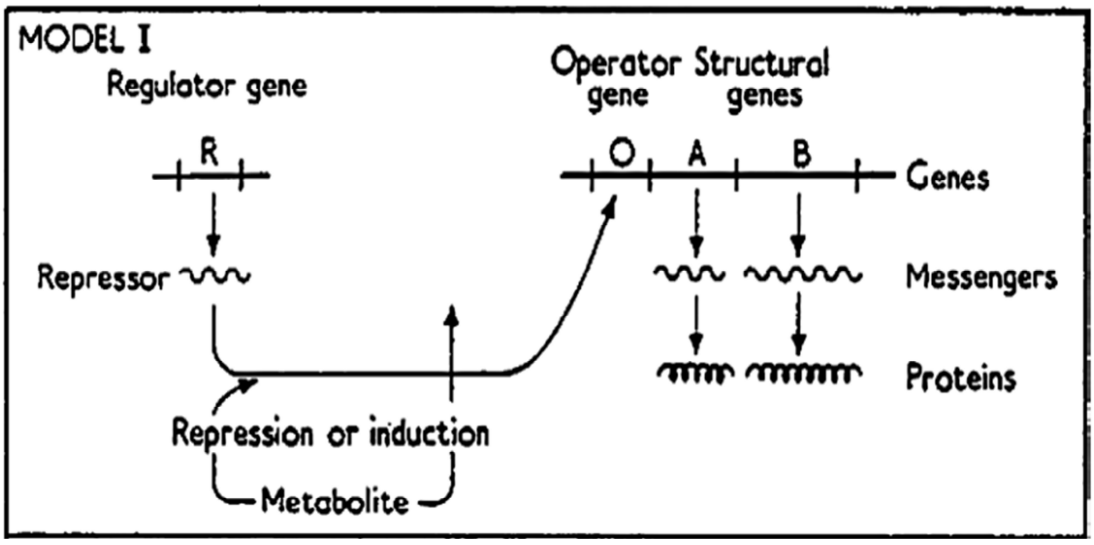


Figure 2.10: Diagram of the operon model described by Jacob and Monod (1961). Reprinted from (Jacob and Monod 1961), with permission from Elsevier.

This interaction between the repressors and the operator blocks or negatively regulates the expression of the structural genes. However, for the repressor molecule to regulate the operon, it must be

inducible, that is, it must be able to turn on or off in response to a particular chemical signal, here known as the inducer (I). Aside from binding to the operator, the repressor molecule is also able to bind to the inducer, which in turn modulates the repressor's affinity for the operator site. When the inducer is available in the media, the repressor dissociates from the operator, thus allowing the expression of the structural genes to occur.

Lactose, a disaccharide sugar, gave name to the *lac* operon. When this sugar is used by *E. coli* as its primary source of carbon, the lac repressor is induced, which in turn allows the expression of the three structural genes, responsible for lactose metabolism (Jacob and Monod 1961) (Figure 2.11).

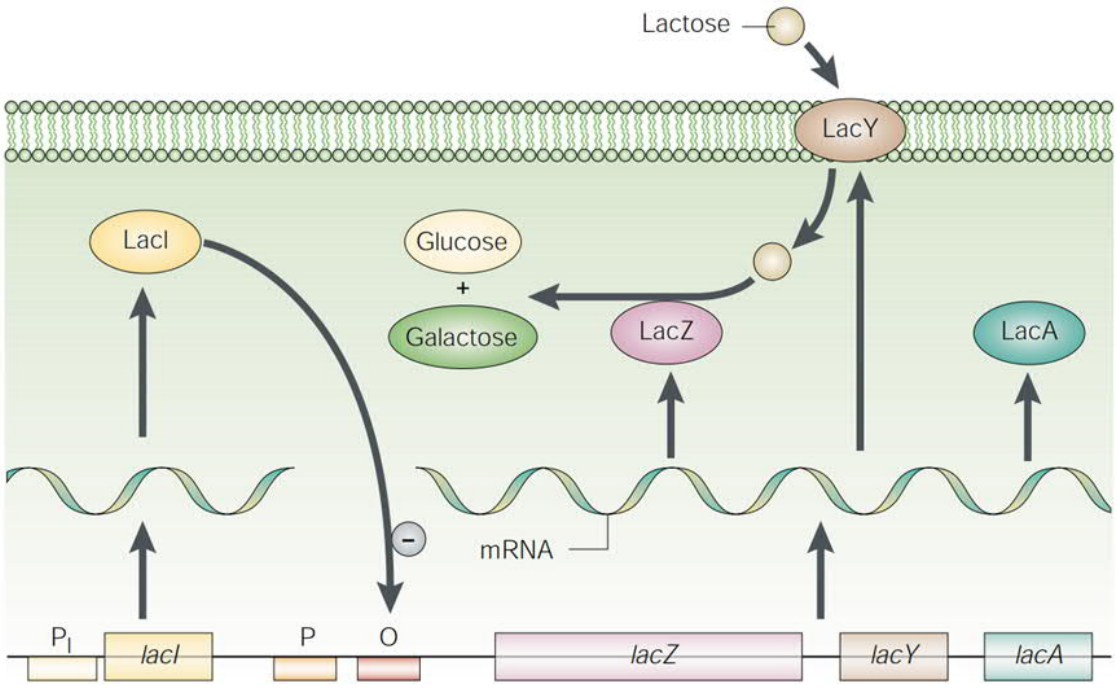


Figure 2.11: Illustration of the lactose operon, from *E. coli*. The three genes responsible for lactose metabolism (*lacZ*, *lacY* and *lacA*) are organized in a cluster, called the lac operon. The coordinated transcription and translation of the lac operon structural genes is regulated by a shared promoter, operator and terminator. The regulator gene (*lacI*) is located upstream of the operon. Reprinted by permission from Macmillan Publishers Ltd: [Nature Reviews Genetics] (Shuman and Shilhavy 2003), copyright (2003).

The structural genes regulated by the lac repressor (LacI) are the *lacZ*, *lacY* and *lacA* and they code for the proteins involved in the metabolism of lactose, namely β-galactosidase, lac permease, and a transacetylase, respectively (Jacob and Monod 1961). β-galactosidase is responsible for breaking lactose into galactose and glucose. This is the first step of the lactose metabolism. Lac permease, a transmembrane protein, is responsible for the transportation of lactose molecules into the cell, and transacetylase transfers an acetyl group from coenzyme A (CoA) to the hydroxyl group of the galactosides (Jacob and Monod 1961; Lewis 2005).

Aside from structural and regulatory proteins, the *lac* operon contains three operator sites, namely O_1, O_2 and O_3. The O_1 site is located immediately downstream of the *lac* promoter, O_2 is located downstream of the *lac* promoter and within the coding region of the *lacZ* gene, and the O_3 is located upstream of the *lac* promoter (Jacob and Monod 1961; Riggs *et al.* 1970; Reznikoff *et al.* 1974; Schlax *et al.* 1995) (Figure 2.12).

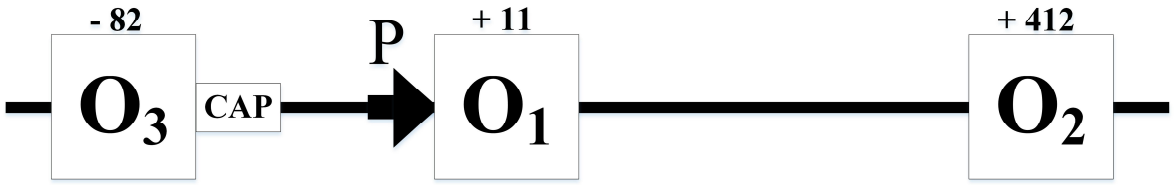


Figure 2.12: Illustration of the operators of the lac operon, O_1, O_2 and O_3. Shown are their location and the CAP binding site, relative to the TSS. Reprinted from (Lewis 2005), with permission from Elsevier.

Additionally, the cAMP receptor protein (CRP) or catabolite activator protein (CAP) binding site is located upstream of the promoter region, in close proximity with the O_3 site (Beckwith *et al.* 1972; Reznikoff *et al.* 1974). LacI is the negative regulator of the *lac* operon (Jacob and Monod 1961). Contrary to LacI, the secondary messenger cAMP, positively regulates the transcription of the *lac* operon (Emmer *et al.* 1970). Jacob and Monod observed that if glucose and lactose are present, cells will not use lactose until glucose has been depleted from the media. Under these conditions, the intracellular levels of cAMP are low, and as glucose concentration decreases, due to consumption, the intracellular levels of cAMP increase. This leads to the formation of a complex between cAMP and the CRP protein, increasing its ability for binding to the CRP binding site. This in turn activates the transcription of the *lac* operon, by increasing the affinity of the RNAp for the *lac* promoter (Zubay *et al.* 1970; Griffiths *et al.* 2004).

The LacI protein consists of two DNA-binding dimers, which forms a dimer of dimers, resulting in a homotetramer (Beyreuther *et al.* 1973; Gilbert and Maxam 1973; Kania and Brown 1976; Miller 1980). The dimers interact with the operator, and each monomer binds, with equal affinity, one inducer molecule (Ohshima *et al.* 1974; Schmitz *et al.* 1976; Lewis 2005). When LacI binds its natural inducer allolactose (Jobe and Bourgeois 1972) or an artificial inducer such as IPTG (Jacob and Monod 1961), its conformation changes, which in turn reduces its affinity for the operator (Barkley *et al.* 1975; Lewis *et al.* 1996; Lewis 2005). This, along with the cAMP/CRP binding, leads to the transcription of the structural genes of the *lac* operon (Zubay *et al.* 1970; Griffiths *et al.* 2004).

When there is no lactose in the media, LacI binds with high affinity to the O_1 operator, thus preventing the expression of the genes responsible for lactose catabolism (Jacob and Monod 1961; Riggs *et al.* 1970; Schlax *et al.* 1995). Initially, it was assumed that the mechanism by which LacI

prevents the transcription of the *lac* operon was through steric hindrance, where the binding of LacI to the O_1 site would physically block the RNAp movement (Schmitz and Galas 1979). Although two additional *lac* operators sites were identified, due to their low affinity for LacI and their distance from the promoter region, it was assumed that they had no significant contribution in the repression process (Pfahl *et al.* 1979). However, it was later shown that the 'full repression' of the *lac* operon requires both O_2 and O_3 operators (Oehler *et al.* 1990). This full repression is achieved by the simultaneous binding of LacI molecules to the O_1 site, and to either O_2 or O_3. This causes the formation of a DNA loop that blocks transcription (Oehler *et al.* 1990) (Figure 2.13).

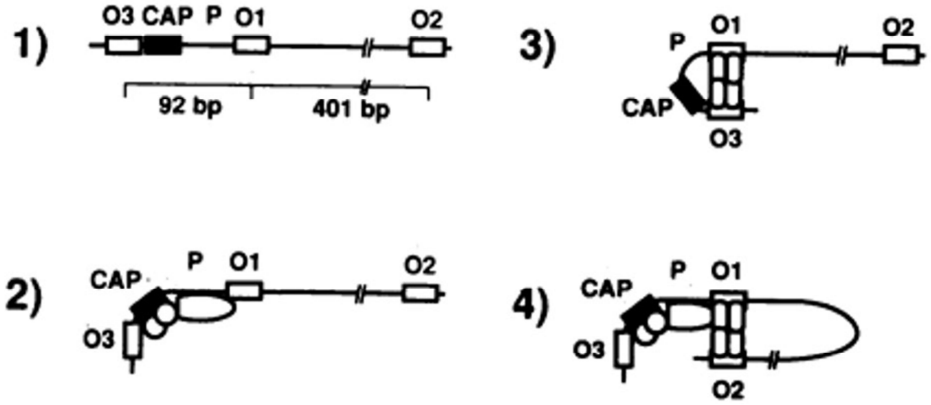


Figure 2.13: A representation of the *lac* operon repression, by LacI. 1) Shows the location of the operators and the CAP binding sites, and their relative distance to the O_1 site. 2) In the absence of LacI, the binding of the CAP protein activates transcription. In the absence of any inducer molecules, the LacI tetramers form a DNA loop, by binding simultaneously to 3) O_1 and O_3 or 4) O_1 and O_2 (Oehler *et al.* 1994). Reproduced with permission from EMBO.

It has been shown that in the LacUV5 promoter the loop formation halts transcription by preventing the RNAp to access the promoter region, positioned within the loop (Becker *et al.* 2013). Using different combinations of operators, it has been also shown that when the O_2 site is removed, the repression strength is slightly reduced when compared to the wild-type *lac*, containing all three operators (Oehler *et al.* 1990, 1994).

The *lac* promoter has been widely used in synthetic and molecular biology studies (Studier and Moffatt 1986; Elowitz and Leibler 2000; Golding and Cox 2004). As studies have shown, the full repression of the *lac* promoter is achieved when the three operators are present (Oehler *et al.* 1990, 1994). Given that most *lac* derivatives promoters used in molecular biology carry only one operator or a combination of two operators, it is expected that their repression mechanism would be less efficient, thus presenting some level of leaky transcription.

Leakiness is defined as transcription events occurring in the absence of an inducer, due to a failure in the repression mechanism. For an inducible promoter, this type of behavior is usually not desirable. Several studies have addressed the issue of engineering promoters with reduced levels of leakiness (Lanzer and Bujard 1988; Penumetcha *et al.* 2010), but to date little is known about the effects of sub-optimal temperatures on the leakiness levels of genes.

In **Publication III**, we used a variant of the *lac* promoter containing the O_1 and the O_3 operator sites ($LacO_3O_1$ promoter) (Oehler *et al.* 1990) to investigate the leakiness rate of this promoter, at sub-optimal temperatures. Using single cell, single molecule techniques, we conducted measurements in the absence of inducers, and with the support of mathematical models, we analyzed the data and dissected which factors contributed the most to the observed leakiness at sub-optimal temperatures.

2.3.3 The Arabinose Operon

In *E. coli*, L-Arabinose can be used as a source of carbon and energy by the arabinose system (Helling and Weinberg 1963; Englesberg *et al.* 1965; Schleif 2010). This system consists of genes responsible for the uptake and catabolism of L-Arabinose, which upon entrance in the cell is converted into D-xylose-5-phosphate, thus entering the pentose phosphate pathway (Schleif 2010) (Figure 2.14).

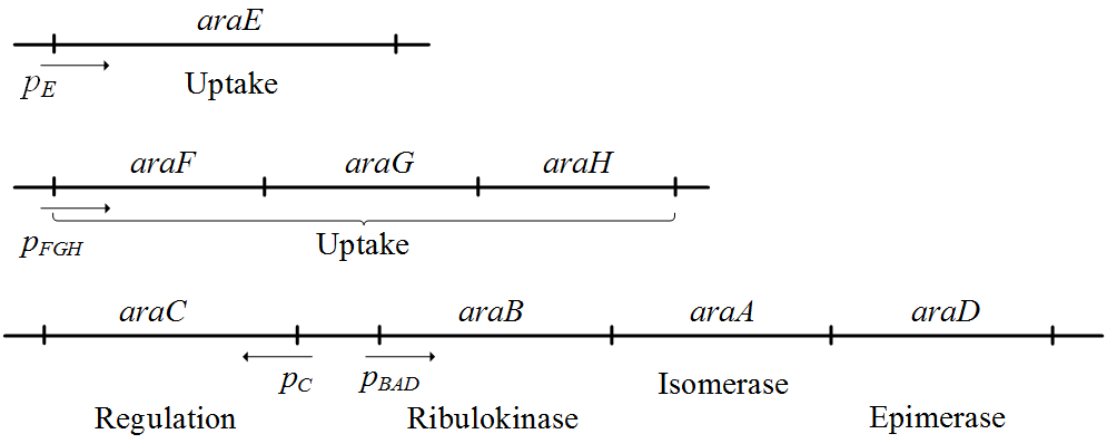


Figure 2.14: The L-Arabinose system of *E. coli*. The genes presented in this illustration code for proteins required for the uptake and catabolism of L-Arabinose. Reprinted from (Schleif 2010), with permission from Oxford University Press.

The L-Arabinose system has two different transport mechanisms responsible for the uptake of L-Arabinose. In the lower affinity transport system, the AraE transporter protein, bound to the inner membrane, uses the electrochemical potential to transport L-Arabinose (Lee *et al.* 1981; Schleif 2010). The AraFGH is part of the high affinity transport system, which is an L-Arabinose specific

ABC transporter using the energy from the hydrolysis of ATP to transport L-Arabinose (Hogg and Englesberg 1969; Schleif 2010) (Figure 2.15).

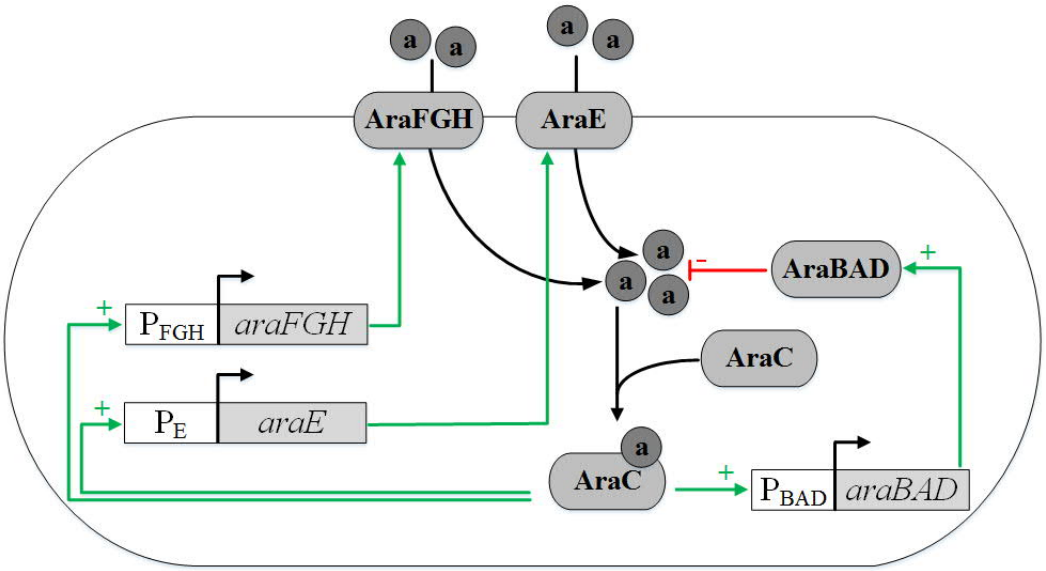


Figure 2.15: The L-Arabinose system in *E. coli*. Represented are the genes responsible for the uptake (*araFGH*, *araE*), metabolism (*araBAD*) and the regulator of the L-Arabinose system (*araC*). Once inside the cell, L-Arabinose binds to AraC, and this complex activates P_E, P_{FGH}, P_{BAD}, expressing the proteins araE, araFGH and araBAD, respectively. The arrows indicate the transport of L-Arabinose, through a negative (red) and positive (green) feedback. Adapted from (Megerle *et al.* 2008), with permission from Elsevier.

AraC is the regulatory protein of the Arabinose system in *E. coli*. It is a dimeric protein that regulates the expression of all proteins in the arabinose system, by activating and repressing the genes responsible for the uptake and catabolism of L-Arabinose (Englesberg *et al.* 1965; Johnson and Schleif 1995; Schleif 2010). In the absence of Arabinose, AraC is bound to the I_1 and O_2 sites, forming a DNA loop, which prevents the RNAp to access the promoter region (P_C and P_{BAD} promoters) and initiate transcription (Lobell and Schleif 1990; Johnson and Schleif 1995; Schleif 2000, 2010).

When L-Arabinose is present in the cell, AraC binds to the I_1 and I_2 half-sites, recruiting the RNAp and thus activating the transcription initiation at the P_{BAD} promoter (Schleif 2000, 2010) (Figure 2.16).

In **Publication I** and **Publication IV**, to study the intake kinetics of IPTG, we used the Lac/ara-1 promoter, which is a synthetic promoter based on the combination of the Lac and AraBAD promoters, from *E. coli* (Lutz and Bujard 1997). To construct this promoter, a mutant Lac promoter, the $P_{lac\text{-}8A}$ which has Thymine replaced by Adenine in the -8 position, was used as the basic structure.

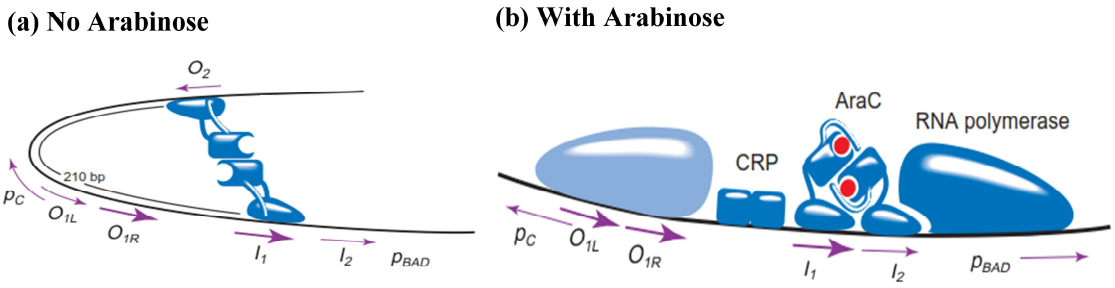


Figure 2.16: Illustration of the regulatory regions of the *araCBAD* genes. I_1 and I_2 are named half-sites, given that only a single subunit of AraC contacts each of them. O_1 consists of two-half-sites and it is an operator for the P_C promoter, which controls the expression of the AraC, while O_2 is a single half-site. The CAP binding site serves both the divergently orientated P_C and P_{BAD} promoters. Reprinted from (Schleif 2000), with permission from Elsevier.

To this sequence, some modifications were eventually made. Namely, the introduction of a symmetrical lac operator sequence (Os) (Sadler *et al.* 1983; Lanzer and Bujard 1988), and the insertion of the wild type sequence of the *lacO1* operator upstream of the promoter region. Additionally, the CRP/cAMP binding site, of the Lac promoter, was replaced by the I_1/I_2 operator sites for the AraC protein. The resulting promoter is inducible by both L-Arabinose and IPTG.

2.4 Gene Expression in T7 Phage

The bacteriophage T7 is an obligate lytic phage that infects *E. coli* and uses this host's resources to produce its progeny. T7 genome consists of a single double-stranded DNA molecule, enclosed within an icosahedral head attached to a short non-contractile tail (Molineux 2005). Sequencing results have shown that it has a 39,937 bp genome coding for 56 genes which are organized into three classes, based on their location, function and the order by which they are transcribed (McAllister and Wu 1978; Dunn and Studier 1983) (Figure 2.17).

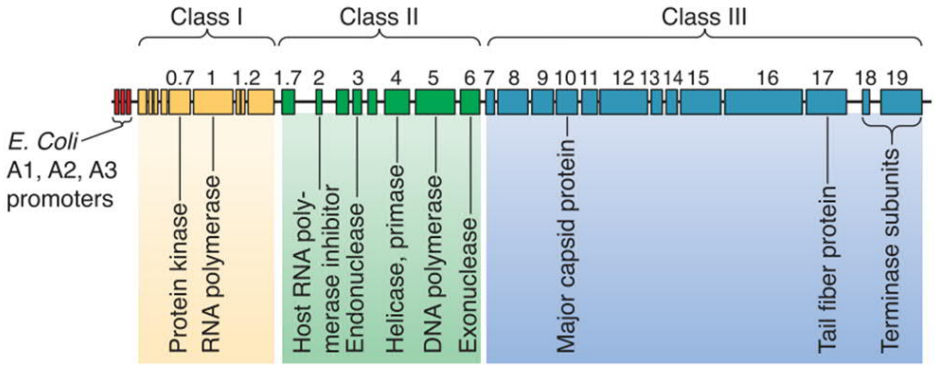


Figure 2.17: The genetic map of the T7 phage. The DNA is depicted as a black horizontal line, while the colored boxes represent the genes and their promoters. In addition, the major proteins are indicated

in the map. Reused from (Qimron *et al.* 2010), with permission from the American Society for Microbiology.

Upon infection, Class I genes are the first to be transcribed, and are responsible for creating favorable conditions for the phage growth (McAllister and Barrett 1977; Dunn and Studier 1983). Class II genes are next and they synthesize DNA replication proteins (McAllister and Wu 1978; Dunn and Studier 1983). The Class III genes, the last to be transcribed, codes for structural proteins involved in viral DNA packaging, virion assembly and lysis functions (McAllister and Wu 1978; Dunn and Studier 1983).

2.4.1 Transcription Initiation in T7 phage

In T7 phage, transcription is unidirectional, from left to right, with the complete process requiring two different RNAps. The Class I genes are transcribed by *E. coli*'s RNAp, while Class II and III are transcribed by T7 RNAp. This is a highly specific polymerase, and unlike *E. coli*'s RNAp, it does not require any additional co-factor for transcription to occur (Chamberlin *et al.* 1970; Chamberlin and Ring 1973; Dunn and Studier 1983). The production of T7 RNAp early during the infection process and the inactivation of *E. coli*'s RNAp, efficiently directs all gene expression resources towards T7's own DNA (McAllister and Barrett 1977; Dunn and Studier 1983).

The two set of T7 genes, have specific and time-dependent roles during infection, thus gene expression has to occur in a way that ensures accurate timing (Chamberlain *et al.* 1970; Dunn and Studier 1983). Here, this accuracy is achieved by coupling the T7 genome entrance into *E. coli* cells, with transcription (Zavriev and Shemyakin 1982; Moffatt and Studier 1988; Garcia and Molineux 1995, 1996). Upon infection, after the initial entrance of the first 850 bp of T7 DNA, the transcription process causes the internalization of the remaining genome (Zavriev and Shemyakin 1982; Moffatt and Studier 1988; Garcia and Molineux 1995, 1996). First, *E. coli* RNAp starts transcribing at its promoters, on the initial 850 bp at a rate of 40-60bp/second, pulling up to 7.5 kb of T7 genome into the cell (Garcia and Molineux 1995). Once T7 RNAp is produced, it starts transcribing its genes, thus pulling the remainder of the T7 genome, at a rate of 200-300bp/second (Garcia and Molineux 1995). This RNAp-mediated DNA entrance also couples gene expression with gene position, meaning that a gene cannot be expressed until it enters the cell.

The T7 RNAp mediated transcription initiation is similar to that of *E. coli*'s RNAp, thus occurring in several steps. First, the RNAp interacts weakly with nonspecific regions of the DNA until the promoter region is found and recognized through specific protein-DNA interactions (Cheetham and Steitz 2000). T7 RNAp recognizes the promoter region through an interaction between the region -7 to -11 in the major DNA groove, and the T7 RNAp amino acids residues in a specific

loop (Muller *et al.* 1989; Li *et al.* 1996; Cheetham *et al.* 1999). This step is known as the closed complex formation. Subsequently, the promoter undergoes RNAp-induced isomerization and melting in the promoter region, forming the open complex (Ujvári and Martin 2000). Next, abortive RNA synthesis events occur, releasing short RNA chains, until the formation of a stable transcribing complex consisting of the nascent RNA transcript, the DNA template, and T7 RNAp. Following the promoter escape, a stable elongation complex is formed (Liu and Martin 2002). After the formation of this stable elongation complex, the gene is transcribed, and termination occurs once T7 RNAp reaches a terminator sequence (Arnaud-Barbe *et al.* 1998).

2.4.2 T7 RNAp Promoters

The two classes of T7 RNAp promoters consist of a highly conserved consensus sequence of 23 bp, extending from -17 to + 6, that can be further separated into two functional domains: an upstream binding region, from -17 to -5 and an initiation region from -4 to +6 (Chapman and Burgess 1987; Chapman *et al.* 1988; Ujvári and Martin 1997).

Studies have reported that transcription from T7 RNAp promoters is affected by factors known to disturb the stability of the DNA helix, such as ionic strength and temperature (McAllister and Carter 1980; Chapman and Wells 1982) and also that transcription from Class II, is weaker than that from Class III promoters (Golomb and Chamberlin 1974; McAllister and Wu 1978). To better understand which step in transcription initiation contributes the most for these apparent differences, Ikeda and colleagues (Ikeda 1992; Ikeda *et al.* 1992) conducted studies where they compared the transcriptional activity of several Class II and III promoters. In line with previous studies, the results showed that the Class II were generally weaker than Class III promoters (Ikeda 1992; Ikeda *et al.* 1992). In addition, they found that the observed differences in strength were not reflected in their maximum transcription rate (Ikeda *et al.* 1992), but rather in the efficiency of promoter clearance (Ikeda 1992). These results, along with computational simulation data, have suggested that the Φ10 promoter is the strongest of the T7 RNAp promoters (Ikeda 1992; Endy *et al.* 1997).

It is known that promoter sequence plays an important role in transcription regulation (Brewster *et al.* 2012; Browning and Busby 2004, 2016). Similar to *E. coli* promoters, T7 promoters carrying elements identical to the consensus sequence are considered to be stronger than those having less similarity (Dunn and Studier 1983; Ikeda 1992). This is the case for the Φ10 (Ikeda 1992; Endy *et al.* 1997). Although *in vitro* studies have shown that temperature affects the transcription from T7 RNAp promoters (McAllister and Carter 1980; Chapman and Wells 1982), currently there is little information on how temperature changes affects its *in vivo* transcription dynamics.

To assess this, in **Publication II** we used a single copy plasmid carrying the T7 Φ10 promoter and measured its RNA production as a function of temperature. Given that this is a constitutive promoter, we used the *E. coli* BL21(DE3) strain which carries a single copy of the gene coding for the T7 RNAp integrated in the chromosome, under the control of the LacUV5 promoter (Studier and Moffatt 1986). We then measured the transcription activity of this promoter, under different temperature conditions, using live-cell microscopy with single-cell, single-RNA sensitivity (Golding *et al.* 2005; Lloyd-Price *et al.* 2016).

2.5 Temperature Shifts and the Transcription of Synthetic Genes in *E. coli*

Microorganisms are able to adapt to changing environmental conditions by detecting changes in their surrounding and by responding accordingly to these changes. This adaptation process is usually characterized by a rapid response, which is necessary for the cell to initiate adaptation, followed by a more continuous response to create conditions for long term survival (Yamanaka 1999; Arsène *et al.* 2000; Gunasekera *et al.* 2008).

Temperature fluctuations require adaptive responses from the cells (Arsène *et al.* 2000; Yamanaka 1999; Barria *et al.* 2013). Some of these responses are relatively fast while others take longer to occur. The latter usually involve the expression of specific proteins. For instance, when there is a sudden up-shift in temperature, known as heat shock, the structure of several proteins is affected. To cope with this, the cell increases the expression level of heat shock proteins (HSPs), which are crucial for cell survival given that they act on preventing protein aggregation and in helping to refold misfolded proteins. Although most studies have focused on *E. coli*'s heat shock response, there are also reports on the changes occurring when cells face a downshift in temperature, known as cold shock. Similarly to heat shock, this process triggers the activation of a specific set of proteins, known as cold-shock proteins, which help the cell in counteracting the effects caused by this reduction in temperature (Yamanaka 1999; Phadtare 2004).

Meanwhile, fast responses include readjustments of the lipid composition of the membrane in a response known as homeoviscous adaptation (Sinensky 1974). This is crucial for the cell survival since temperature shifts lead to changes in the membrane fluidity (Yamanaka 1999; Phadtare 2004; Barria *et al.* 2013), thus affecting cellular processes such as the transportation of molecules in/out of the cell (Phadtare 2004; 2010). A recent study (Oliveira *et al.* 2016b) has reported that temperature downshift reduces the cytoplasm viscosity, which shows additional effects of temperature on the cell's functionality. These effects combined with compromised membrane functions, can affect the intake of molecules from the environment, as well as their movement and that of other cellular components, in the cell's cytoplasm. As a consequence, it can alter the timely expression of genes

requiring activation by external molecules, which can be translated in a delay of the molecule to enter and eventually reach the promoter region and activate the gene of interest.

The effect of temperature on transcription initiation and its rate-limiting steps has been addressed in several studies. Using *in vitro* techniques, Buc and McClure (1985) have demonstrated that for the LacUV5 promoter, a third rate-limiting step appears at temperatures lower than 20^0C. It was hypothesized that this step was isomerization, which was believed to have a short duration at 37^0C. These observations showed that the steps comprising the transcription initiation process respond differently to temperature changes, and that some are temperature dependent. A more recent *in vivo* study demonstrated that for the TetA promoter, a third rate-limiting step emerges at temperatures lower than 24^0C (Muthukrishnan 2014), which is in line with the previous report (Buc and McClure 1985). Unlike the behavior observed for LacUV5 and TetA promoters, another recent study has demonstrated that for the Lac/ara-1 promoter, temperature changes affect the closed complex formation, suggesting that the interaction between the RNAp and the TSS is the most affected by temperature (Oliveira *et al.* 2016a). The changes observed in the rate-limiting steps of these promoters, will also reflect in their rate of RNA production. Overall, these studies suggest that the dynamics of the rate-limiting steps of these promoters cause them to respond differently, in their RNA production kinetics, to temperature changes.

Given that gene expression regulation occurs mostly at the stage of transcription initiation, these studies provide a more detailed picture on how temperature changes affect gene expression, and which steps are more susceptible to these changes.

Synthetic biology is currently a particularly attractive field, that allows the engineering of simpler-than-natural synthetic gene networks, from well characterized genes and proteins (Endy 2005; Nandagopal and Elowitz 2011). Over the years, a number of successfully implemented synthetic gene circuits have been reported (Elowitz and Leibler 2000; Gardner *et al.* 2000; Stricker *et al.* 2008), however the stability and tunability of some of these circuits appears to be a major issue (Elowitz and Liebler 2000; Oliveira *et al.* 2015). This shows that even when the design of these networks is based on known components, there are additional factors that must be taken into consideration (Brophy and Voigt 2014). For instance, how do these components respond to environmental perturbations such as temperature fluctuations?

Different studies have addressed this question, by assessing the effects of temperature on the stability of a synthetic genetic circuit, such as the oscillator (Stricker *et al.* 2008), or the repressilator (Oliveira *et al.* 2015). In the latter, it was shown that when subject to changing temperature conditions, one of the components of the repressilator loses its functionality, hampering the circuit's

ability to track time (Oliveira *et al.* 2015). This loss in functionality occurred with increasing temperature and was shown to also affect the stability of a switch using the same protein as one of the components. Such studies are important in helping to identify which components of a circuit are more sensitive to temperature fluctuations.

Temperature can also be used as a means to control the activity of a synthetic genetic circuit, as reported in (Isaacs *et al.* 2003). Here, using a temperature sensitive variant of the repressor protein from the lambda (λ) phage (Arkin *et al.* 1998), it was shown that by varying temperature, the stability of this protein can be tuned, thus altering the level of activity of the circuit. Another study has reported the construction of a synthetic genetic clock exhibiting temperature compensation, meaning that the periodicity of the clock is not temperature dependent (Hussain *et al.* 2014). This was achieved by genetically modifying the repressor protein of the target promoter, so that it loses its repression ability at high temperatures. The resulting clock, exhibited a near constant oscillation period for temperatures ranging from 30^0C to 41^0C (Hussain *et al.* 2014).

These studies show how the temperature sensitivity of a component from a genetic circuit can have a significant effect on its overall stability. These synthetic genetic circuits are often constructed using well-characterized components, which are then inserted into host strains. Thus, factors that can affect the physiology or fitness of the host strain can also affect the activity/stability of these circuits. The choice of a promoter to control the expression of a gene remains one of the most important factors in synthetic biology. Therefore, understanding a promoter's responsiveness to environmental changes, such as temperature fluctuations, can help in designing genetic circuits with more predictable and robust functions. Especially given that future synthetic biology applications, will require higher accuracy and robustness of these components (Brödel *et al.* 2016; Bervoets *et al.* 2018), regardless of environmental factors, copy number (Segall-Shapiro *et al.* 2018), etc.

In this thesis, we investigated the temperature dependence of the transcription initiation dynamics of synthetic genes in *E. coli*. We assessed whether this dependency is affected by the promoter structure, the robustness of its repression mechanism, and also how the intake kinetics of an extracellular inducer is affected by temperature. For this, we made use of three promoters (T7 phage Φ10, Lac/ara-1 and $LacO_3O_1$) and measured their transcription activity, at various temperatures, using single RNA detection techniques and image analysis tools.

3 Experimental and Theoretical Methods for Studying the Dynamics of Transcription Initiation

This chapter provides an overview of the experimental and theoretical methods usually employed for studying the *in vivo* dynamics of transcription initiation, with emphasis on those used in this thesis. These methods include single molecule techniques for RNA detection, live cell imaging using fluorescent proteins, and methods for validating these measurements. Finally, this chapter presents the modelling strategies employed to study the dynamics of transcription initiation.

3.1 Fluorescent Proteins

Due to their non-invasive, high specificity and sensitivity for live cell imaging, the use of fluorescent proteins has become critical for conducting *in vivo* cell biology measurements. The possibility of expressing these proteins fused with a desired target protein, while inside a cell, have made them very convenient for numerous studies, which explains why they have become one of the most used tools in Biology (Tsien 1998).

Although they present several advantages, there are some features that one should consider when choosing a fluorescent protein to use in an imaging experiment (Shaner *et al.* 2005). Namely, the fluorescent protein should present a brightness significantly above the background, photostability, and exhibit a minimal crosstalk in its excitation and emission channels, so that it can be accurately detected for the duration of the experiment. Also, when fused with a protein of interest, the fluorescent protein should not affect the proper functioning of the native protein in the chosen system. In addition, the fluorescent protein should be insensitive to the changes in environmental conditions so as to not compromise the interpretation of the results. In this regard, some of the early and wild type fluorescent proteins are temperature and pH sensitive (Tsien 1998; Shaner *et al.* 2005).

E.g., the wild type GFP folds better at room temperature or lower, than at high temperatures (Tsien 1998), and the yellow fluorescent protein (YFP) is sensitive to pH and Chloride (Griesbeck *et al.* 2001). To overcome these limitations, several mutagenesis studies have been conducted in order to improve their folding and maturation properties at high temperatures (Cormack *et al.* 1996; Nagai *et al.* 2002; Pédelacq *et al.* 2006). These studies have led to a significant rise in the variety of fluorescent proteins available that fold and mature more efficiently, thus increasing their applicability in different environment, organisms and, organelles. Conversely, the fact that some fluorescent proteins are sensitive to environmental factors can be of value to certain studies, in that the changes can be used as a means to sense environmental changes (Kneen *et al.* 1998; Matsuyama *et al.* 2000; Shen *et al.* 2014).

One of the most common illumination scheme used in fluorescence microscopy is a wide-field epi-illumination, where the entire depth of the sample is excited, which makes the out-of-focus fluorescent molecules to also contribute to the background fluorescence. As a consequence, the images have low resolution and high signal/noise ratio, which can compromise the correct interpretation of the data. Several different methods have been developed to avoid excess out-of-focus illumination, and to also increase the image resolution, such as the Confocal microscopy (Pawley 2006), Total Internal Reflection (TIRF) microscopy (Axelrod 1981), and Highly Inclined and Laminated Optical (HILO) sheet microscopy (Tokunaga *et al.* 2008). Confocal Microscopy is based on the reduction of the focal volume by using a pinhole which guarantees that the detected light comes only from the focal point of the sample, thus resulting in a reduction of the out of focus information (Pawley 2006). TIRF microscopy offers a better image sectioning than confocal microscopy, because here only a thin section of the sample's surface is illuminated (Axelrod 1981). The light in TIRF creates a thin lamina of evanescent wave that penetrates the interface between the coverglass and the sample. As such, TIRF only illuminates 50-200nm deep into the sample, thus allowing to probe molecules that are in close proximity with the coverglass surface, such as molecules that are close the cell membrane. In order to illuminate regions deeper than the TIRF imaging range, without a significant reduction in the signal-to-noise ratio, the HILO microcopy was developed (Tokunaga *et al.* 2008). Here, the light is refracted into the sample at a high inclination angle, thus illuminating an angled layer within the sample. This results in a lower out-of-focus fluorescence (Tokunaga *et al.* 2008).

The methodologies described in the previous section show the different imaging techniques that allows the *in vivo* visualization of fluorescent molecules. However, some studies also require the visualization of cells where these fluorescent molecules are inserted in. Phase Contrast microscopy is one of such techniques. It allows the visualization of transparent living organisms, by employing an optical mechanism that converts differences in refractive index and cell density, into detectable

variations in light that can then be visualized as differences in the contrast of the image (Zernike 1942). This technique allows the *in vivo* visualization of cells, such as *E. coli*.

3.2 Single-Molecules Approaches for Studying Transcription

Most of our current knowledge about transcription comes from biochemical and biophysical studies, usually conducted *in vitro*. Although these studies have contributed significantly to our understanding of the transcription process, the *in vivo* characterization of this process is of particular importance, given that a live cell is a dynamic system with a large number of molecules and reactions interconnected into a complex network, and many of its systems have proven to behave differently in an *in vitro* context.

The development of *in vivo* and real time observations techniques has allowed the dissection of the transcription process in live cells, providing detailed spatiotemporal information, which was not possible previously. In particular, the recent advancements in microscopy techniques, fluorescent proteins and single-molecule measurements, have allowed the tagging of individual molecules in a single cell, such as the *in vivo* tagging of RNA molecules using fluorescent proteins (Golding and Cox 2004). Currently, there are different ways of achieving this, but the most popular techniques usually involve the use of complementary oligonucleotides or RNA binding probes labelled with fluorophores, such as RNA binding proteins, which are able to bind to specific RNA motifs (Pitchiaya *et al.* 2014). This popularity is mostly due to their ability in tagging and detecting endogenous RNAs, as well as exogenous constructs (Raj and van Oudenaarden 2009).

The fluorescence *in situ* hybridization (FISH) method, was one of the first methods to achieve single RNA sensitivity (Raj and van Oudenaarden 2009). This method is based on the use of fluorescently labelled oligonucleotides probes that hybridizes, with high sequence complementarity, to a unique region of its target RNA (Raj and van Oudenaarden 2009). FISH has the advantage of allowing the simultaneous detection of different RNAs, and it can also be used when conducting studies on cell-to-cell variability, or on the spatial localization of RNA (Raj and van Oudenaarden 2009; Montero Llopis *et al.* 2010).

The FISH method can provide insights about the several steps comprising the transcription process, but it requires the fixation of cells and the permeabilization of the cell membrane (Raj and van Oudenaarden 2009), thus making it not suitable for conducting *in vivo* studies of the transcription dynamics in single cells, in real time.

3.2.1 The MS2-GFP Tagging Method

The fusion of RNA binding proteins with fluorescent proteins, has allowed the tracking and detection of single RNA molecules in live cell, as soon as they are produced (Bertrand *et al.* 1998; Golding and Cox 2004; Golding *et al.* 2005). The method is based on fusing the gene of interest with a tandem array of repeats, each coding for an RNA binding site region for the tagging protein, and then fusing that protein with a fluorescent protein so that when they bind to the RNA motif, it becomes visible due to the accumulation of multiple fluorescent molecules in the same spot. For this, both genetic constructs need to be inserted into a plasmid or integrated in the genome of a cell.

The most common RNA binding protein used for *in vivo* RNA detection is the MS2 phage coat protein, which interacts with a stem-loop structure in the viral RNA to encapsulate the viral gene (Peabody 1993; Valegard *et al.* 1997; Bertrand *et al.* 1998). This system consists of two constructs (Figure 3.1). The first construct carries the gene of interest and the gene coding for the RNA containing the multiple binding sites (BS) for the MS2 protein, and is usually placed in a single copy plasmid (Golding and Cox 2004; Golding *et al.* 2005). In order to cross validate the transcription activity, an additional gene, usually coding for a fluorescent protein such as the mRFP1 or mCherry can be inserted before the multiple binding sites, and after the promoter of interest (Golding *et al.* 2005; Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016) (Figure 3.1A).

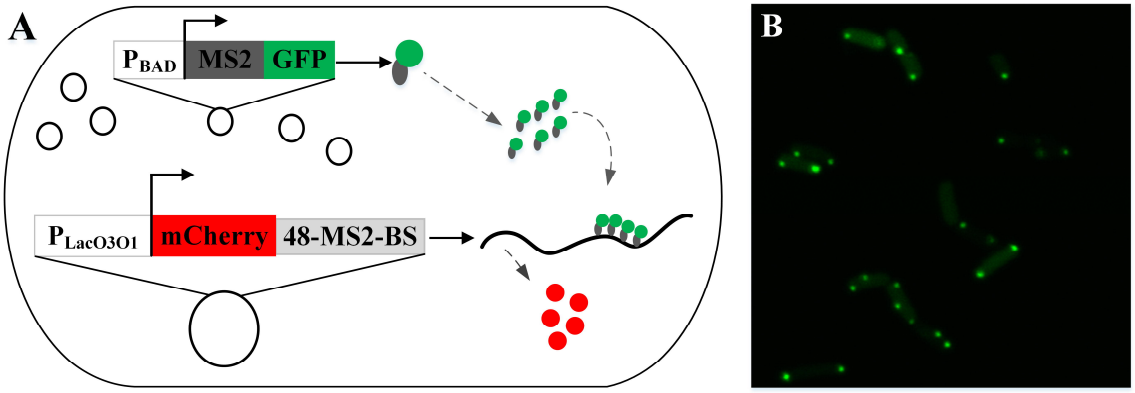


Figure 3.1: *In vivo* detection of RNA in *E. coli* cells, using the MS2-GFP tagging method. (A) Illustration of the MS2-GFP tagging method. The bottom construct represents a single copy plasmid carrying the 48 binding sites (target RNA) for the MS2-GFP proteins (reporter proteins). Here, the target RNA is under the control of the lac_{O3O1} promoter. The MS2-GFP proteins (gray and green balls) are produced from a medium-copy plasmid and are under the control of the P_{BAD} promoter. (B) An example of fluorescent cells using confocal microscopy. Here, these cells are expressing both the target RNA and the reporter MS2-GFP proteins. RNA molecules are visible as fluorescent spots, and the background of the cells is due to the unbound MS2-GFP diffusing freely inside the cell.

To increase their stability, random sequences need to be introduced between the multiple binding sites. The second construct, is usually placed in a multi-copy plasmid coding for the MS2-GFP proteins and is under the control of an inducible promoter (Golding *et al.* 2005; Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016) (Figure 3.1A). To determine the transcription dynamics of the promoter of interest, the reporter protein MS2-GFP should be expressed before activating of the target gene. The intracellular level of MS2-GFP should be high enough, so that it can rapidly bind to all binding sites of the target RNA as soon as they are transcribed. The binding of multiple MS2-GFP to the same target RNA makes it much brighter than the fluorescence of the unbound MS2-GFP that are freely diffusing inside the cell (Golding *et al.* 2005; Xie *et al.* 2008). Under confocal microscopy, the MS2-GFP that are bound to the target RNA appear as bright spots, and can be detected with single-molecule sensitivity (Golding and Cox 2004; Golding *et al.* 2005) (Figure 3.1B).

Since the target RNA is wrapped by the MS2-GFP proteins, which protects it from the natural degradation (Talbot *et al.* 1990; Fusco *et al.* 2003; Golding and Cox 2004; Golding *et al.* 2005), these "RNA spots" do not loose intensity overtime, thus allowing to study the dynamics of transcription initiation without the influence of degradation (Kandavalli *et al.* 2016).

Aside from the MS2 proteins, other proteins have been also used for RNA detection. One is the PP7, from the PP7 bacteriophage (Chao *et al.* 2008; Larson *et al.* 2011). Another is the Lambda $(\lambda)_N$ peptide from λ phage (Daigle and Ellenberg 2007; Lange *et al.* 2008). These proteins have different binding sites and thus their use could allow the simultaneous imaging of up to three different RNA targets in the same cell (Lange *et al.* 2008; Hocine *et al.* 2013).

In this thesis, we used the MS2-GFP tagging method to measure RNA production from different promoters in live *E. coli* cells. In all publications, the target gene was inserted into a single copy F-plasmid. In **Publication I** and **Publication IV**, the target promoter was followed by a tandem array of 96 MS2 binding sites, while in **Publication II**, **III** we made use of only 48 BS.

3.2.2 Constructing a Single-Copy Plasmid Carrying the Binding Sites for the MS2-GFP Reporter Proteins

As mentioned in the beginning of the previous section, the use of fluorescently tagged RNA binding proteins has enabled the *in vivo* detection of RNA molecules. This is achieved by fusing fluorescent proteins with these RNA binding proteins, and currently several techniques have facilitated these constructions. One of such techniques is the Gibson Assembly® Method (Gibson *et al.* 2009; Gibson 2011). This is a cloning method that allows the assembling of multiple overlapping DNA molecules, in a single isothermal reaction. This method is based on the activity of three enzymes:

a 5’ exonuclease that chews the 5’ ends of double-stranded DNA generating single-stranded complementary DNA overhangs, a DNA polymerase which fills in the gaps of the annealed sequence, and a DNA ligase that will seal the resulting nicks (Figure 3.2).

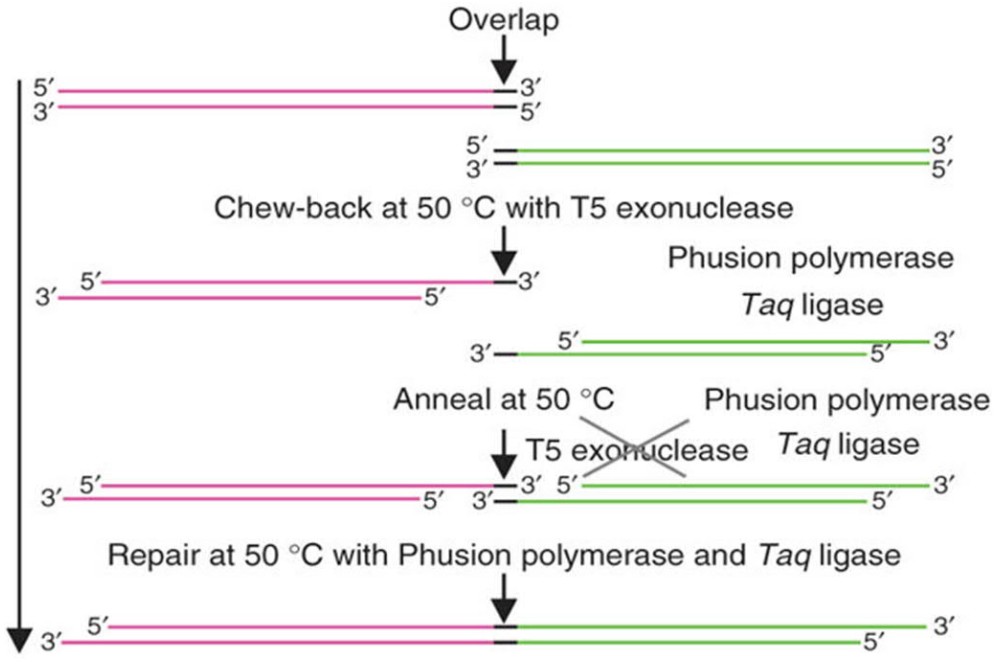


Figure 3.2: An illustration representing the one-step isothermal reaction of the Gibson Assembly® Method. Reprinted by permission from Macmillan Publishers Ltd: [Nature Methods] (Gibson *et al.* 2009), copyright (2009).

For this to occur, the DNA fragments must be designed such that they contain at least 40 bp overlap with the adjacent DNA fragment. This can be easily achieved, via PCR by using primers where these overlapping sequences are introduced. Given that these fragments are usually assembled with a vector to form a circular product, these vectors must also carry terminal ends overlapping with the DNA fragments to which they will be joined with. In a single reaction, the 5’ exonuclease creates single stranded 3’ overhangs, by “chewing back” the DNA at the 5’ end. This facilitates the annealing process between the fragment and the vector, which are complementary at one end (overlapping region). The DNA polymerase then fills the gaps within the annealed fragments and the DNA ligase seals the nicks in the assembled DNA.

In **Publication II**, we evaluated how the transcription initiation dynamics of a viral promoter is affected by temperature changes. To conduct this study, we used the Gibson Assembly® Method to construct a single copy plasmid carrying the T7 phage Φ10 promoter, controlling the expression of the target gene, the 48BS for the MS2-GFP reporter protein (Figure 3.3). Briefly, the coding sequence for the T7 Φ10 promoter and terminator were amplified by PCR, from the pRSET-EmGFP plasmid (ThermoFisher Scientific).

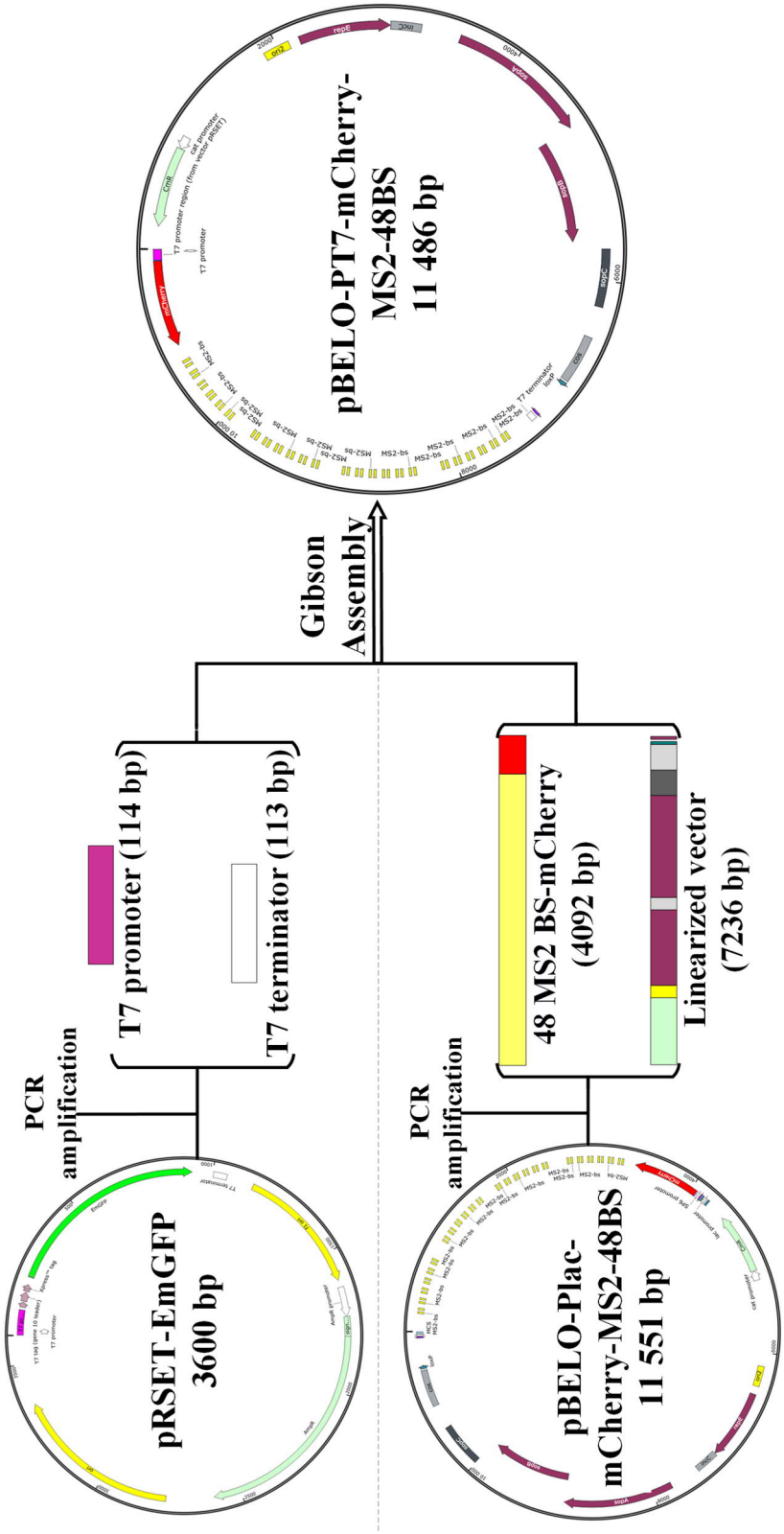


Figure 3.3: An illustration showing the steps involved in the construction of the single copy plasmid carrying the T7 Φ10 promoter and the target gene, using the Gibson Assembly® Method. (The plasmids presented in this figure were generated using SnapGene® 2.3.5).

Given that this is a viral promoter, requiring a different terminator than the one present in the single copy plasmid (pBELO-BAC11), the vector was amplified in a way to originate two fragments (referred to as 48 MS2-mCherry and Linearized vector), thus facilitating the simultaneous ligation of the terminator and the promoter, to the vector's fragments. All fragments were then ligated, in a single reaction, following the guidelines described in (Gibson *et al.* 2009).

3.2.1 Live Cell Imaging and Time-Lapse Microscopy

The MS2-GFP RNA tagging and detection system allows the observation of *in vivo* transcription dynamics, at the single cell and single RNA levels in real time, using time-lapse microscopy (Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016; Mäkelä *et al.* 2017). In these types of experiments, cells are usually placed on an agarose gel pad, which is located between the microscope slide and a coverslip. This agarose gel contains the nutrients necessary for the growth of the cells, as well as inducers of gene expression for both the target and reporter genes (Golding *et al.* 2005; Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016). In addition, to ensure accurate temperature conditions throughout the duration of the experiment, a temperature regulated chamber can also be used (Oliveira *et al.* 2016b). For studies requiring the accurate measurement of transcription events, from the moment of induction, a peristaltic pump can be used, to provide the cells with fresh media supplemented with the appropriate inducers. This also allows steady state growth for the duration of the experiment, under the microscope (Choi *et al.* 2008; Mäkelä *et al.* 2017).

Figure 3.4 shows a schematic illustration of the imaging and the temperature control chamber used for imaging the cells while under microscope. This approach was implemented for the studies described in **Publications I, II, III,** and **IV** where cells were supplied with fresh media containing the appropriate inducers for the target and reporter genes.

In **Publications I** and **IV**, to study the intake kinetics of an inducer from the environment, cells (under the microscope) were supplied with fresh media containing the inducers for the target and reporter genes. Here, the target gene was activated under the microscope, which allows to observe and determine the time taken for the first RNA to be produced. In **Publication II**, the dynamics of the transcription initiation from the T7 Φ10 promoter was studied by providing the cells with the media containing the inducer for the reporter and target system, thus activating both genes under the microscope. In **Publication III**, to measure the transcription rate of the of the $LacO_3O_1$ promoter, under full induction, cells were provided with fresh media containing the appropriate concentrations of the reporter and target inducers. It is worth of mentioning that in all experiments a temperature chamber was used to ensure the correct temperature throughout the duration of the experiment.

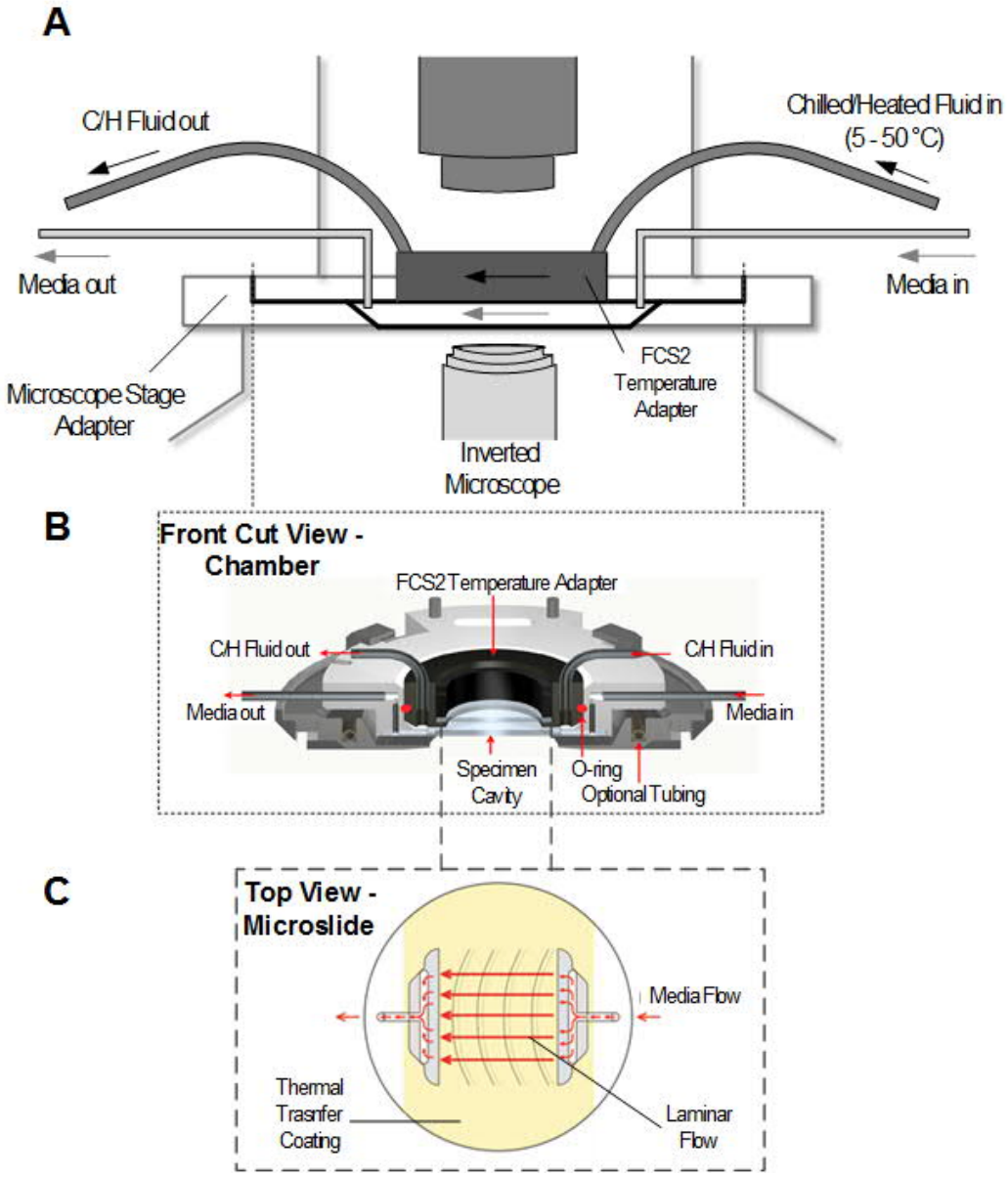


Figure 3.4: (A) Schematic illustration of the CFCS2 microfluidics and the temperature control system for cell cultures, while under microscope observation. The CFCS2 chamber is mounted on the stage of an inverted microscope, and it comprises two independent fluidic systems. One is a thermo-chiller device (not shown), connected to two inlets and outlets of the CFCS2 chamber, which controls the temperature of the system (i.e. the temperature of the metal chamber and the optical cavity, where cells are placed) through the flow of heat/chilled fluidics, whose temperature can range from 5 ºC to 50 ºC ± 0.2 ºC. The second device, a micro-perfusion device (not shown), connected to one inlet and one outlet of the CFCS2, constantly provides the cells with fresh media and chemicals required for cell growth. (B) An illustrative front cut view of the optical cavity of the cooled FCS2 adapter (CFCS2). The CFCS2 is a modified version of the FCS2 system, in that it has an additional, independent tubing system to facilitate the circulation of a heat/chilled fluid, that increases/reduces the temperature of the metal base and of the optical cavity of the chamber. (C) Schematic top view of the

micro-aqueduct slide, which is placed inside the optical cavity. The slide allows laminar flow of fluids, when a uniform and rapid exchange of media is required across the cell population. Images shown in (B) and (C) are adapted from Bioptechs Inc. (http://www.bioptechs.com). Figure from **Publication IV**.

3.3 Methods for Validating Live Cell Imaging Measurements

The MS2-GFP tagging method allows the *in vivo* measurement of transcription events, with single-molecule sensitivity. Given the level of information one can extract using this methodology, currently there is no other methodology capable of independently validate these measurements with the same sensitivity. However, some methods can be employed, which allow a partial validation of these measurements.

Quantitative polymerase chain reaction (qPCR) allows to quantify the amount of RNA produced, at a population level (Schmittgen and Livak 2008). Unlike the traditional PCR method, where the detection of the amplified sequence is performed after the PCR reaction has been completed, in qPCR, this detection is done in real time (Schmittgen and Livak 2008). Here, a modified version of the standard PCR reaction protocol is employed, where sequence-specific primers for the gene of interest are used, and the amplified product is detected in real-time by using fluorescent probes or fluorescent DNA binding dyes. This information is then used to generate the cycle values (C_T), by determining the number of PCR cycles where the fluorescence level crosses a certain threshold value, which is usually the background fluorescence level.

This C_T value is inversely proportional to the amount of amplification product in the reaction, that is, if a particular gene is abundant in the sample, the amplification of this product is observed in the early cycles, while if the amount is very low, the amplification is observed in later cycles (Schmittgen and Livak 2008). From these C_T values it is then possible to determine the fold change in mRNA level using, e.g., the Livak's $2^{-\Delta\Delta CT}$ method (Livak and Schmittgen 2001). In this method, genes that are not affected by the experimental conditions are usually used as a reference gene, which is an internal control used for the relative quantification of the target mRNA levels and/or production rates (Livak and Schmittgen 2001).

In this thesis, qPCR was used as a method to validate the RNA levels obtained from microscopy measurements. In **Publications II** (Figure 3.5), **Publication III** and **Publication IV** this method was used to validate the functionality of the induction mechanisms of the promoters used. Additionally, in **Publication II** and **Publication IV** the qPCR results were also used, in combination with microscope measurements, to produce τ-plots (section 4.3, Chapter 4).

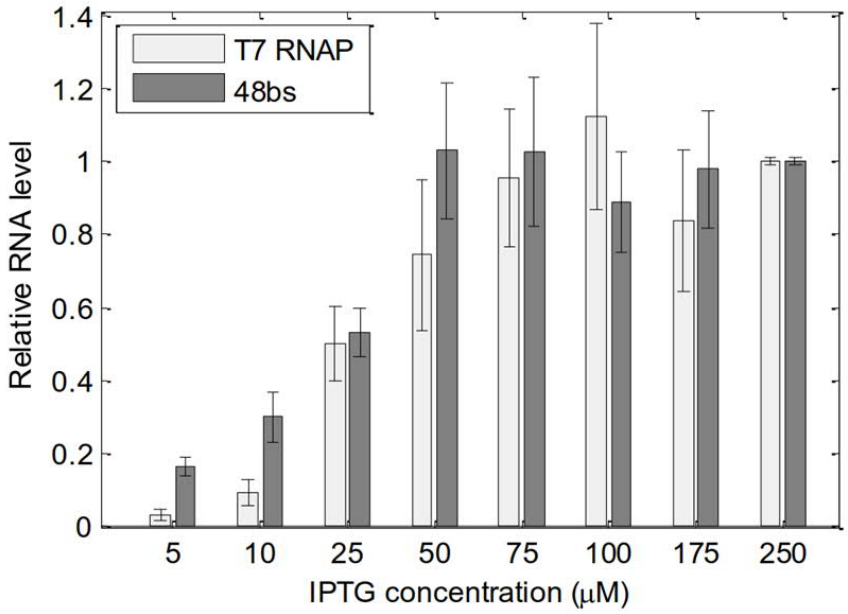


Figure 3.5: An example of qPCR measurements to quantify the relative RNA levels of a gene of interest. Here, the relative levels of *T7 RNAp* (light greys) and the target gene *mCherry* (dark grey) were measured at 37^0C, for varying IPTG concentrations. The standard errors are also shown for each conditions and were obtained from three technical replicates. The *16SrRNA* gene was used as a reference. Figure from **Publication II**.

3.4 Modelling Transcription Initiation in *E. coli*

When conducting studies in *E. coli*, due to the limitations of the measurements techniques, sometimes it is not possible to directly observe in great detail all the steps of a particular process, in its native form. Thus, to overcome these limitations we make use of models, which can be defined as simplified representations of a system or process. These models require tuning, which means finding the model parameter values that best reproduce the observed data. In the end, they require validation. This can be done by making a prediction with the model and verifying that prediction empirically. Models are particularly useful to generate predictions on how a system will behave under different conditions, when it is not possible to observe it experimentally.

Modelling strategies have been widely used to investigate the gene expression process in *E. coli*. Being a chemical process, it can be described as a set of chemical reactions. In general, chemical reactions can be represented as shown below:

$$A + B \xrightarrow{k} C \tag{3.1}$$

Here, one molecule of the species A and B react to form a molecule of the species C, at a rate of *k*.

When modeling transcription, the most important steps, i.e. those affecting the RNA production kinetics, should be included (Ribeiro *et al.* 2006). In its simplest form, the transcription process can be modelled as a single-step reaction, as follows:

$$RNAp + Pro \xrightarrow{k} Pro + RNAp + RNA \tag{3.2}$$

In this reaction, RNAp is the RNAp holoenzyme, Pro is the promoter, and RNA is the RNA produced from this reaction. It is important to mention that in this model, there is no regulatory molecule controlling the activity of the promoter, thus the gene is transcribed in a constitutive manner.

This model is an oversimplification of the transcription process from the point of view of the current state of the art of measurement techniques. For instance, it does not consider the existence of multiple rate-limiting or reversible steps in the transcription process. Thus, in order to account for such factors, a more detailed model is required.

In *E. coli*, the transcription initiation process has been described as a complex multi-step process where the RNAp has to find the promoter, open the DNA and initiate the RNA synthesis (McClure 1985; Saecker *et al.* 2011). This process can be described as follows:

$$RNAp + Pro \xleftrightarrow{K_B} RP_C \xrightarrow{k_f} RP_O \rightarrow\rightarrow\rightarrow RNA \tag{3.3}$$

The model (3.3) was first suggested by Walter, Zillig and colleagues (Walter *et al.* 1967; Chamberlin 1974; McClure 1985). It includes the initial binding of the RNAp holoenzyme to a promoter with a binding constant, K_B, to form the closed complex, RP_C, which then isomerizes with a rate constant of k_f to form the active open complex, RP_O (McClure 1985). From that point onwards, the process becomes nearly irreversible and much faster, being thus represented by the symbol '→→→'.

Our current knowledge of the steps during transcription initiation (Buc and McClure 1985; McClure 1985; Saecker *et al.* 2011), and how these can affect the rate of RNA production (Kandavalli *et al.* 2016) is extensive. These effects can be accounted in the model (3.3) by changing the values of the rate constants.

It has been reported that in prokaryotes, the transcription of highly expressed genes can, under certain conditions, occur in bursts, due to the buildup of positive supercoiling with transcription events (Chong *et al.* 2014). While the chromosome presents topologically constrained segments

that allow this supercoiling buildup to occur (Hardy and Cozzarelli 2005; Higgins 2016), plasmids lack such constrains, only presenting transient ones (for instance, due to transient protein binding) (Leng *et al.* 2011). The exception to these are plasmids encoding membrane associated proteins (Pruss and Drlica 1986; Lynch and Wang 1993), carrying tandem copies of multiple DNA-binding sites (Fulcrand *et al.* 2016), or when expressed in strains lacking DNA topoisomerase I (Samul and Leng 2007). In these plasmids, these topologically constrained segments are more efficient than the transient ones in leading to the appearance of supercoiling buildup with transcription events (Liu and Wang 1987; Deng *et al.* 2005). When these 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).

In this thesis, all the genes studied were inserted into single copy plasmids lacking these topological constrains. Due to this, the models of transcription initiation presented do not account for transcriptional bursts.

In **Publication II**, we investigated how the transcription kinetics of the T7 phage Φ 10 promoter changes with temperature. For that, we assumed the transcription model proposed in (McClure 1980, 1985; Buc and McClure 1985). The model is as follows:

$$R + P \underset{k'_{cc}}{\overset{k_{cc}}{\rightleftarrows}} Pr_{cc} \xrightarrow{k_{oc}} Pr_{oc} \xrightarrow{\infty} Pr + R + RNA \tag{3.4}$$

In (3.4), R is an active T7 RNAp, P is the free promoter, Pr_{CC} is a formed closed complex and the Pr_{OC} is a formed open complex. The closed complex formation occurs at a rate of k_{CC}, and once it is formed, the promoter can be unbound by the R at a rate of k'_{CC} or it can undergo open complex formation, at a rate k_{OC}. It is worth to mention that this model does not include an ON-OFF mechanism, given that Φ 10 is a constitutive promoter.

For promoters regulated by repressor molecules, an additional layer of complexity would need to be added, such as the binding of the repressor to the promoter region. In **Publication III**, we studied the temperature dependence of the repression mechanism of the $LacO_3O_1$ promoter. To assess this, we measured the leakiness of this promoter, meaning the RNA production events in the absence of any inducer molecules, at different temperatures. The activity of this promoter is controlled by its repressor, the LacI (Oehler *et al.* 1990). Thus, we assumed the following model of transcription and repression mechanism:

$$\mathrm{Pro}_{OFF}.\mathrm{Rep} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} \mathrm{Pro}_{ON} + \mathrm{Rep} \tag{3.5}$$

$$\mathrm{Pro}_{ON} + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro}_{ON} + \mathrm{RNAp} + \mathrm{RNA} \tag{3.6}$$

Reaction (3.5) models the repression mechanism where Pro is the promoter, Rep is the repressor (here, the LacI tetramers), k_{on} and k_{off} are the rates of the repressor binding and unbinding to the promoter, respectively. Contrary to the models used in the previous publication, reaction (3.6) models transcription as a single step process, where k_t is the rate by which the RNAp finds a promoter, and after this, it produces an RNA. We used a simplified model which does not include the rate-limiting steps of transcription initiation because we found that it did not affect our conclusions, and thus opted for the simplest model.

From these models, the average RNA production rate (λ_{RNA}), can be defined as:

$$\lambda_{RNA} = \mathrm{P}_{ON} \times (k_t \times RNAp) \tag{3.7}$$

Where P_{ON} is the probability that the promoter is "ON", and from (3.5) and (3.6) it can be:

$$\mathrm{P}_{ON} = \frac{k_{on}}{\mathrm{Rep} \times k_{off} + k_{on}} \tag{3.8}$$

From these quantities and the variables that they depend upon, the RNA production rate (λ_{RNA}) equals:

$$\lambda_{RNA} = \left(\frac{k_{on}}{Rep \times k_{off} + k_{on}} \right) \times RNAp \times k_t \tag{3.9}$$

Equation (3.9) shows which factors determine the mean rate of RNA production. These are the RNAp concentration, the transcription rate (k_t) of the free, active promoter, and the fraction of time that the promoter is free from repressors, which depends on the repressors numbers (*Rep*) and their rates of binding (k_{on}) and unbinding (k_{off}) to the promoter.

Given the limitations to experimentally dissect the parameters associated with the repression mechanism, we defined β as the 'repression strength', being equal to the inverse of the time that the promoter is free from repressors:

$$\beta = \frac{Rep \times k_{off} + k_{on}}{k_{on}} \tag{3.10}$$

From (3.10), when the system is fully induced, assuming that all repressor molecules are virtually inactive due to the action of inducer molecules (IPTG), β is equal to 1. Thus, it is possible to define the average rate of RNA production, under full induction (λ_{RNA}^{Act}) and under full repression ($\lambda_{RNA}^{\mathrm{Rep}}$) to be, respectively:

$$\lambda_{RNA}^{Act} \cong k_t \times RNAp \tag{3.11}$$

$$\lambda_{RNA}^{\mathrm{Rep}} \cong \beta^{-1} \times k_t \times RNAp \tag{3.12}$$

Both equations inform on the parameters determining the mean rate of RNA production of an active promoter (3.11) and the leakiness production of RNA (3.12). Given that it is possible to obtain empirically the values for λ_{RNA}^{Act}, and the relative RNAp levels in the cells, it is possible to determine the value for k_t. After determining k_t for active promoters, we can then calculate which factors in (3.12) were the most affected by temperature, thus contributing the most for the leakiness of this promoter.

For this, we work with the following formula, obtained from (3.12):

$$\frac{\beta}{\beta_{control}} = \frac{RNAp}{RNAp_{control}} \times \frac{k_t}{k_{t(control)}} \times \left(\frac{\lambda_{RNA}^{\mathrm{Rep}}}{\lambda_{RNA}^{\mathrm{Re}\,p}(control)} \right)^{-1} \tag{3.13}$$

In short, from (3.12), by obtaining empirically the values of k_t, $\lambda_{RNA}^{\mathrm{Rep}}$, and RNAp in a given condition relative to a control condition, one can estimate the degree to which the repression strength of this mechanism was affected by changing conditions.

Similarly to previous publication, the promoter used in **Publication I** is also regulated by repressor molecules, and as such, its transcription can be modeled as follows:

$$R + \mathrm{Pr} \xleftrightarrow{K_R} \mathrm{Pr}R \tag{3.14}$$

$$\mathrm{Pr} \xrightarrow{k_c} \mathrm{Pr_c} \tag{3.15}$$

$$\mathrm{Pr_c} \xrightarrow{k_o} \mathrm{Pr_o} \xrightarrow{\infty} \mathrm{Pr} + \mathrm{M} \tag{3.16}$$

Reaction (3.14) models the bind and unbinding of the repressor (R) to the promoter (Pr), with a rate constant K_R, to form the promoter repressor complex (PrR). When free from repressors, an RNAp holoenzyme can bind to this promoter, leading to the formation of the closed complex ($\mathrm{Pr_c}$) (represented in reaction 3.15), with a rate k_c. This is followed by the formation of the open complex with a rate constant k_o (3.16). The formation of the open complex is followed by promoter escape, transcription elongation and the release of a RNA molecule (M).

3.4.1 Model of Inducer Intake and Active Transcription

In **Publication I** and **IV**, we characterized the intake kinetics of IPTG molecules at, respectively, optimal and sub-optimal temperatures in the absence of a transporter protein. In both studies, we used the Lac/ara-1 promoter, assuming the model depicted in reactions (3.14-3.16).

Since the *E. coli* strain used in these studies (DH5αPro) does not produce LacY (Lutz and Bujard 1997), the intake process of inducer molecules is expected to be diffusive-like. Also, as mentioned in section 2.3.1 (Chapter 2), *E. coli*'s membrane consists of two layers, the outer and the inner

membrane, separated by the periplasmic space (Zimmermann and Rosselet 1977; Alberts *et al.* 2002). Given this, the activation process is expected to have at least two-rate limiting steps: the entrance of inducer molecules from the media to the periplasmic space, and the entrance of these molecules to the cell's cytoplasm. Thus, the activation of this gene can be modelled by a two-step process, as follows:

$$\mathrm{I}_{env} \xrightarrow{Int_1} \mathrm{I}_{peri} \tag{3.17}$$

$$\mathrm{I}_{peri} \xrightarrow{Int_2} \mathrm{I} \tag{3.18}$$

Reaction (3.17) represents the entrance of an inducer molecule (I) into the periplasm of the cell, and reaction (3.18) the entrance of an inducer molecule from the periplasm to the cell's cytoplasm, where it interacts with repressor molecules leading to transcription activation of the target gene. Thus, it can be modelled as follows:

$$\mathrm{Rep} + \mathrm{I} \underset{k_{inact}}{\overset{k_{act}}{\rightleftarrows}} \mathrm{Rep.I} \tag{3.19}$$

$$\mathrm{P}_{OFF} + \mathrm{I} \xrightarrow{k_{act}} \mathrm{P}_{ON} + \mathrm{Rep.I} \tag{3.20}$$

$$\mathrm{P}_{OFF} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} \mathrm{P}_{ON} + \mathrm{Rep} \tag{3.21}$$

In reaction (3.19), an inducer (I) binds to the repressor (Rep), forming the complex (Rep.I), which cannot repress a promoter. Here, the repressors are LacI tetramers, and IPTG is the inducer that binds to the repressors reducing their binding affinity to the promoter (Lewis 2005). In reaction (3.20), the inducer binds to the repressor, which is bound to the promoter, thus freeing the promoter from the repressors. Reaction (3.21) represents the repression of the promoter, by free repressors, and it also represents the possibility of a repressor unbinding the promoter, without a direct interaction with inducer molecules.

Finally, active transcription by a free promoter can be modelled as a multi-step process (Ribeiro *et al.* 2006; McClure 1985):

$$\mathrm{R} + \mathrm{P}_{ON} \xrightarrow{k_{cc}} \mathrm{RP}_{\mathrm{CC}} \xrightarrow{k_{oc}} \mathrm{RP}_{\mathrm{OC}} \xrightarrow{\infty} \mathrm{P}_{ON} + \mathrm{R} + \mathrm{RNA} \tag{3.22}$$

In reaction (3.22), R is the RNA polymerase, and once bound to the promoter, it forms the closed complex ($\mathrm{RP_{CC}}$), which is followed by the formation of the open complex ($\mathrm{RP_{OC}}$), elongation and finally production of RNA and release of RNAp. Note that, except for the open complex formation, which once initiated is nearly irreversible (McCLure 1985), the steps represented in reaction (3.22) are considered to be reversible.

4 Computational Tools for Image and Data Analysis

This chapter presents the computational tools implemented in this thesis for image analysis. These tools were used for cell segmentation and RNA spot detection and quantification from microscope images. The measurement of time intervals is discussed in the final section of this chapter.

4.1 Image Analysis and RNA Quantification

When conducting transcription dynamics studies using single-molecules measurement techniques, the amount of images acquired is usually high, and in cases where time-lapse measurements are conducted, this number increases significantly. Thus, image analysis and signal processing tools are required so as to accurately and unbiasedly estimate the number of RNA molecules in each cell at any given time.

The first step in image analysis is cell segmentation, where cells are detected and segmented from the background. This process is done from phase contrast images using the software “iCellFusion” (Santinha *et al.* 2015), which in a first step performs automatic segmentation of the cells, but then allows the results to be manually corrected, when necessary. The level of accuracy is thus the highest possible in present days, although laborious.

After the segmentation process is concluded, the phase contrast images have to be aligned with their corresponding fluorescent images (Figure 4.1). This registration process is performed automatically by iCellFusion. Manual correction (by control-point mapping) is also possible.

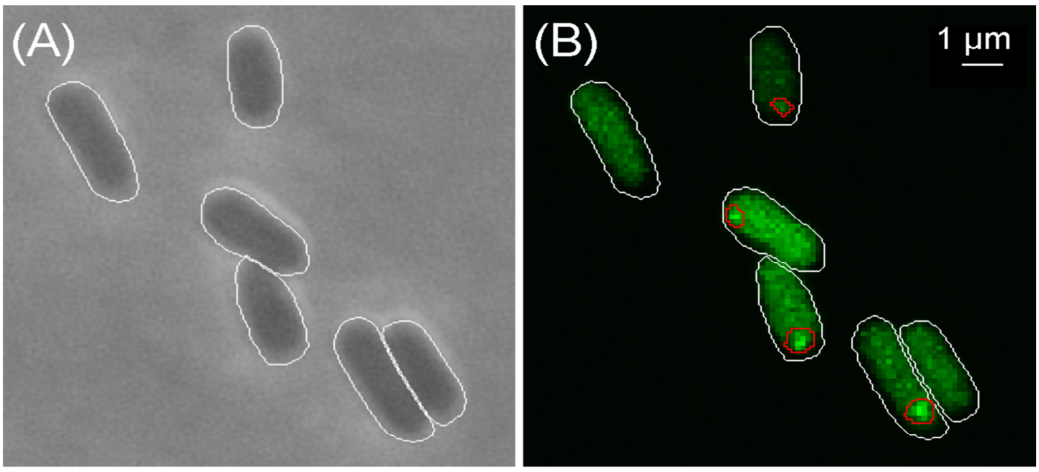


Figure 4.1: Example of phase contrast images, along with the result of cell segmentation (A), and confocal images of cells with the MS2-GFP-RNA spots (B), which were detected and segmented using the spot detection method.

In addition, when analyzing time series data, it is necessary to establish a temporal relationship between the cells of consecutive frames. The software "CellAging" (Häkkinen *et al.* 2013) achieves this by determining the cell lineages from overlapping areas of segments between consecutive images. After, the fluorescent MS2-GFP-RNA spots in each cell are detected using the Kernel Density Estimation (KDE) method for spot detection (Häkkinen *et al.* 2013).

To quantify the RNA numbers in each cell, we used the method described by Golding and colleagues (2005). Briefly, the number of RNA molecules in each cell can be extracted from the distribution of total spots intensity in the cell (an intensity histogram from all spots intensity), by normalizing it with the intensity of a single RNA molecule, which is equivalent to the first peak in the histogram (Golding *et al.* 2005).

In this thesis, the spot and cell segmentation methods were applied in all publications, along with the RNA quantification methods. In all publications the RNA quantification methods were used to extract the RNA numbers from time series data, and in addition, in **Publication III** these numbers were also obtained from population data.

4.2 Measurement of Time Intervals and Estimation of Inducer Intake Times

Time-lapse or time series measurements allow the observation of transcription events in real time (Lloyd-Price *et al.* 2016; Kandavalli *et al.* 2016; Mäkelä *et al.* 2017). The information extracted from these experiments can thus be used, for instance, to build detailed models of RNA production dynamics, at a single cell level.

Time series data carry far more information than just a single time point distribution of RNA numbers that can be obtained from a population of cells. The MS2-GFP tagging of RNA allows for the detection of RNA molecules as soon as they are produced (Golding *et al.* 2005), and given that these RNAs are protected from degradation (Talbot *et al.* 1990; Fusco *et al.* 2003; Golding and Cox 2004; Golding *et al.* 2005), the total intensity of these RNA spots increases overtime, as RNA production events occur (Mäkelä *et al.* 2017). Thus, the appearance of a new RNA molecule in the cell results in positive "jumps" in the total spot fluorescence intensity in the cell. These jumps in intensity can then be used to measure the time intervals between consecutive productions of RNAs, from which one can extract a distribution of time intervals of RNA production in individual cells (Mäkelä *et al.* 2017). Based on this distribution, a model of transcription with a specific number of rate-limiting steps (and their order of magnitude) can be automatically selected using the likelihood ratio test, which evaluates the goodness of fitness between models (Häkkinen and Ribeiro 2015).

In **Publication II**, we obtained the time intervals between RNA production from the T7 phage Φ 10 promoter. Based on the method described above, we then found that a two steps model of transcription initiation was the one that best fitted the data.

The 'jump detection' method can be also used to determine the time taken for the first RNA to appear in each cell under observation, following induction. From the experimental point of view, this requires the use of microfluidics or a peristaltic pump, which allows to induce the cells while under the microscope. The image acquisition of the time series can then be started simultaneously with the induction of the target gene. Then, the time for the first RNA to appear, denoted t_0, is counted from the start of induction (Mäkelä *et al.* 2017). Figure 4.2 shows an example of the time for the first RNA to appear (t_0) in each cell following induction, as well as of time intervals between consecutive RNA productions, known as Δt.

It is worth noting that the time taken for the RNA to appear in the cell also includes the time for the cells to uptake inducer molecules from the media and the time that it takes to produce the first RNA once the promoter has been activated (Mäkelä *et al.* 2017). Thus, this method can be used to estimate the time taken for inducer molecules to enter the cell, as shown in **Publication I**, where this method was used to infer the time taken for IPTG molecules to enter the cells, in a regime of high concentration of extracellular inducer molecules.

Given that it is not possible to directly measure the time taken for inducer molecules to enter the cell's cytoplasm, and activate the promoter of interest, in **Publication IV** a new methodology for inferring this information from time series data is presented.

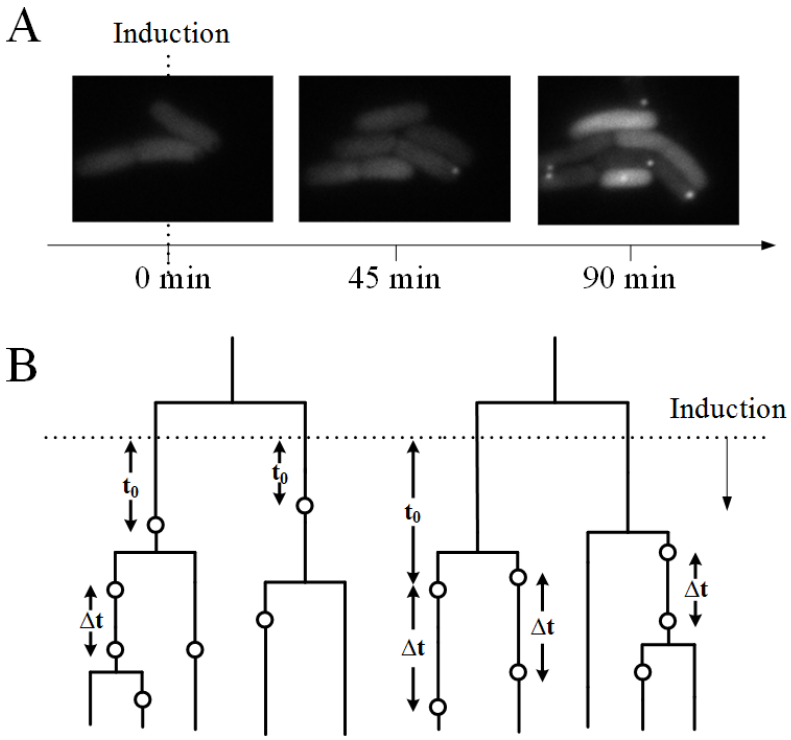


Figure 4.2: Detection of RNA production events from individual cells time series data. (A) Shows the moment when the cells are placed under the microscope (t=0) and continuously supplemented with fresh medium containing the inducers of the reporter and target genes. (B) Illustrates RNA production events (circles) in cell lineages. The waiting time for the first RNA to be produced (t_0) and the intervals between subsequent production events (Δt) are also shown. The dotted lines represent the time where the inducer of the target gene was introduced. Figure from **Publication IV**.

According to the model presented in section 3.4.1 (Chapter 3), following the addition of inducers in the media, the mean time for the first RNA to appear in the cell (t_0) depends on the time for the inducers to enter the cell (reactions (3.17) and (3.18)), (t_{int}), and also on the time for the RNA to be produced by an active promoter, (reactions (3.19)-(3.22)), (Δt). Given that under full induction this time equals the time between consecutive RNA productions in active promoters (Mäkelä *et al.* 2017), we have:

$$t_0 = t_{int} + \Delta t \tag{4.1}$$

Finally, given this, since the inducer intake and the production of the first RNA are independent and consecutive processes, it should be possible to determine the t_{int} by deconvolving Δt from t_0. This is possible if at least 1 of the 2 distributions composing t_0 are known. E.g., knowing t_0 and one of its components, Δt, allows finding the distribution of values of t_{int}, the other component.

These results where then empirically validated by using the Lineweaver-Burk equation (Lineweaver and Burk 1934) to estimate the mean intake times. From equation (4.1) and the model of

gene expression (reactions 3.17-3.22, Chapter 3), as the amount of inducer is increased in the media, in a first stage, the inducers inside the cell will also increase, and both the t_{int} and the Δt will decrease. However, beyond a certain concentration of inducers, additional increases in this concentration will not lead to an increase in the rate of RNA production, given that the regime of full induction has been already reached.

Based on reaction (3.22), this is due to the rate-limiting steps in transcription initiation and the limited number of RNA polymerases inside the cell. On the other hand, the time taken for the cells to intake inducer molecules from the media should decrease with increasing concentration of these molecules, even under full induction. This means that, in theory, for an infinite number of inducer molecules the intake time will become infinitely fast and should equal zero. Thus, following the introduction of inducer molecules in the media (full induction, i.e. IPTG = ∞), the mean time taken for the first RNA (t_0) to be produced will be equal to the duration of interval between consecutives transcription events (Δt).

In **Publication IV**, this method was used to determine the mean intake time for IPTG molecules to enter the cell. For that, from single-cell time-series data, at a given temperature, we determined the Δt and t_0, for two different concentrations of IPTG (both in the regime of full induction). It is worth mentioning that, under this regime, Δt does not differ between the two conditions, thus t_{int} is the only component of t_0 that is affected. From the t_0 measurements at these two induction levels (for a given temperature), the Lineweaver–Burk equation can be used to extrapolate the t_0 value for IPTG = ∞, which allows estimating the mean IPTG intake time, at that particular temperature.

This methodology is similar to the usage of τ plots, where by plotting the results of the measurements of transcription rate against the inverse of RNAp concentrations it is possible to extract the duration of events following the open complex formation (Buc and McClure 1985; McClure 1985).

Table 4.1: Table showing the mean t_{int} ($\mu_{t_{int}}$) obtained from deconvolution and using the Lineweaver-Burk equation. It is worth mentioning that although the mean t_{int} obtained using Lineweaver-Burk equation is larger (by 35-50%) than the deconvolved $\mu_{t_{int}}$, this is expected given that it usually underestimates the peak value of the probability density function (Sheu and Ratcliff 1995).

T(°C)	Deconvolved $\mu_{\widehat{t_{int}}}$(s)	$\mu_{t_{int}}$ (s) using Lineweaver-Burk Eq.
24	1548	2434
37	986	1322
41	1083	1459

4.3 Tau (τ) Plots

As described in the previous chapters, transcription initiation is a multi-step process that starts with the binding of the RNAp to the promoter, which leads to the closed complex formation followed by isomerization to form a transcriptionally active open complex (Chamberlin 1974). The kinetics of these steps have been measured directly, using abortive initiation techniques and *in vitro* transcription initiation assays (Buc and McClure 1985; McClure 1985; Lutz *et al.* 2001). For instance, the rate of the open complex formation has been determined by the mean time taken by the components of transcription, namely the RNAp and the promoter, to reach a steady state rate of production of abortive products. These measurements have shown that there is a sequence-specific lag time before reaching steady state, which was interpreted as the time taken for the RNAp to bind the promoter and form the closed complex. It is the fact that this lag time changes with the concentration of RNAp that distinguishes the closed complex formation from the open complex formation (Buc and McClure 1985).

This dependence allows the construction of a 'τ plot', which depicts the positive linear relationship between the lag times (inverse of the rate constant of closed complex formation) and the inverse of the RNAp concentration (McClure 1980). In a τ plot, the slope of the line between the data points corresponds to the mean time for the completion of the closed complex formation. Meanwhile, the point where the line intercepts with the y axis corresponds to the mean time for the open complex formation, since it corresponds to having an infinite concentration of RNAp's in the system (McClure 1980).

The use of *in vitro* techniques to study the kinetics of the steps in transcription initiation has the advantage of allowing to measure transcription for a wide range of concentrations of RNAp. On the other hand, changing RNAp concentrations in live cells is expected to disturb the cell significantly (Gummesson *et al.* 2009). This makes it difficult to assess *in vivo*, the effect of changing RNAp levels on RNA production rates for a given promoter. However, a recent work showed that the two major impediments could be overcome, and established a method for dissecting the *in vivo* kinetics of the steps involved in transcription initiation (Lloyd-Price *et al.* 2016). Namely, first, it was shown that RNAp concentrations could be changed to a degree, without altering tangibly the cell growth rates. This not only shows that the cells are not being placed under harmful stress but also that differences in cell division rates do not disturb the estimations. Additionally, it was shown that, within this range, the rate of RNA production changed linearly with the inverse of the RNAp concentration. Second, this method assumes that the fraction of RNA polymerases free for transcription is approximately constant within this range of conditions and as such, the intracellular concentration of free RNAps can be assessed from the total RNAp concentration. Note that if this

condition was not valid, a Lineweaver-Burk plot of the inverse of RNAp concentration versus the rate of RNA production would result in a curve. Thus, the occurrence of a line is evidence that (i) the relative free RNAp concentrations can be assessed from the total RNAp concentrations, and that (ii) no factors other than the changes in the free RNAp concentration are affecting transcription of the target promoter (Lloyd-Price *et al.* 2016).

Given this, after defining which media conditions result in specific intracellular levels of RNAp, it is possible to determine, the relative RNA production rate of a promoter, which is inversely proportional to the mean duration of the time intervals between consecutive RNA production. It is then possible to fit the general model of transcription initiation to the empirical data, which accounts for the multi-steps comprising this process, and estimate the *in vivo* duration of the open and closed complex for a particular promoter.

In **Publication II**, we made use of this strategy to assess how the duration of the closed and open complex formation of the T7 phage Φ10 promoter change with temperature. For this, we obtained empirical data on how its transcription activity changes with varying concentration of T7 RNAp. As mentioned above, this change should affect the kinetics of the closed complex, but not of the following steps (Lloyd-Price *et al.* 2016). Here, instead of using media richness to change the T7 RNAp levels, we implemented different concentrations of IPTG, since the gene coding for T7 RNAp was placed under the control of the LacUV5 promoter (Studier and Moffatt 1986). Meanwhile, the relative levels of T7 RNAp and the target gene were determined by qPCR.

From reaction (3.4) (Chapter 3), the mean time interval between consecutive RNA production (Δt) is:

$$\Delta t(RNAp) = \frac{(k'_{CC} + k_{OC})}{RNAp k_{CC} k_{OC}} + \frac{1}{k_{OC}} = \frac{1+K}{RNAp k_{CC}} + \frac{1}{k_{OC}} \tag{4.2}$$

In (4.2), RNAp is the concentration of T7 RNAp in the cell while K is the ratio between k'_{CC} and k_{OC}. From (4.2) Δt is given by:

$$\Delta t(RNAp) = \tau(RNAp) + \tau_{OC} \tag{4.3}$$

Where τ(RNAp) is the mean time for an RNAp to commit to the open complex formation, and τ_{OC} is the mean time for the completion of the open complex formation. From (4.3), the inverse of the interval between consecutive RNA production events changes linearly with the inverse of T7 RNAp level (1/RNAp). Also, from (4.2) and (4.3):

$$\tau_{OC} = \Delta t(RNAp = \infty) \tag{4.4}$$

From qPCR results, we can infer the relative rate of RNA production, given an infinite amount of T7 RNAps in the cell. This rate should correspond to the fraction of time of the transcription initiation process that corresponds to the open complex formation alone (Lloyd-Price *et al.* 2016), as depicted in Equation (4.4). Figure 4.3 shows the resulting τ-plots for each temperature assessed in **Publication II**.

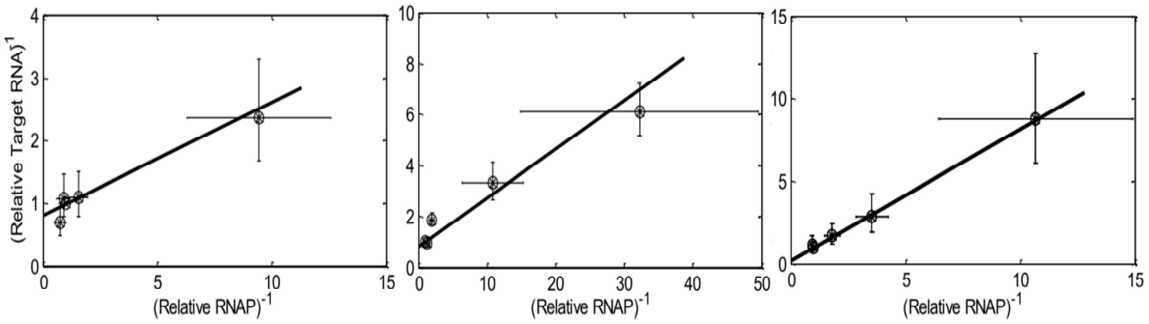


Figure 4.3: Tau plots for the T7 Φ 10 promoter activity at different temperatures: (A) 43^{0}C, (B) 37^{0}C and (C) 20^{0}C. Figure from **Publication IV**.

5 Conclusions and Discussion

Temperature changes affect several molecular processes in *E. coli*, and upon sensing these fluctuations, the cell undergoes changes which allows it to respond accordingly. This response includes a modulation in gene expression, known to occur mostly at the transcription level.

In this thesis, we investigated the temperature dependence of the dynamics of the process of transcription of synthetic genes at the single RNA level, using *in vivo* single RNA-detection techniques and tailored image and signal processing techniques. We first investigated at the single RNA level, the temperature dependence of the transcription initiation dynamics, using *in vivo* single RNA-detection techniques and tailored image and signal processing methods. For this, we first assessed the effects of the intake kinetics of inducer molecules on the expression dynamics of its target genes at optimal (**Publication I**) and suboptimal temperatures (**Publication IV**). Next, we studied how temperature affects the kinetics of the transcription initiation process from a viral promoter, which is structurally different from *E. coli* promoters (**Publication II**). Finally, we evaluated how the leakiness rate of the repressed $LacO_3O_1$ promoter is affected by suboptimal temperatures (**Publication III**.

In **Publication I** we characterized the mechanism of the intake kinetics of IPTG molecules from the media to the cell cytoplasm, in the absence of transporter proteins. To conduct this study, we made use of *in vivo* measurements techniques that allow for the induction of the target gene simultaneously with the image acquisition. Using this technique, we then determined the intervals between consecutive RNA productions and the time for the first RNA to appear in the cell, following induction. Given that the time for the first RNA to appear in the cell also includes the time for the cells to intake inducer molecules from the media (Mäkelä *et al.* 2017), we used a model of inducer intake through a bilayer membrane, coupled with a multi-step model of transcription, to assess how the intake kinetics changes, for different extracellular levels of inducer. We found that

for high extracellular concentrations of inducer, the intake at the outer membrane was diffusive-like, with the intake rate changing linearly with the concentration of inducer. Additionally, the model of a bilayer membrane fitted well the observed data, with the inducer molecules taking, on average, approximately 32 minutes to travel from the periplasm to the cell cytoplasm. And it also differs widely between cells.

These results show that the entrance of inducer molecules from the media to the cell cytoplasm is both time consuming as well as highly noisy, which indicates that, first, it is a major component of the cell's response time, and second, it is a major source of cell-to-cell diversity in RNA and protein numbers for a long period of time following induction. Relevantly, the methodology presented in this publication can be applied to study the intake kinetics of any inducer molecule, which provides clues on which membrane features are involved in this process, and to study how suboptimal conditions, which are known to affect the membrane fluidity, affect the intake kinetics.

In **Publication II**, we studied the transcription initiation kinetics of the T7 Φ10 promoter as a function of temperature. For that, we constructed a single-copy plasmid carrying the Φ10 promoter controlling the expression of the target gene, coding for 48BS for MS2-GFP reporter proteins. Given that Φ10 is a constitutive promoter, we used the strain BL21 (DE3). This strain carries a single copy of the gene co

ding for the T7 RNAp under the control of the LacUV5 promoter (Studier and Moffatt 1986), thus allowing to control, indirectly, the expression of Φ10.

From time-lapse microscopy measurements, we compared the transcription kinetics of Φ10 at different temperatures (20^0C, 37^0C and 43^0C). We found that the mean time intervals between RNA production events in individual cells increases with decreasing temperature, proving the temperature dependence of this process.

Importantly, the kinetics of RNA production at the different temperature was not controlled solely by the T7 RNAp numbers, i.e. Φ10's transcription activity was not maximized at 37^0C, while the relative levels of T7 RNAp were. To investigate this, we evaluated how the kinetics of the underlying steps in transcription initiation change with temperature and T7 RNAp levels. The results indicated that the rate of open complex formation was faster at 20^0C than at higher temperatures. This suggested that at higher temperatures, the open complex might be less stable, thus more likely to be reversible. Additionally, the reversibility of the closed complex increased with increasing temperature. However, at 20^0C, this reversibility appears to be negligible, probably due to a more stable closed complex formation as well as a faster open complex formation.

Overall, these results showed that at lower temperatures, the mean and the noise in RNA production from this promoter are lower, likely due to a stabilization of the closed complex formation in these conditions. Although structurally different from *E. coli* promoters, the initiation kinetics of the Φ10 promoter presents some similarities with those reported for *E. coli* promoters, namely TetA. E.g., the transcription initiation is also sub-Poissonian, with two rate-limiting steps, with its open complex formation being temperature dependent (Muthukrishnan 2014). However, for the Φ10 promoter, the noise in transcription decreases with decreasing temperature. Although it is not known which features are responsible for this behavior, from the adaptation point of view, this could be advantageous for the phage, when facing fluctuating temperature conditions. As such, understanding how this is made possible may be of help in engineering genes with a more robust behavior at suboptimal temperatures. Especially given that some of our results suggest that the mean transcription rate should not decrease significantly with decreasing temperature, provided a constant number of T7 RNAp.

In **Publication III**, we studied the temperature dependence of the transcriptional leakiness of the $LacO_3O_1$ promoter under full repression. For this, we used a single copy plasmid with $LacO_3O_1$ controlling the expression of the target gene and a specially tailored temperature chamber to allow critically low temperatures while under microscope observations.

First, to compare the transcription kinetics of $LacO_3O_1$ when not induced at 24^0C and 37^0C, we conducted microscope measurements of the leakiness rate of this promoter. We found that this rate at 24^0C was 8.5 times higher than at 37^0C, indicating the temperature dependence of the repression mechanism of this promoter.

Next, to understand which processes were most affected by the sub-optimal temperatures, we first assumed a standard model of transcription and repression for this promoter. From the model, we defined which parameters contributed to this leakiness at low temperature. These are the intracellular concentration of RNAp, the transcription rate of a free promoter and the fraction of time that the promoter is free from repressors (which depends on the number of repressors and the rate by which they bind and unbind to the promoter). The former two parameters can be measured, and thus, measurements were taken at each condition and the relative difference between the conditions obtained. From these, along with the absolute changes in leakiness rate with temperature, we finally obtained the degree of change in the efficiency of repression of this promoter (the third parameter) when changing temperature from 37^0C to 24^0C.

We observed that, first, the relative RNAp levels were only slightly higher at 24^0C than at 37^0C, and the transcription rate of the fully induced promoter was also only mildly higher at 37^0C. Neither parameter changes could explain the differences in absolute leakiness rates between the two

conditions. From this we established that, out of the three parameters, the repression efficiency was the most affected by temperature, and the one that contributed the most to the increased leakiness rate at the lower temperature condition, as it changed by one order of magnitude from 37^0C to 24^0C. Given this, it would be of interest to better dissect which components of the repression mechanism are more affected by this decrease in temperature.

In **Publication IV** we characterized, at a single cell level, the intake kinetics of IPTG molecules from the media to the cell's cytoplasm, as a function of temperature. For this, we first made use of a methodology that allows the induction of the target gene in simultaneous with image acquisition. From this, we then determined the time taken for the first RNA to be produced (t_0), and the time between consecutive RNA production events (Δt) in individual cells, for all tested temperatures. Given that it is not possible to directly determine the time taken for inducer molecules to enter the cell (intake time), we made use of two independent methodologies to obtain this value, with the results of both methods being consistent with one another.

These results showed that the mean time for the first RNA to be produced increases, while the variability associated with this process decreases, for temperatures lower than 37^0C. This trend was also observed for the time taken for inducer molecules to enter the cell. Since the intake process can be well modeled by a two-step process (at 37^0C) (Mäkelä *et al.* 2017), we hypothesized that the observed reduction in variability, with decreasing temperature, could be due to the appearance of additional rate-limiting steps at lower temperatures. To test this hypothesis, we estimated the number and the duration of these steps for all tested temperatures. The results were in agreement with our hypothesis, with higher order models (more than 1 step) fitting better the observed data. Additionally, the 2-steps model fitted better the data for cells at 37^0C, which is in line with previous results (Mäkelä *et al.* 2017). For the lower temperatures, models with 3 or more steps fit better the data, suggesting that one or more steps had become rate-limiting with decreasing temperature. The fact that the 4-step model was not enough to model the data for the lowest temperature tested (24^0C), can be interpreted as evidence for changes in the intake kinetics caused by a reduction in temperature.

Given this, we hypothesized that this behavior might be a result of changes in the physical properties of processes associated with the machinery responsible for the intake of molecules, and consequently, with the intracellular movement of these molecules. For instance, it is known that low temperatures lead to an increase in cytoplasm viscosity, which in turn can affect the movement of molecules and cellular components (Oliveira *et al.* 2016b). In addition, temperature downshift can induce physical changes in cellular membranes which in turn can alter the diffusive intake of these inducer molecules. As a consequence, this can alter the single-cell distribution of intake times, which can have a significant impact on the RNA and proteins numbers, thus altering the degree of

heterogeneity of cell populations. In the future, it would be of great interest to further dissect the causes for the reduced cell-to-cell diversity observed with decreasing temperatures, which we believe results from the emergence of rate-limiting steps in the intake process.

Overall, the studies presented in this thesis allow concluding that, in *E. coli*, temperature shifts have a major effect on the transcription dynamics of synthetic genes. Further, these effects depend on the gene's repression mechanism and also on its structure. As a result, these genes, whether alone or integrated in a circuit, might not behave as expected when facing temperature fluctuations. Thus, these properties should be taken into consideration when using synthetic promoters in conditions where temperature fluctuations may occur, particularly when engineering synthetic genes or genetic circuits. In addition, we expect our results to contribute to a better understanding of the dynamics of synthetic genes at critically low temperatures and to near-future efforts in Synthetic Biology.

Bibliography

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. and Walter, P. *Molecular Biology of the Cell*. Garland Science, USA, (2002).

Arkin, A., Ross, J. and McAdams, H. H. ‘‘Stochastic kinetic analysis of developmental pathway bifurcation in phage Lambda-infected Escherichia coli cells’’. *Genetics* 149.4 (1998), pp. 1633–1648.

Arnaud-Barbe, N., Cheynet-Sauvion, V., Oriol, G., Mandrand, B. and Mallet, F. ‘‘Transcription of RNA templates by T7 RNA polymerase.’’ *Nucleic Acids Research* 26.15 (1998), pp. 3550–3554.

Arsène, F., Tomoyasu, T. and Bukau, B. ‘‘The heat shock response of Escherichia coli.’’ *International Journal of Food Microbiology* 55.1 (2000), pp. 3–9.

Axelrod, D. ‘‘Cell-substrate contacts illuminated by total internal reflection fluorescence’’. *Journal of Cell Biology* 89.1 (1981), pp. 141–145.

Babu, M. M. and Teichmann, S. A. ‘‘Evolution of transcription factors and the gene regulatory network in Escherichia coli.’’ *Nucleic Acids Research* 31.4 (2003), pp. 1234–1244.

Barkley, M. D., Riggs, A. D., Jobe, A. and Burgeois, S. ‘‘Interaction of effecting ligands with lac repressor and repressor-operator complex.’’ *Biochemistry* 14.8 (1975), pp. 1700–1712.

Barria, C., Malecki, M. and Arraiano, C. M. ‘‘Bacterial adaptation to cold.’’ *Microbiology* 159.12 (2013), pp. 2437–43.

Becker, N. A., Peters, J. P. and Maher, L. J. ‘‘Mechanism of promoter repression by Lac repressor-DNA loops.’’ *Nucleic Acids Research* 41.1 (2013), pp. 156–166.

Beckwith, J., Grodzicker, T. and Arditti, R. ‘‘Evidence for two sites in the lac promoter region.’’ *Journal of Molecular Biology* 69.1 (1972), pp. 155–160.

Bernstein, J. A., Khodursky, A. B., Lin, P.-H., Lin-Chao, S. and Cohen, S. N. ‘‘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 of the United States of America* 99.15 (2002), pp. 9697–9702.

Bertrand-Burggraf, E., Lefèvre, J. F. and Daune, M. ‘‘A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322.’’ *Nucleic Acids Research* 12.3 (1984), pp. 1697–1706.

Bertrand, E., Chartrand, P., Schaefer, M., Shenoy, S. M., Singer, R. H. and Long, R. M. ‘‘Localization of ASH1 mRNA Particles in Living Yeast.’’ *Molecular Cell* 2.4 (1998), pp. 437–445.

Bervoets, I., Van Brempt, M., Van Nerom, K., Van Hove, B., Maertens, J., De Mey, M. and Charlier, D. ‘‘A sigma factor toolbox for orthogonal gene expression in Escherichia coli.’’ *Nucleic Acids Research* 46.4 (2018), pp. 2133–2144.

Beyreuther, K., Adler, K., Geisler, N. and Klemm, A. ‘‘The amino-acid sequence of lac repressor.’’ *Proceedings of the National Academy of Sciences of the United States of America* 70.12 (1973), pp. 3576–3580.

Blattner, F. R., Plunkett III, G., Bloch, C. A., Perna, N. T., Burland, V., Riley, M., Collado-Vides, J., Glasner, J. D., Rode, C. K., Mayhew, G. F., Gregor, J., Davis, N. W., Kirkpatrick, H. A., Goeden, M. A., Rose, D. J., Mau, B. and Shao, Y. ‘‘The Complete Genome Sequence of Escherichia coli K-12.’’ *Science* 277.5331 (1997), pp. 1453–1462.

Brewster, R. C., Jones, D. L. and Phillips, R. ‘‘Tuning Promoter Strength through RNA Polymerase Binding Site Design in Escherichia coli’’. *PLoS Computational Biology* 8.12 (2012), p. e1002811.

Brödel, A. K., Jaramillo, A. and Isalan, M. ‘‘Engineering orthogonal dual transcription factors for multi-input synthetic promoters.’’ *Nature Communications* 7 (2016), p. 13858.

Brophy, J. A. N. and Voigt, C. A. ‘‘Principles of genetic circuit design.’’ *Nature Methods* 11.5 (2014), pp. 508–520.

Browning, D. F. and Busby, S. J. W. ‘‘The regulation of bacterial transcription initiation.’’ *Nature Reviews Microbiology* 2.1 (2004), pp. 57–65.

Browning, D. F. and Busby, S. J. W. ‘‘Local and global regulation of transcription initiation in bacteria.’’ *Nature Reviews Microbiology* 14.10 (2016), pp. 638–650.

Buc, H. and McClure, W. R. ‘‘Kinetics of open complex formation between Escherichia coli RNA polymerase and the lacUV5 promoter. Evidence for a sequential mechanism involving three steps.’’ *Biochemistry* 24.11 (1985), pp. 2712–2723.

Busby, S. and Ebright, R. H. ‘‘Transcription activation by catabolite activator protein (CAP)’’. *Journal of Molecular Biology* 293.2 (1999), pp. 199–213.

Cardinale, S. and Arkin, A. P. ‘‘Contextualizing context for synthetic biology - identifying causes of failure of synthetic biological systems.’’ *Biotechnology Journal* 7.7 (2012), pp. 856–866.

Chamberlin, M. J. ‘‘The Selectivity of Transcription.’’ *Annual Review of Biochemistry* 43.1 (1974), pp. 721–775.

Chamberlin, M., McGrath, J. and Waskell, L. ‘‘New RNA polymerase from Escherichia coli infected with bacteriophage T7.’’ *Nature* 228 (1970), pp. 227–231.

Chamberlin, M. and Ring, J. ‘‘Characterization of T7-specific ribonucleic acid polymerase.’’ *Journal of Biological Chemistry* 248.6 (1973), pp. 2245–2250.

Chao, J. A., Patskovsky, Y., Almo, S. C. and Singer, R. H. ‘‘Structural basis for the coevolution of viral RNA-protein complex.’’ *Nature Structural and Molecular Biology* 15.1 (2008), pp. 103–105.

Chapman, K. A. and Burgess, R. R. ‘‘Construction of bacteriophage T7 late promoters with point mutations and charactherization by in vitro trasncription properties.’’ *Nucleic Acids Research* 15.13 (1987), pp. 5413–5432.

Chapman, K. A., Gunderson, S. I., Anello, M., Wells, R. D. and Burgess, R. R. ‘‘Bacteriophage T7 late promoters with point mutations: Quantitative footprinting and in vivo expression.’’ *Nucleic Acids Research* 16.10 (1988), pp. 4511–4524.

Chapman, K. A. and Wells, R. D. ‘‘Bacteriophage T7 late promoters: Construction and in vitro transcription properties of deletion mutants.’’ *Nucleic Acids Research* 10.20 (1982), pp. 6331–6340.

Cheetham, G. M. T., Jeruzalmi, D. and Steitz, T. A. ‘‘Structural basis for initiation of transcription from an RNA polymerase-promoter complex.’’ *Nature* 399.6731 (1999), pp. 80–83.

Cheetham, G. M. T. and Steitz, T. A. ‘‘Insights into transcription: structure and function of single-subunit DNA-dependent RNA polymerases.’’ *Current Opinion in Structural Biology* 10.1 (2000), pp. 117–123.

Cho, B.-K., Zengler, K., Qiu, Y., Park, Y. S., Knight, E. M., Barrett, C. L., Gao, Y. and Palsson, B. ‘‘The transcription unit architecture of the Escherichia coli genome.’’ *Nature Biotechnology* 27.11 (2009), pp. 1043–1049.

Choi, P. J., Cai, L., Frieda, K. and Xie, X. S. ‘‘A stochastic single-molecule event triggers phenotype switching of a bacterial cell.’’ *Science* 322.5900 (2008), pp. 442–446.

Chong, S., Chen, C., Ge, H. and Xie, X. S. ‘‘Mechanism of Transcriptional Bursting in Bacteria.’’ *Cell* 158.2 (2014), pp. 314–326.

Cormack, B. P., Valdivia, R. H. and Falkow, S. ‘‘FACS-optimized mutants of the green fluorescent protein (GFP).’’ *Gene* 173.1 (1996), pp. 33–38.

Crick, F. ‘‘Central Dogma of Molecular Biology.’’ *Nature* 227.5258 (1970), pp. 561–563.

Daigle, N. and Ellenberg, J. ‘‘Lambda N-GFP: an RNA reporter system for live-cell imaging.’’ *Nature Methods* 4.8 (2007), pp. 633–636.

Decad, G. M. and Nikaido, H. ‘‘Outer membrane of gram-negative bacteria. XII. Molecular-sieving function of cell wall.’’ *Journal of Bacteriology* 128.1 (1976), pp. 325–336.

Deng, S., Stein, R. A. and Higgins, N. P. ‘‘Organization of supercoil domains and their reorganization by transcription’’. *Molecular Microbiology* 57.6 (2005), pp. 1511–1521.

Dunn, J. J. and Studier, F. ‘‘Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements.’’ *Journal of Molecular Biology* 166.4 (1983), pp. 477–535.

Eliasson, A., Bernander, R., Dasgupta, S. and Nordström, K. ‘‘Direct visualization of plasmid DNA in bacterial cells.’’ *Molecular Microbiology* 6.2 (1992), pp. 165–170.

Elowitz, M. B. and Leibler, S. ‘‘A synthetic oscillatory network of transcriptional regulators.’’ *Nature* 403.6767 (2000), pp. 335–338.

Elowitz, M. B., Levine, A. J., Siggia, E. D. and Swain, P. S. ‘‘Stochastic gene expression in a single cell’’. *Science* 297.5584 (2002), pp. 1183–1186.

Emmer, M., DeCrombrugghe, B., Pastan, I. and Perlman, R. ‘‘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 (1970), pp. 480–487.

Endy, D. ‘‘Foundations for engineering biology.’’ *Nature* 438.7067 (2005), pp. 449–453.

Endy, D., Kong, D. and Yin, J. ‘‘Intracellular kinetics of a growing virus: A genetically structured simulation for bacteriophage T7.’’ *Biotechnology and Bioengineering* 55.2 (1997), pp. 375–389.

Englesberg, E., Irr, J., Power, J. and Lee, N. ‘‘Positive Control of Enzyme Synthesis by Gene C in the Positive Control of Enzyme Synthesis by Gene C in the L-Arabinose System.’’ *Journal of Bacteriology* 90.4 (1965), pp. 946–957.

Faith, J. J., Hayete, B., Thaden, J. T., Mogno, I., Wierzbowski, J., Cottarel, G., Kasif, S., Collins, J. J. and Gardner, T. S. ‘‘Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles.’’ *PLoS Biology* 5.1 (2007), p. e8.

Finkelstein, A. ‘‘Water and nonelectrolyte permeability of lipid bilayer membranes’’. *The Journal of General Physiology* 68.2 (1976), pp. 127–135.

Fulcrand, G., Dages, S., Zhi, X., Chapagain, P., Gerstman, B. S., Dunlap, D. and Leng, F. ‘‘DNA supercoiling, a critical signal regulating the basal expression of the lac operon in Escherichia coli’’. *Scientific Reports* 6.19243 (2016), pp. 1–12.

Fusco, D., Accornero, N., Lavoie, B., Shenoy, S. M., Blanchard, J. M., Singer, R. H. and Bertrand, E. ‘‘Single mRNA molecules demonstrate probabilistic movement in living mammalian cells.’’ *Current Biology* 13.2 (2003), pp. 161–167.

Garcia, L. R. and Molineux, I. J. ‘‘Rate of translocation of bacteriophage T7 DNA across the membranes of Escherichia coli.’’ *Journal of Bacteriology* 177.14 (1995), pp. 4066–4076.

Garcia, L. R. and Molineux, I. J. ‘‘Transcription independent DNA translocation of bacteriophage T7 DNA into Escherichia coli.’’ *Journal of Bacteriology* 178.23 (1996), pp. 6921–6929.

Gardner, T. S., Cantor, C. R. and Collins, J. J. ‘‘Construction of a genetic toggle switch in Escherichia coli.’’ *Nature* 403.6767 (2000), pp. 339–342.

Gibson, D. G. ‘‘Enzymatic assembly of overlapping DNA fragments.’’ *Methods in Enzymology* 498 (2011), pp. 349–361.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison III, C. A., and Smith, H. O. ‘‘Enzymatic assembly of DNA molecules up to several hundred kilobases.’’ *Nature Methods* 6.5 (2009), pp. 343–345.

Gilbert, W. and Maxam, A. ‘‘The nucleotide sequence of the lac operator.’’ *Proceedings of the National Academy of Sciences of the United States of America* 70.12 (1973), pp. 3581–3584.

Goldberg, A. L. ‘‘Degradation of abnormal proteins in Escherichia coli.’’ *Proceedings of the National Academy of Sciences of the United States of America* 69.2 (1972), pp. 422–426.

Golding, I. and Cox, E. C. ‘‘RNA dynamics in live Escherichia coli cells.’’ *Proceedings of the National Academy of Sciences of the United States of America* 101.31 (2004), pp. 11310–11315.

Golding, I., Paulsson, J., Zawilski, S. M. and Cox, E. C. ‘‘Real-time kinetics of gene activity in individual bacteria.’’ *Cell* 123.6 (2005), pp. 1025–1036.

Golomb, M. and Chamberlin, M. ‘‘A preliminary map of the major transcription units read by T7 RNA polymerase on the T7 and T3 bacteriophage chromosomes.’’ *Proceedings of the National Academy of Sciences of the United States of America* 71.3 (1974), pp. 760–764.

Griesbeck, O., Baird, G. S., Campbell, R. E., Zacharias, D. A. and Tsien, R. Y. ‘‘Reducing the environmental sensitivity of yellow fluorescent protein.’’ *Journal of Biological Chemistry* 276.31 (2001), pp. 29188–29194.

Griffiths, A. J., Miller, J. H., Suzuki, D. T., Lewontin, R. C. and Gelbart, W. M. *An introduction to Genetic Analysis*. W. H. Freeman, USA, (2004).

Gummesson, B., Magnusson, L. U., Lovmar, M., Kvint, K., Persson, O., Ballesteros, M., Farewell, A. and Nyström, T. ‘‘Increased RNA polymerase availability directs resources towards growth at the expense of maintenance.’’ *The EMBO Journal* 28.15 (2009), pp. 2209–2219.

Gunasekera, T. S., Csonka, L. N. and Paliy, O. ‘‘Genome-wide transcriptional responses of Escherichia coli K-12 to continuous osmotic and heat stresses.’’ *Journal of Bacteriology* 190.10 (2008), pp. 3712–3720.

Hardy, C. D. and Cozzarelli, N. R. ‘‘A genetic selection for supercoiling mutants of Escherichia coli reveals proteins implicated in chromosome structure’’. *Molecular Microbiology* 57.6 (2005), pp. 1636–1652.

Häkkinen, A., Muthukrishnan, A. B., Mora, A., Fonseca, J. M. and Ribeiro, A. S. ‘‘CellAging: a tool to study segregation and partitioning in division in cell lineages of Escherichia coli.’’ *Bioinformatics* 29.13 (2013), pp. 1708–1709.

Häkkinen, A. and Ribeiro, A. S. ‘‘Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data.’’ *Bioinformatics* 31.1 (2015), pp. 69–75.

Harley, C. B. and Reynolds, R. P. ‘‘Analysis of E. coli promoter sequences.’’ *Nucleic Acids Research* 15.5 (1987), pp. 2343–2361.

Hazel, J. R. ‘‘Thermal Adaptation in Biological-Membranes - Is Homeoviscous Adaptation the Explanation?’’ *Annual Review of Physiology* 57.94 (1995), pp. 19–42.

Helling, R. B. and Weinberg, R. ‘‘Complementation Studies of Arabinose genes in Escherichia coli.’’ *Genetics* 48.10 (1963), pp. 1397–1410.

Higgins, N. P. ‘‘Species-specific supercoil dynamics of the bacterial nucleoid’’. *Biophysical Reviews* 8 (2016), pp. 113–121.

Hocine, S., Raymond, P., Zenklusen, D., Chao, J. A. and Singer, R. H. ‘‘Single-molecule analysis of gene epxression using two-color RNA labeling in live yeast.’’ *Nature Methods* 10.2 (2013), pp. 119–121.

Hogg, R. W. and Englesberg, E. ‘‘L-arabinose binding protein from Escherichia coli B/r.’’ *Journal of Bacteriology* 100.1 (1969), pp. 423–432.

Horvath, I., Glatz, A., Varvasovszki, V., Torok, Z., Pali, T., Balogh, G., Kovacs, E., Nadasdi, L., Benko, S., Joo, F. and Vigh, L. ‘‘Membrane physical state controls the signaling mechanism of the heat shock response in Synechocystis PCC 6803: Identification of hsp17 as a “fluidity gene”.’’ *Proceedings of the National Academy of Sciences of the United States of America* 95.7 (1998), pp. 3513–3518.

Hsu, L. M. ‘‘Promoter clearance and escape in prokaryotes.’’ *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1577.2 (2002), pp. 191–207.

Hussain, F., Gupta, C., Hirning, A. J., Ott, W., Matthews, K. S., Josic, K. and Bennett, M. R. ‘‘Engineered temperature compensation in a synthetic genetic clock’’. *Proceedings of the National Academy of Sciences of the United States of America* 111.3 (2014), pp. 972–977.

Ikeda, R. A., Lin, A. C. and Clarke, J. ‘‘Initiation of transcription by T7 RNA polymerase at its natural promoters.’’ *Journal of Biological Chemistry* 267.4 (1992), pp. 2640–2649.

Ikeda, R. A. ‘‘The efficiency of promoter clearance distinguishes T7 class II and class III promoters.’’ *Journal of Biological Chemistry* 267.16 (1992), pp. 11322–11328.

Inoue, K., Matsuzaki, H., Matsumoto, K. and Shibuya, I. ‘‘Unbalanced membrane phospholipid compositions affect transcriptional expression of certain regulatory genes in Escherichia coli.’’ *Journal of Bacteriology* 179.9 (1997), pp. 2872–2878.

Isaacs, F. J., Hasty, J., Cantor, C. R. and Collins, J. J. ‘‘Prediction and measurement of an autoregulatory genetic module’’. *Proceedings of the National Academy of Sciences of the United States of America* 100.13 (2003), pp. 7714–7719.

Jacob, F. and Monod, J. ‘‘Genetic regulatory mechanism in the sysnthesis of proteins.’’ *Journal of Molecular Biology* 3.3 (1961), pp. 318–356.

Jensen, P. R., Westerhoff, H. V and Michelsen, O. ‘‘The use of lac-type promoters in control analysis.’’ *European Journal of Biochemistry* 211.1–2 (1993), pp. 181–191.

Jobe, A. and Bourgeois, S. ‘‘Lac repressor–operator interaction: VI. The natural inducer of the lac operon.’’ *Journal of Molecular Biology* 69.3 (1972), p. 397–408.

Johnson, C. M. and Schleif, R. F. ‘‘In vivo induction kinetics of the arabinose promoters in Escherichia coli.’’ *Journal of Bacteriology* 177.12 (1995), pp. 3438–3442.

Kaback, H. R. ‘‘The lac carrier protein in Escherichia coli.’’ *Journal of Membrane Biology* 76.2 (1983), pp. 95–112.

Kandavalli, V. K., Tran, H. and Ribeiro, A. S. ‘‘Effects of σ factor competition are promoter initiation kinetics dependent.’’ *Biochimica et Biophysica Acta (BBA)- Gene Regulatory Mechanisms* 1859.10 (2016), pp. 1281–1288.

Kania, J. and Brown, D. T. ‘‘The functional repressor parts of a tetrameric lac repressor-beta-galactosidase chimaera are organized as dimers.’’ *Proceedings of the National Academy of Sciences of the United States of America* 73.10 (1976), pp. 3529–3533.

Kannan, G., Wilks, J. C., Fitzgerald, D. M., Jones, B. D., BonDurant, S. S. and Slonczewski, J. L. ‘‘Rapid acid treatment of Escherichia coli: transcriptomic response and recovery.’’ *BMC Microbiology* 8.1 (2008), p. 1.

Kneen, M., Farinas, J., Li, Y. and Verkman, A. S. ‘‘Green fluorescent protein as a noninvasive intracellular pH indicator.’’ *Biophysical Journal* 74.3 (1998), pp. 1591–1599.

Lange, S., Katayama, Y., Schmid, M., Burkacky, O., Brauchle, C., Lamb D. C. and Jansen, R. P. ‘‘Simultaneous transport of different localized mRNA species revealed by live-cell imaging.’’ *Traffic* 9.8 (2008), pp. 1256–1267.

Lanzer, M. and Bujard, H. ‘‘Promoters largely determine the efficiency of repressor action.’’ *Proceedings of the National Academy of Sciences of the United States of America* 85.23 (1988), pp. 8973–8977.

Larson, D. R., Zenklusen, D., Wu, B., Chao, J. A. and Singer, R. H. ‘‘Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene.’’ *Science* 332.6028 (2011), pp. 475–478.

Lee, D. J., Minchin, S. D. and Busby, S. J. W. ‘‘Activating Transcription in Bacteria’’. *Annual Review of Microbiology* 66.1 (2012), pp. 125–152.

Lee, J. H., Al-Zarban, S. and Wilcox, G. ‘‘Genetic characterization of the araE gene in Salmonella typhimurium LT2.’’ *Journal of Bacteriology* 146.1 (1981), pp. 298–304.

Lee, P. S. and Lee, K. H. ‘‘Escherichia coli - A Model System That Benefits from and Contributes to the Evolution of Proteomics.’’ *Biotechnology and Bioengineering* 84.7 (2003), pp. 801–814.

Leng, F., Chen, B. and Dunlap, D. D. ‘‘Dividing a supercoiled DNA molecule into two independent topological domains’’. *Proceedings of the National Academy of Sciences of the United States of America* 108.50 (2011), pp. 19973–19978.

Lewis, M. ‘‘The lac repressor.’’ *Comptes Rendus - Biologies* 328.6 (2005), pp. 521–548.

Lewis, M., Chang, G., Horton, N. C., Kercher, M. A., Pace, H. C., Schumacher, M. A., Brennan, R. G. and Lu, P. ‘‘Crystal structure of the lactose operon repressor and its complexes with DNA and inducer.’’ *Science* 271.5253 (1996), pp. 1247–1254.

Li, T., Ho, H. H., Maslak, M., Schick, C. and Martin, C. T. ‘‘Major groove recognition elements in the middle of the T7 RNA polymerase promoter.’’ *Biochemistry* 35.12 (1996), pp. 3722–3727.

Lineweaver, H. and Burk, D. ‘‘The determination of enzyme dissociation constants’’. *Journal of the American Chemical Society* 56.3 (1934), pp. 658–666.

Liu, C. and Martin, C. T. ‘‘Promoter clearance by T7 RNA polymerase. Initial bubble collapse and transcript dissociation monitored by base analog fluorescence.’’ *Journal of Biological Chemistry* 277.4 (2002), pp. 2725–2731.

Liu, L. F. and Wang, J. C. ‘‘Supercoiling of the DNA template during transcription.’’ *Proceedings of the National Academy of Sciences of the United States of America* 84.20 (1987), pp. 7024–7027.

Livak, K. J. and Schmittgen, T. D. ‘‘Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCT method.’’ *Methods* 25.4 (2001), pp. 402–408.

Lloyd-Price, J., Startceva, S., Chandraseelan, J. G., Kandavalli, V., Goncalves, N., Oliveira, S. M. D., Häkkinen, A. and Ribeiro, A. S. ‘‘Dissecting the stochastic transcription initiation process in live Escherichia coli.’’ *DNA Research* 23.3 (2016), pp. 203–214.

Lobell, R. and Schleif, R. F. ‘‘DNA looping and unlooping by AraC protein.’’ *Science* 250.4980 (1990), pp. 528–532.

Los, D. A. and Murata, N. ‘‘Membrane fluidity and its roles in the perception of environmental signals.’’ *Biochimica et Biophysica Acta - Biomembranes* 1666.1–2 (2004), pp. 142–157.

López-García, P. and Forterre, P. ‘‘DNA topology and the thermal stress response, a tale from mesophiles and hyperthermophiles.’’ *BioEssays* 22.8 (2000), pp. 738–746.

Lutz, R. and Bujard, H. ‘‘Independent and tight regulation of trasncriptional units in Escherichia coli via the LacR/O, the TetR/O and Arac/I1-I2 regulatory elements.’’ *Nucleic Acids Research* 25.6 (1997), pp. 1203–1210.

Lutz, R., Lozinski, T., Ellinger, T. and Bujard, H. ‘‘Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator.’’ *Nucleic Acids Research*. 29.18 (2001), pp. 3873–3881.

Lynch, A. S. and Wang, J. C. ‘‘Anchoring of DNA to the bacterial cytoplasmic membrane through cotranscriptional synthesis of polypeptides encoding membrane proteins or proteins for export: A mechanism of plasmid hypernegative supercoiling in mutants deficient in DNA topoisomerase I’’. *Journal of Bacteriology* 175.6 (1993), pp. 1645–1655.

Mäkelä, J., Kandavalli, V. and Ribeiro, A. S. ‘‘Rate-limiting steps in transcription dictate sensitivity to variability in cellular components’’. *Scientific Reports* 7.1 (2017), p. 10588.

Marbach, A. and Bettenbrock, K. ‘‘Lac operon induction in Escherichia coli: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA.’’ *Journal of Biotechnology* 157.1 (2012), pp. 82–88.

Matsuyama, S., Llopis, J., Deveraux, Q. L., Tsien, R. Y. and Reed, J. C. ‘‘Changes in intramitochondrial and cytosolic pH: early events that modulate caspase activation during apoptosis’’. *Nature Cell Biology* 2 (2000), pp. 318–325.

Megerle, J. A., Fritz, G., Gerland, U., Jung, K. and Rädler, J. O. ‘‘Timing and dynamics of single cell gene expression in the arabinose utilization system’’. *Biophysical Journal* 95.4 (2008), pp. 2103–2115.

McAllister, W. T. and Barrett, C. L. ‘‘Roles of the Early Genes of Bacteriophage T7 in Shutoff of Host Macromolecular Synthesis.’’ *Journal of Virology* 23.3 (1977), pp. 543–553.

McAllister, W. T. and Carter, A. D. ‘‘Regulation of promoter selection by the bacteriophage T7 RNA polymerase in vitro.’’ *Nucleic Acids Research* 8.20 (1980), pp. 4821–4838.

McAllister, W. T. and Wu, H.-L. ‘‘Regulation of transcription of the late genes of bacteriophage T7.’’ *Proceedings of the National Academy of Sciences of the United States of America* 75.2 (1978), pp. 804–808.

McClure, W. ‘‘Rate-limiting steps in RNA chain initiation.’’ *Proceedings of the National Academy of Sciences of the United States of America* 77.10 (1980), pp. 5634–5638.

McClure, W. R. ‘‘Mechanism and control of transcription initiation in prokaryotes.’’ *Annual Review of Biochemistry* 54.1 (1985), pp. 171–204.

McClure, W. R., Cech, C. L. and Johnston, D. E. ‘‘A steady state assay for the RNA polymerase initiation reaction.’’ *The Journal of Biological Chemistry* 253.24 (1978), pp. 8941–8948.

Miller, J. H. ‘The lacI Gene: Its Role in lac Operon Control and Its Use as a Genetic System’. Cold Spring Harbor Monograph Archive 7 (1980), pp. 31-88.

Miller, O. L., Hamkalo, B. A. and Thomas, C. A. ‘‘Visualization of Bacterial Genes in Action.’’ *Science* 169.3943 (1970), pp. 392–395.

Moffatt, B. A. and Studier, F. W. ‘‘Entry of bacteriophage T7 DNA into the cell and escape from host restriction.’’ *Journal of Bacteriology* 170.5 (1988), pp. 2095–2105.

Molineux, I. J. ‘‘The T7 group.’’ In *The Bacteriophages*. Oxford University Press, UK, (2005) pp. 277–301.

Montero Llopis, P., Jackson, A. F., Sliusarenko, O., Surovtsev, I., Heinritz, J., Emonet, T. and Jacobs-Wagner, C. ‘‘Spatial organization of the flow of genetic information in bacteria.’’ *Nature* 466.7302 (2010), pp. 77–81.

Muller, D. K., Martin, C. T. and Coleman, J. E. ‘‘T7 RNA polymerase interacts with its promoter from one side of the DNA helix.’’ *Biochemistry* 28.8 (1989), pp. 3306–3313.

Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O. and Darst, S. A. ‘‘Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex.’’ *Science* 296.5571 (2002), pp. 1285–1290.

Muthukrishnan, A.B. 2014, “*Studies of the plasticity of transcription in Escherichia coli using single-molecule, in vivo detection techniques*”, PhD thesis, Tampere University of Technolgy.

Nagai, T., Ibata, K., Park, E. S., Kubota, M., Mikoshiba, K. and Miyawaki, A. ‘‘A Variant of Yellow Fluorescent Protein with Fast and Efficient Maturation for Cell-Biological Applications.’’ *Nature Biotechnology* 20 (2002), pp. 87–90.

Nandagopal, N. and Elowitz, M. B. ‘‘Synthetic Biology: Integrated Gene Circuits.’’ *Science* 333.6047 (2011), pp. 1244–1248.

Nikaido, H. ‘‘Molecular basis of bacterial outer membrane permeability revisited’’ *Microbiology and Molecular Biology Reviews* 67.4 (2003), pp. 593–656.

Nudler, E. and Gottesman, M. E. ‘‘Transcription termination and anti-termination in E. coli.’’ *Genes to Cells* 7.8 (2002), pp. 755–768.

Oehler, S., Amouyal, M., Kolkhof, P., von Wilcken-Bergmann, B. and Müller-Hill, B. ‘‘Quality and position of the three lac operators of E. coli define efficiency of repression.’’ *The EMBO Journal* 13.14 (1994), pp. 3348–3355.

Oehler, S., Eismann, E. R., Krämer, H. and Müller-Hill, B. ‘‘The three operators of the lac operon cooperate in repression.’’ *The EMBO Journal* 9.4 (1990), pp. 973–979.

Ohshima, Y., Mizokoshi, T. and Horiuchi, T. ‘‘Binding of an inducer to the lac repressor.’’ *Journal of Molecular Biology* 89.1 (1974), pp. 127–136.

Oliveira, S. M. D., Chandraseelan, J. G., Häkkinen, A., Goncalves, N. S. M., Yli-Harja, O., Startceva, S. and Ribeiro, A. S. ‘‘Single-cell kinetics of a repressilator when implemented in a single-copy plasmid’’. *Molecular Biosystems* 11.7 (2015), pp. 1939–1945.

Oliveira, S. M. D., Häkkinen, A., Lloyd-price, J., Tran, H., Kandavalli, V. and Ribeiro, A. S. ‘‘Temperature-Dependent Model of Multi-step Transcription Initiation in Escherichia coli Based on Live Single-Cell Measurements.’’ *PLoS Computational Biology* 12.10 (2016a), p. e1005174.

Oliveira, S. M. D., Neeli-Venkata, R., Goncalves, N. S. M., Santinha, J. A., Martins, L., Tran, H., Mäkelä, J., Gupta, A., Barandas, M., Häkkinen, A., Lloyd-Price, J., Fonseca, J. M. and Ribeiro, A. S. ‘‘Increased cytoplasm viscosity hampers aggregate polar segregation in Escherichia coli’’. *Molecular Microbiology* 99.4 (2016b), pp. 686–699.

Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. and van Oudenaarden, A. ‘‘Multistability in the lactose utilization network of Escherichia coli.’’ *Nature* 427.6976 (2004), pp. 737–740.

Pawley, J. B. *Handbook of Biological Confocal Microscopy* (3rd ed.). Berlin: Springer, 2006.

Peabody, D. S. ‘‘The RNA binding site of bacteriophage MS2 coat protein.’’ *The EMBO Journal* 12.2 (1993), pp. 595–600.

Pédelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. and Waldo, G. S. ‘‘Engineering and characterization of a superfolder green fluorescent protein.’’ *Nature Biotechnology* 24.1 (2006), pp. 79–88.

Penumetcha, P., Lau, K., Zhu, X., Davis, K., Eckdahl, T. T. and Campbell, A. M. ‘‘Improving the Lac System for Synthetic Biology.’’ *Bios* 81.1 (2010), pp. 7–15.

Pfahl, M., Gulde, V. and Bourgeois, S. ‘‘‘‘Second” and “third operator” of the lac operon: An investigation of their role in the regulatory mechanism.’’ *Journal of Molecular Biology* 127.3 (1979), pp. 339–334.

Phadtare, S. ‘‘Recent developments in bacterial cold-shock response.’’ *Current Issues in Molecular Biology* 6.2 (2004), pp. 125–136.

Phadtare, S. and Severinov, K. ‘‘RNA remodeling and gene regulation by cold shock proteins’’. *RNA Biology* 7.6 (2010), pp. 788–795.

Pitchiaya, S., Heinicke, L. A., Custer, T. C. and Walter, N. G. ‘‘Single molecule fluorescence approaches shed light on intracellular RNAs.’’ *Chemical Reviews* 114.6 (2014), pp. 3224–3265.

Pruss, G. J. and Drlica, K. ‘‘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 (1986), pp. 8952–6.

Qimron, U., Tabor, S. and Richardson, C. C. ‘‘New details about bacteriophage T7-host interactions.’’ *Microbe* (2010), pp. 117–122.

Raj, A. and van Oudenaarden, A. ‘‘Single-molecule approaches to stochastic gene expression.’’ *Annual Review of Biophysics* 38 (2009), pp. 255–270.

Ramakrishnan, V. ‘‘Ribosome structure and the mechanism of translation.’’ *Cell* 108.4 (2002), pp. 557–572.

Ramos, S. and Kaback, H. R. ‘‘The relationship between the electrochemical proton gradient and active transport in Escherichia coli membrane vesicles.’’ *Biochemistry* 16.5 (1977), pp. 854–859.

Reznikoff, W. S., Siegele, D. A., Cowing, D. W. and Gross, C. A. ‘‘The regulation of transcription initiation in bacteria.’’ *Annual Review of Genetics* 19.1 (1985), pp. 355–387.

Reznikoff, W. S., Winter, R. B. and Hurley, C. K. ‘‘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 (1974), pp. 2314–2318.

Ribeiro, A. S., Zhu, R. and Kauffman, S. A. ‘‘A General Modeling Strategy for Gene Regulatory Networks with Stochastic Dynamics.’’ *Journal of Computational Biology* 13.9 (2006), pp. 1630–1639.

Riggs, A. D., Suzuki, H. and Bourgeois, S. ‘‘Lac repressor-operator interaction: I. Equilibrium studies.’’ *Journal of Molecular Biology* 48.1 (1970), pp. 67–83.

Sadler, J. R., Sasmor, H. and Betz, J. L. ‘‘A perfectly symmetric lac operator binds the lac repressor very tightly.’’ *Proceedings of the National Academy of Sciences of the United States of America* 80.22 (1983), pp. 6785–6789.

Saecker, R. M., Record, M. T. J. and deHaseth, P. L. ‘‘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 (2011), pp. 754–771.

Samul, R. and Leng, F. ‘‘Transcription-coupled Hypernegative Supercoiling of Plasmid DNA by T7 RNA Polymerase in Escherichia coli Topoisomerase I-Deficient Strains’’. *Journal of Molecular Biology* 374.4 (2007), pp. 925–935.

Santinha, J., Martins, L., Häkkinen, A., Lloyd-Price, J., Oliveira, S. M. D., Gupta, A., Annila, T., Mora, A., Ribeiro, A. S. and Fonseca, J. R. ‘‘iCellFusion: Tool for fusion and analysis of live-cell

images from time-lapse multimodal microscopy.'' in *Biomedical Image Analysis and Mining Techniques for Improved Health Outcomes, IGI Global*. IGI Global (2015), pp. 71–99.

Schlax, P. J., Capp, M. W. and Record, M. T. ''Inhibition of transcription initiation by lac repressor.'' *Journal of Molecular Biology* 245.4 (1995), pp. 331–350.

Schleif, R. ''Regulation of the L-arabinose operon of Escherichia coli.'' *Trends in Genetics* 16.12 (2000), pp. 559–565.

Schleif, R. ''AraC protein, regulation of the L-arabinose operon in Escherichia coli, and the light switch mechanism of AraC action.'' *FEMS Microbiology Reviews* 34.5 (2010), pp. 779–796.

Schmittgen, T. D. and Livak, K. J. ''Analyzing real-time PCR data by the comparative CT method.'' *Nature Protocols* 3.6 (2008), pp. 1101–1108.

Schmitz, A. and Galas, D. J. ''The interaction of RNA polymerase and lac repressor with the lac control region.'' *Nucleic Acids Research* 6.1 (1979), pp. 111–137.

Schmitz, A., Schmeissner, U., Miller, J. H. and Lu, P. ''Mutations affecting the quaternary structure of the lac repressor.'' *Journal of Biological Chemistry* 251.11 (1976), pp. 3359–3366.

Segall-Shapiro, T. H., Sontag, E. D. and Voigt, C. A. ''Engineered promoters enable constant gene expression at any copy number in bacteria.'' *Nature Biotechnology* 36.4 (2018), pp. 352–358.

Shaner, N. C., Steinbach, P. A. and Tsien, R. Y. ''A guide to choosing fluorescent proteins.'' *Nature Methods* 2.12 (2005), pp. 905–909.

Shen, Y., Rosendale, M., Campbell, R. E. and Perrais, D. ''pHuji, a pH-sensitive red fluorescent protein for imaging of exo- and endocytosis.'' *Journal of Cell Biology* 207.3 (2014), pp. 419–432.

Sheu, C. F. and Ratcliff, R. ''The application of fourier deconvoution to reaction time data: a cautionary note''. *Psychological Bulletin* 118.2 (1995), pp. 285–299.

Shuman, H. A. and Silhavy, T. J. ''The art and design of genetic screens: Escherichia coli''. *Nature Reviews Genetics* 4.6 (2003), pp. 419–431.

Sinensky, M. ''Homeoviscous Adaptation-A Homeostatic Process that Regulates the Viscosity of Membrane Lipids in Escherichia coli''. *Proceedings of the National Academy of Sciences of the United States of America* 71.2 (1974), pp. 522–525.

Stoebel, D. M., Hokamp, K., Last, M. S. and Dorman, C. J. ‘‘Compensatory evolution of gene regulation in response to stress by Escherichia coli lacking RpoS.’’ *PLoS Genetics* 5.10 (2009), p. e1000671.

Stricker, J., Cookson, S., Bennett, M. R., Mather, W. H., Tsimring, L. S. and Hasty, J. ‘‘A fast, robust and tunable synthetic gene oscillator.’’ *Nature* 456.7221 (2008), pp. 516–519.

Studier, F. W. and Moffatt, B. A. ‘‘Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes.’’ *Journal of Molecular Biology* 189.1 (1986), pp. 113–130.

Talbot, S. J., Goodman, S., Bates, S. R. E., Fishwick, C. W. G. and Stockley, P. G. ‘‘Use of synthetic oligoribonucleotides to probe RNA-protein interactions in the MS2 translational operator complex.’’ *Nucleic Acids Research* 18.12 (1990), pp. 3521–3528.

Tao, H., Bausch, C., Richmond, C., Blattner, F. R. and Conway, T. ‘‘Functional Genomics: Expression Analysis of Escherichia coli Growing on Minimal and Rich Media.’’ *Journal of Bacteriology* 181.20 (1999), pp. 6425–6440.

Tokunaga, M., Imamoto, N. and Sakata-Sogawa, K. ‘‘Highly inclined thin illumination enables clear single-molecule imaging in cells.’’ *Nature Methods* 5.2 (2008), pp. 159-161.

Tsien, R. Y. ‘‘The green fluorescent protein.’’ *Annual Review of Biochemistry* 67 (1998), pp. 509–544.

Ujvári, A. and Martin, C. ‘‘Identification of a minimal binding element within the T7 RNA polymerase promoter.’’ *Journal of Molecular Biology* 273.4 (1997), pp. 775–781.

Ujvári, A. and Martin, C. ‘‘Evidence for DNA bending at the T7 RNA polymerase promoter.’’ *Journal of Molecular Biology* 295.5 (2000), pp. 1173–1184.

Valegard, K., Murray, J. B., Stonehouse, N. J., van den Worm, S., Stockley, P. G. and Liljas, L. ‘‘The three-dimensional structure of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions.’’ *Journal of Molecular Biology* 270.5 (1997), pp. 724–738.

Vigh, L., Maresca, B. and Harwood, J. L. ‘‘Does the membrane’s physical state control the expression of heat shock and other genes?’’ *Trends in Biochemical Sciences* 23.10 (1998), pp. 369–374.

von Hippel, P. H., Bear, D. G., Morgan, W. D. and McSwiggen, J. A. ‘‘Protein-Nucleic Acid Interactions in Transcription: A Molecular Analysis.’’ *Annual Review of Biochemistry* 53.1 (1984), pp. 389–446.

Walter, G., Zillig, W., Palm, P. and Fuchs, E. ‘‘Initiation of DNA-Dependent RNA synthesis and the effect of heparin on RNA polymerase.’’ *European Journal of Biochemistry* 3.2 (1967), pp. 194–201.

Wang, F. and Greene, E. C. ‘‘Single-molecule studies of transcription: From one RNA polymerase at a time to the gene expression profile of a cell.’’ *Journal of Molecular Biology* 412.5 (2011), pp. 814–831.

Willetts, N. S. ‘‘Intracellular protein breakdown in growing cells of Escherichia coli.’’ *The Biochemical Journal* 103.2 (1967), pp. 462–466.

Willey, J. M., Sherwood, L. M. and Woolverton, C. J. ‘*Prescott, Harley, & Klein’s Microbiology*’. McGraw Hill, USA, 2008.

Wösten, M. M. S. M. ‘‘Eubacterial sigma-factors.’’ *FEMS Microbiology Reviews* 22.3 (1998), pp. 127–150.

Xie, X. S., Choi, P. J., Li, G.-W., Lee, N. K. and Lia, G. ‘‘Single-molecule approach to molecular biology in living bacterial cells.’’ *Annual Review of Biophysics* 37 (2008), pp. 417–444.

Yamanaka, K. ‘‘Cold Shock Response in Escherichia coli.’’ *Journal of Molecular Microbiology and Biotechnology* 1.2 (1999), pp. 193–202.

Young, B. A., Gruber, T. M., Gross, C. A. and Francisco, S. ‘‘Views of Transcription Initiation.’’ *Cell* 109.4 (2002), pp. 417–420.

Yu, J., Xiao, J., Ren, X., Kaiqin, L. and Xie, X. S. ‘‘Probing Gene Expression in Live Cells, One Molecule at a Time.’’ *Science* 311.5767 (2006), pp. 1600–3.

Zavriev, S. and Shemyakin, M. ‘‘RNA polymerase-dependent mechanims for the stepwise T7 phage DNA transprot from the virion into E. coli.’’ *Nucleic Acids Research* 10.5 (1982), pp. 1635–1652.

Zernike, F. ‘‘Phase Contrast, a new method for the microscopic observation of transparent objects’’. *Physica* 9.10 (1942), pp. 974–986.

Zimmermann, W. and Rosselet, A. ‘‘Function of the outer membrane of Escherichia coli as a permeability barrier to beta-lactam antibiotics.’’ *Antimicrobial Agents and Chemotherapy* 12.3 (1977), pp. 368–372.

Zubay, G., Schwartz, D. and Beckwith, J. ‘‘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 (1970), pp. 104–11.

PUBLICATIONS

PUBLICATION I

H. Tran, S.M.D. Oliveira, N.S.M. Goncalves, and A.S. Ribeiro, “Kinetics of the cellular intake of a gene expression inducer at high concentrations”, *Molecular Biosystems*, 11:2579-2587, 2014.

Molecular BioSystems

PAPER

Cite this: *Mol. BioSyst.*, 2015, 11, 2579

Kinetics of the cellular intake of a gene expression inducer at high concentrations†

Huy Tran, Samuel M. D. Oliveira, Nadia Goncalves and Andre S. Ribeiro*

Received 7th April 2015,
Accepted 20th July 2015

DOI: 10.1039/c5mb00244c

www.rsc.org/molecularbiosystems

From *in vivo* single-event measurements of the transient and steady-state transcription activity of a single-copy *lac-ara-1* promoter in *Escherichia coli*, we characterize the intake kinetics of its inducer (IPTG) from the media. We show that the empirical data are well-fit by a model of intake assuming a bilayer membrane, with the passage through the second layer being rate-limiting, coupled to a stochastic, sub-Poissonian, multi-step transcription process. Using this model, we show that for a wide range of extracellular inducer levels (up to 1.25 mM) the intake process is diffusive-like, suggesting unsaturated membrane permeability. Inducer molecules travel from the periplasm to the cytoplasm in, on average, 31.7 minutes, strongly affecting cells' response time. The novel methodology followed here should aid the study of cellular intake mechanisms at the single-event level.

1. Introduction

Many genes in *Escherichia coli* are kept inactive by constitutive repressors, unless specific inducers appear in the media.[1–3] The kinetics of the transcriptional response to the introduction of inducers into the media depends both on the genetic target system[4–6] as well as on the mechanisms of the intake of the inducer into cells' cytoplasm. By regulating the kinetics of the intake as a function of the inducer numbers in the media, the intake system allows cells, among other things, to adjust to fluctuations in the inducer extracellular concentration. One of the best studied intake mechanisms is the one responsible for the intake of lactose and its analogues, such as isopropyl β-D-1-thiogalactopyranoside (IPTG).[7–11]

Early studies of this system focused on the observation of the target gene's expression at the steady state, as a function of the inducer concentration in the media.[1,8,9] More recent studies have focused on the transient dynamics of the inducible gene, following the introduction of inducers into the media[10] so as to study the intake mechanism of the inducer molecules. These studies were, in general, conducted in the regime of low IPTG concentration (usually below 0.5 mM), where the mean expression rate of the target gene exhibits a close-to-linear dependence on the intracellular inducer level.[8,9,11] In this regime, the dynamics of intake of IPTG is in agreement with the existence of positive feedback, *i.e.*, upon entering cells, IPTG activates the lac operon, thus triggering the production of *lacY*, a permease protein that enhances the intake of IPTG.[8,9,11]

Meanwhile, in the regime of high concentrations, in which *lacY* no longer is the major contributor of intake,[8,9] the behavior of the intake process of IPTG is less explored, as the transient period is shorter and thus less well captured using standard measurement techniques (*e.g.* qPCR or GFP expression). Also, direct measurements of inducer levels in cells and media[12] are only accurate on high-density cultures able to deplete the media of inducers, which causes the cellular intake kinetics to vary over time.

The advent of *in vivo* single RNA molecule measurement techniques, based on the tagging of RNA by MS2d-GFP fluorescent proteins,[13] now allows exploring this regime in detail, since it allows measuring fast responses due to detecting RNA molecules as soon as these are produced. In addition, it is possible to maintain a constant concentration of inducers in the media during measurements. This technique has recently been used to characterize the transcription kinetics of some promoters in *E. coli*,[6,14] revealing that, *e.g.*, $P_{lac\text{-}ara\text{-}1}$ transcription initiation is a multi-stepped process and IPTG mainly affects one of the two rate-limiting steps, likely the closed complex formation.[15,16]

Here, using live, single-cell, time-lapse microscopy and MS2-tagging of RNA that allows the detection of each RNA soon after production,[13,17] we measure the time that it takes cells to produce the first target RNA, following the introduction of inducer into the media, as a function of the extracellular inducer concentration in the regime of high concentrations (from 0.25 mM to 1.25 mM).[5] We then use methods of statistical inference to derive from the empirical data a deterministic model of inducer intake through a bilayer membrane,[18] coupled with a stochastic, multi-step model of transcription.[19] By fitting the model to the moments

Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology, FI-33101 Tampere, Finland.
E-mail: andre.ribeiro@tut.fi; Fax: +358 33115498; Tel: +358 408490736

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c5mb00244c

of appearance of the first RNA in each cell, we evaluate the significance of the time it takes inducers to cross the two layers of the cells' wall in the waiting times for RNA appearances, as a function of the extracellular inducer concentration. Given this, we characterize the intake mechanism of IPTG, in the range of IPTG concentrations tested.

2. Methods

Bacterial strain and plasmids

The *E. coli* strain used here is DH5α-PRO, generously provided by Ido Golding, University of Illinois, USA. This strain contains two genes, *lacI* and *tetR*, which are constitutively overexpressed under the control of P^{q}_{laci} and PN25 promoters.[20] The native lac operon (*lacZYA*) is mutated to prevent the production of additional permease proteins (*lacY*) and the activation of the lactose metabolic system. The cells also contain two constructs: pROTET-K133 carrying $P_{LtetO-1}$-MS2d-GFP and pIG-BAC, a single-copy plasmid, containing $P_{lac\text{-}ara\text{-}1}$-mRFP1-MS2-96bs (see Fig. 1).

Media and growth conditions

Cells were grown overnight at 30 °C with aeration and shaking in Luria-Bertani (LB) medium, supplemented with the necessary antibiotics. Cells were then diluted in fresh M63 medium. When reaching an optical density of OD600 ≈ 0.3–0.5, cells were pre-incubated for 45 min with 100 ng ml^{-1} anhydrotetracycline (aTc) to produce enough matured MS2d-GFP proteins to detect RNAs at the start of the microscopy measurements. During microscopy, cells were kept in M63 (ESI†) as, by inspection, we observed that it reduces leaky expression (compared to LB media).

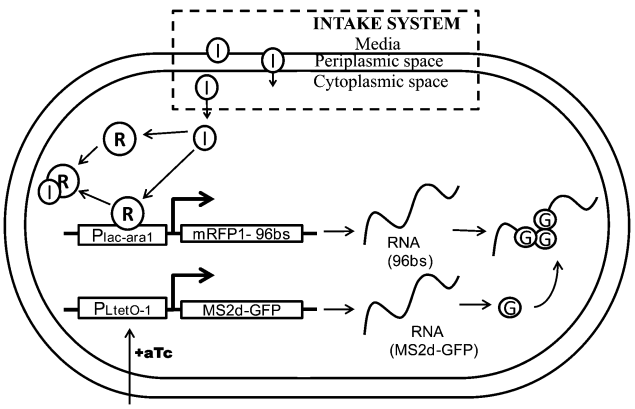


Fig. 1 Diagram of the inducer intake system, target gene and RNA tagging system: IPTG molecules (*I*) in the media enter the cytoplasm by passing through two membrane layers, with a periplasmic space in between. When in the cytoplasm, they neutralize *lacI* repressors (*R*) by forming inducer–repressor complexes (*RI*). Once the repression of $P_{lac\text{-}ara1}$ is hampered, the target gene is free to express. It codes for an RNA that includes an mRFP1 coding region and an array of 96 MS2-binding sites.[17] MS2d-GFP expression is controlled by $P_{LtetO-1}$ promoters and anhydrotetracycline (aTc). Once transcribed, the target transcript is bound by multiple tagging MS2d-GFP proteins (*G*) and rapidly appears as a bright spot under the confocal microscope.[13]

Microscopy and image analysis

Protocols for microscopy sessions are fully described in ref. 5. Time series are 3 hours long, with cells being imaged once per minute. For image analysis, we use semi-automatic cell segmentation and RNA spot detection strategies[21,22] (ESI†).

The moment when the first RNA appears in each cell (denoted as t_0) and the subsequent intervals between consecutive RNA production events (denoted as Δt) are extracted from the time series of total spot intensities as in ref. 5, 6, 14, 15, 23 and 24. The method is described in the ESI.† Also, in Section VII of the ESI,† we provide an empirical validation for the method of detecting, from the time series of total spot intensity in each cell, the moments when novel RNA molecules first appear. Data for t_0 were collected from the first 2 hours of the measurements, while data for Δt are collected from the third hour alone, in order to ensure that cells are fully induced by the time these intervals are collected.

Data analysis

Our empirical data, extracted from the microscopy, consist of the time for the appearance of the first RNA (t_0) and the subsequent intervals between consecutive transcription events (Δt) in each cell. Due to cell divisions and limited measurement time, along with t_0 and Δt being of the order of hundreds of seconds,[5,15] larger values of t_0 and Δt might not be detectable, resulting in the underestimation of their mean values. To exemplify this consider that, in the absence of induction, the measured leaky RNA production of $P_{lac\text{-}ara\text{-}1}$ is <0.1 RNA per h per cell,[5] suggesting a Δt's true mean of (at least) 10 hours. However, since our measurement time for Δt is 1 hour long, the mean of the few measured intervals would be smaller than 1 hour, resulting in the underestimation of the true mean value of Δt.

To address this problem, we make use of the information from the lack of production events. Namely, we make use of 'right censored' data from each cell, which consists of the time from the last production event until cell division or until the end of the time series. Combining the data from observed production events (actual sampled values of t_0 and Δt) with the right-censored data (from the lack of productions) results in data that more properly inform on the true distributions of t_0 and Δt (Fig. 2). This happens because, as one conditions the actual samples with censored data (ESI†), the bias on the actual samples (favoring shorter durations) is removed.

The methodology followed in the collection of the censored data is described in the ESI.† For t_0, the actual and censored samples are denoted as t_0' and c_0', respectively. For Δt, the denotations are $\Delta t'$ and $\Delta c'$, respectively. When fitting the theoretical models of t_0 and Δt to the empirical data using the maximum likelihood estimation, we search for the model parameters that maximize the probability of obtaining both the actual samples and the censored samples.[25,26] To measure the goodness of fit of the estimation, we find the model distribution of t_0 and Δt subject to censoring and use statistical tests to verify whether the actual samples can be drawn from the distribution with the estimated parameters (ESI†).

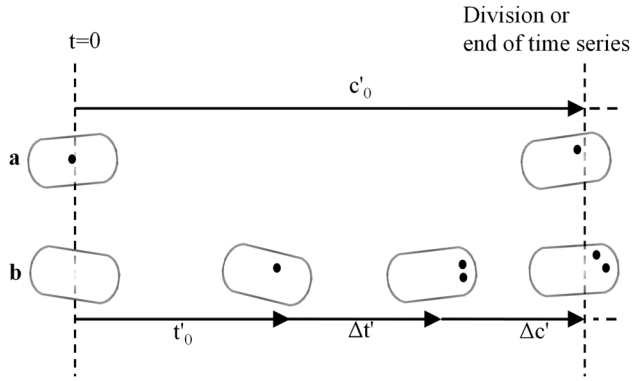


Fig. 2 Collection of t_0 and Δt samples subject to censoring: from cells in the initial population that do not produce any RNA during their lifetime (a), we obtain censored samples of t_0, denoted c_0', whose value equals the cell lifetime. From cells that produce at least one RNA (b), we obtain actual samples, denoted t_0', of t_0 equal to the earliest moment of RNA appearance. Regarding the samples of Δt, the actual samples ($\Delta t'$) are the intervals between consecutive transcription events, whereas the censored samples ($\Delta c'$) are the intervals between the moment of appearance of the last RNA and either the moment of division or the end of the measurement.

Model of transcription

Recent evidence suggests that transcription dynamics in *E. coli* at optimal temperatures (37 °C) from the lac-ara-1 promoter, as well as in a few other promoters, is well modeled by a multi-step sequential process with two rate-limiting elementary steps,[6,15] at least when these are inserted on single-copy plasmids. Namely no significant 'bursts' in transcription were reported.[5,6,15,23,24] This may be because, under these conditions, the recently reported phenomenon of buildup of positive supercoiling with transcription events, which may lead to short-length transcriptional bursts,[27] is too weak due to the 'lack of topological barriers' on the plasmid.[27] Further, this phenomenon is expected to affect tangibly only highly expressed genes, while in our measurements we recorded mean intervals between transcription events longer than 1000 s under full induction.

As such, and following the modeling strategy used in ref. 6, 15, 19 and 28, we model transcription as a non-bursty, two rate-limiting step process, with the following set of reactions (1)–(3):

$$R + \mathrm{Pr} \overset{K_R}{\longleftrightarrow} \mathrm{Pr}\,R \tag{1}$$

$$\mathrm{Pr} \xrightarrow{k_c} \mathrm{Pr_c} \tag{2}$$

$$\mathrm{Pr_c} \xrightarrow{k_o} \mathrm{Pr_o} \xrightarrow{\infty} \mathrm{Pr} + \mathrm{M} \tag{3}$$

Reaction (1) models the fast binding/unbinding with a dissociation constant, K_R, of a single *lacI* tetramer, denoted as *R*, to $P_{lac\text{-}ara\text{-}1}$,[20,29] which is denoted as Pr when free for transcription and as Pr*R* when in the repressed state, *i.e.* bound by a repressor.

Transcription initiation is modeled as a two-rate-limiting-step process[14] by reaction (2), which models the formation of the closed complex (Pr_c) at a rate k_c, and by reaction (3) which models the formation of an open complex (Pr_o), from the closed complex, at a rate k_o, followed by the promoter escape (assumed to be infinitely fast following the completion of the open complex[30]). Since the duration of both the promoter escape and of transcription elongation is negligible when compared to transcription initiation,[31] we assume that a complete RNA molecule (M) is released 'immediately' after the formation of an open complex.

From (1), the probability that the promoter will be in the unrepressed state equals:

$$f_R(t) = \frac{K_R}{R(t) + K_R} \tag{4}$$

where $f_R(t)$ takes values from 0 to 1, representing the activity level of the promoter at time *t*.

Assuming that the binding/unbinding of repressors is a much faster process than the closed complex formation, reactions (1) and (2) can be combined as follows:[32]

$$\mathrm{Pr} \xrightarrow{k_c \times f_R(t)} \mathrm{Pr_c} \tag{5}$$

where the regulation function $f_R(t)$ is a hill function with coefficient 1 and parameter K_R. This assumption is supported by recent *in vivo* measurements of the binding/unbinding rates of *lacI* from its operator sites at the Lac promoter (mean binding time to the DNA of 59 s and mean residence time on the DNA of the order of milliseconds[31]) along with estimations of the duration of the closed complex formation of $P_{lac\text{-}ara\text{-}1}$ *in vivo* (mean higher than 300 s[15]).

Given this model, the mean RNA production interval, following a transient induction time, is given as:

$$\overline{\Delta t} = \frac{1}{f_R(\infty) k_c} + \frac{1}{k_o} \tag{6}$$

Note that this model of transcript production dynamics assumes that the promoter copy number equals 1, since the plasmid coding for the RNA target for MS2d-GFP is a single-copy plasmid (see ''Bacterial strain and plasmids'' section).

Model of inducer number dynamics

E. coli being Gram-negative, the membrane has two layers: the outer membrane and the inner membrane, with the periplasmic space in between.[33] Given high IPTG extracellular abundance, in the absence of feedback mechanisms,[11,20] the inducer levels in the periplasm (I_m) and cytoplasm (I) can be accurately estimated[34] from:

$$\frac{\delta I_m}{\delta t} = k_{outer} - I_m \times k_{inner} \tag{7}$$

$$\frac{\delta I}{\delta t} = I_m \times k_{inner} - I \times d_I \tag{8}$$

Eqn (7) and (8) describe the irreversible intake of inducers from the media into the periplasm at the rate k_{outer} and the subsequent transport of inducers from the periplasm to the cytoplasm at a rate k_{inner}. k_{outer} varies with extracellular inducer concentration but, for each measurement condition, it remains constant during the measurement period, due to the absence of *lacY* permease.

Meanwhile, d_I is the decay rate of intracellular IPTG. Note that, since IPTG is not hydrolysable by the cells and is inefficiently

transported out due to the weak expression of sugar efflux transporters,[35] its concentration in the cytoplasm is expected to dilute mostly through cell growth,[36] rather than consumption or efflux. As such, d_I is estimated from the cell growth rate alone (ESI†). Given this, by solving (7) and (8) (ESI†), one finds the inducer level in the cytoplasm over time to be:

$$\begin{aligned} I(t) &= \frac{k_{\text{outer}}\left(d_I e^{-k_{\text{inner}}t} - k_{\text{inner}} e^{-d_I t} + k_{\text{inner}} - d_I\right)}{d_I(k_{\text{inner}} - d_I)} \\ &= \frac{k_{\text{outer}}}{d_I} \times S(k_{\text{inner}}, t) \end{aligned} \tag{9}$$

where S, the normalized function of the inducer level ranging from 0 to 1, describes the shape of $I(t)$:

$$S(k_{\text{inner}}, t) = \frac{d_I e^{-k_{\text{inner}}t} - k_{\text{inner}} e^{-d_I t} + k_{\text{inner}} - d_I}{k_{\text{inner}} - d_I} \tag{10}$$

The inducer level at equilibrium ($t \to \infty$) is therefore given by:

$$I(\infty) = \frac{k_{\text{outer}}}{d_I} \times S(k_{\text{inner}}, \infty) = \frac{k_{\text{outer}}}{d_I} \tag{11}$$

From eqn (11), the transport rate at the inner membrane, k_{inner}, does not affect intracellular inducer levels and, consequently, the induction strength at equilibrium. However, a finite k_{inner}'s value results in a delay in the entrance of inducers into the cells, which increases the "waiting time" for the synthesis of the first RNA, t_0, following the introduction of inducers into the media.

Model of inducer repressor interactions

IPTG is an indirect activator, as it binds to *lacI* tetramers reducing greatly their binding affinity to the promoter.[37] Reaction (12) describes the fast binding/unbinding between inducers and repressors with the dissociation constant K_I:

$$I + R \overset{K_I}{\longleftrightarrow} RI \tag{12}$$

As the number of intracellular inducers (even under weak induction) is much greater than that of repressors,[12,38] the number of free inducers at any given time can be approximated by the total amount of inducers in the cells, $I(t)$. Due to the high rate of the forward reaction and inducer abundance, when the intracellular inducer concentration changes, we assume that reaction (12) reaches equilibrium before any binding event between R and Pr (1) can occur. The amount of repressors at any given time is therefore expected to be:

$$\begin{aligned} R(t) &= K_I \times RI(t)/I(t) = K_I \times (R_{\max} - R(t))/I(t) \\ &= \frac{K_I R_{\max}}{K_I + I(t)} \end{aligned} \tag{13}$$

Here R, I, Pr and PrR are considered 'fast species', due to their fast rates of interaction. Thus, their impact on the dynamics of the slow species, Pr_c, and consequently M, is determined solely by their mean level.[32] The inducible promoter's activity level over time is thus:

$$f_R(t) = \frac{K_R}{R(t) + K_R} = \frac{(K_I + I(t))K_R}{K_I R_{\max} + (K_I + I(t))K_R} \tag{14}$$

We define R_K and I_K as the relative level of repressors and inducers (both bound and unbound) at equilibrium, respectively, as follows:

$$R_K = \frac{R_{\max}}{K_R} \tag{15}$$

$$I_K = \frac{I(\infty)}{K_I R_K} = \frac{k_{\text{outer}}}{d_I K_I R_K} \tag{16}$$

Combining eqn (9), (14), (15) and (16), we obtain:

$$f_R(t) = \frac{1 + I_K R_K S(k_{\text{inner}}, t)}{R_K + 1 + I_K R_K S(k_{\text{inner}}, t)} \tag{17}$$

In (15), R_K is the ratio between the total number of repressors and the amount required to repress the promoter's activity to half in the absence of inducers. For the strain studied (DH5α-PRO), R_K is much greater than 1.[20] Meanwhile, I_K is the ratio between the total number of intracellular inducers at equilibrium ($I(\infty)$) and the amount required to induce the promoter's activity to half. With d_I, K_I, R_K being invariant to extracellular inducer concentrations, I_K is determined only by k_{outer}. Both prior to induction and when steadily induced, the promoter's activity is therefore given by:

$$f_R(0) = \frac{1}{R_K + 1} \tag{18}$$

$$f_R(\infty) = \frac{1 + I_K R_K}{R_K + 1 + I_K R_K} \sim \frac{I_K}{1 + I_K} \tag{19}$$

From (18) and (19), we can learn about the leakiness in the expression system and the mean RNA synthesis rate ($1/\Delta t$) for a given level of induction.

Model distribution of t_0

From the hybrid model of deterministic inducer and repressor dynamics coupled with stochastic transcription dynamics, we use the chemical master equation (CME)[34] to calculate the first moment of an open complex formation completion in each cell, which is immediately followed by the release of a transcript[30,31] (ESI†).

We have also simulated an all-stochastic model of inducer and repressor dynamics (with the extracellular inducer number (I_m) at 1 mM and the repressor number ($R_{\max}$) set arbitrarily high), coupled with stochastic transcription using the stochastic simulation algorithm.[39] We did not observe any statistical difference between the sample distributions acquired from the simulations (with 1000 samples of t_0) and the distributions calculated using the CME, assuming the hybrid model, indicating that the intrinsic noise in the dynamics of inducers and repressors does not affect the expression dynamics of the target gene.

By finding the time evolution of the promoter state described by the CME for each pair of values of R_K and I_K, one can calculate the time distribution of the appearance of the first RNA in each cell, following the formation of the open complex.

3. Results and discussion

As described in ref. 5, the time for the appearance of the first RNA in a cell, following the introduction of inducers in the media, includes not only the ''intake time'' (time for the inducer to enter the cell), but also the time for a promoter to produce a single RNA. The latter can be extracted from the distribution of intervals between consecutive RNA productions by an active promoter.[5]

3.1 Kinetics of transcription of the active $P_{lac\text{-}ara\text{-}1}$

To characterize the kinetics of transcription initiation of the single-copy promoter $P_{lac\text{-}ara\text{-}1}$ under full induction, we observed for 1 hour 1463 cells induced by 1 mM IPTG and 1052 cells induced by 1.25 mM IPTG in M63 media. We performed a KS test comparing the two distributions of intervals between RNA productions and found no significant difference between them (p-value larger than 0.01). As such, we consider these two sets of cells to be equally fully induced and merged the two sets of data. The resulting distribution of actual intervals is shown in Fig. 3.

To the merged collection of the 759 actual samples and 1083 censored samples (Methods) of intervals extracted from both sets, we fitted a 2-step model of transcription initiation (Methods). The pair of steps $\sim$1751 s and $\sim$337 s long was the best fit (p-value $\sim$1 from Pearson's chi-squared test (dashed line in Fig. 3)). The margin of error of the inferred value for each step was $\sim$15%, with a confidence of 90%. Notably, the inferred mean of Δt ($\sim$2088 s, from actual and censored samples) agrees with reports of the *in vivo* RNA synthesis rate of this promoter under full induction in LB media,[15] confirming the reaching of a fully induced state in M63 media.

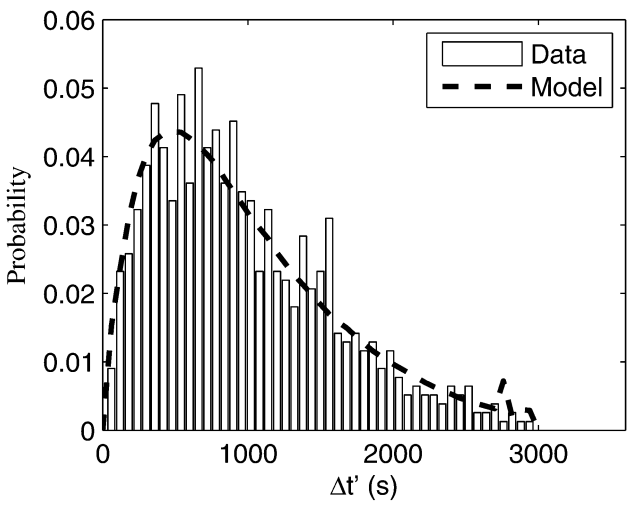


Fig. 3 Distribution of the actual samples of RNA production intervals $\Delta t'$ (bars) and distribution curve (dashed line) inferred from the transcription model of two sequential steps. The samples are obtained from cells subject to 1 mM and 1.25 mM of IPTG in the media.

Finally, based on the conclusions from previous studies,[4,15] here onwards we assume that the longer of the two steps inferred is the closed complex formation, while the shorter is the open complex formation. Accordingly, we set in the model the rates of formation of these complexes in reactions (2) and (3), respectively, as: $k_c \sim 1/1751\ s^{-1}$, $k_o \sim 1/337\ s^{-1}$.

3.2 Time of appearance of the first RNA

To introduce empirical information regarding the intake mechanism into the model, we use the time for the appearance of the first RNA in each cell for different inducer concentrations (denoted $[IPTG]_{media}$), following the introduction of the inducer in the media. In particular, these concentrations were: 0.25 mM, 0.5 mM, 0.75 mM, 1 mM and 1.25 mM. Note that cells grew exponentially[40] during the measurements at a rate of d_I of $\sim 8.25 \times 10^{-5}\ s^{-1}$ (corresponding to a doubling time of $\sim$140 minutes) under all conditions (ESI†), thus it is reasonable to assume that the analyzed cells' physiology is unaffected by the inducer levels in the range tested.

From the time lapse microscopic images, we recorded when the first RNA appeared in each cell following induction. The data are shown in Table 1, for each condition. As expected, the mean t_0' ($\mu_{t_0'}$) decreases with increasing $[IPTG]_{media}$. We performed KS tests of comparison between the empirical distributions in each condition. The resulting p-values are smaller than 10^{-4}, indicating that these differ in a statistical sense.

We also performed measurements at higher IPTG concentrations (2 and 4 mM), but the cells exhibited numerous inclusion bodies (ESI†), likely due to an increase in the rate of protein misfolding.[41] As these may introduce pleiotropic effects,[42] the data were not used.

3.3 Inference of the intracellular relative levels of repressors and inducers and of the intake rates

Given the rates of dilution, d_I, and closed and open complex formations (k_c and k_o, respectively) derived in the previous sections, the model distribution of t_0 can be fully characterized by the intracellular relative numbers of repressors (R_K) and inducers (I_K) along with the transport rate of the inner membrane (k_{inner}).

Table 1 Measurements of t_0 for different IPTG concentrations ($[IPTG]_{media}$). For each condition, the table shows the number of actual samples (t_0') and censored samples (c_0') collected, along with the mean ($\mu_{t_0'}$), standard deviation ($\sigma_{t_0'}$) and the normalized variance $\left(\mu_{t_0'}^2/\sigma_{t_0'}^2\right)$ calculated from the actual samples

$[IPTG]_{media}$ (mM)	No. of t_0'	No. of c_0'	$\mu_{t_0'}$ (s)	$\sigma_{t_0'}$ (s)	$\mu_{t_0'}^2/\sigma_{t_0'}^2$
0.25	114	60	4056	1703	0.18
0.50	210	128	3713	1599	0.19
0.75	120	129	3054	1413	0.21
1.00	199	105	3248	1550	0.23
1.25	80	38	3253	1311	0.16

R_K and k_{inner} are defined by the cell strain and thus are invariant between conditions. Meanwhile, I_K is determined by the inducer intake rate k_{outer} at the outer membrane, determined by the external inducer concentration. Using the maximum likelihood method, we fitted the model of intake to the data (including both actual and censored samples). The parameters to infer are R_K, k_{inner} and I_K for each condition. The inferred values are shown in Table 2.

Note that, for all conditions, the inferred values of I_K are significantly greater than 1 (Table 2). Using eqn (19), we find that the promoter activity $f_R(\infty)$ is close to full induction under all conditions studied, from 84% (at 0.25 mM) to 93% (at 1.25 mM).

Using these inferred values of R_K, k_{inner} and I_K for each condition, we estimated the distribution of t_0 subject to censoring (ESI†). We plotted these in Fig. 4. Also shown are the distributions of the actual samples, t_0'.

Transport rate of IPTG through the inner membrane

We inferred the transport rate of IPTG through the inner membrane, k_{inner}, to be 5.3×10^{-4} s^{-1} (with 90% confidence, k_{inner} is between 1.6×10^{-4} s^{-1} to 8.9×10^{-4} s^{-1}). Thus, each inducer takes on average 31.7 min (with 90% confidence, between 18 min and 104 min) to travel from the periplasm to the cytoplasm. These numbers show that this event is time-consuming, in that it affects RNA numbers at early stages of induction.[5,43] As a side note, the stochasticity of these events is also visible from the data. *E.g.*, we observed RNAs first appearing as early as 5 min and as long as 120 min, after the introduction of inducers (note that the upper bound is also affected by variability in transcription time length).[6]

The relative repressor level, R_K

From the inferred repressor level ($R_K \sim 42$ and, with 90% confidence between 21 and 120), we expect the promoter activity to change by $\sim$43 fold between no induction and full induction (12). This is in the same order of magnitude as the data from *in vitro* measurements on $P_{lac\text{-}ara\text{-}1}$'s range of activity in DH5α-PRO ($\sim$100 fold[20]). Meanwhile, the leaky expression of the target gene (prior to induction) can be estimated to be $\sim$0.03 RNA per h per cell (using (12)), in agreement with the measured leakiness ($<$0.1 RNA per h per cell).

Table 2 Results of fitting the model of intake with t_0 measurements in five conditions. The table shows the estimation of R_K and k_{inner} for the strain used, and then the inferred value of I_K per condition. Also shown is the *p*-value of the Pearson's chi-squared test for estimation of the goodness of fit. We assume that for *p*-values greater than 0.01, the distributions cannot be distinguished

Variables	Inferred value	*p* value
k_{inner}	5.3×10^{-4} s^{-1}	
R_K	42	
I_K (0.25 mM)	5.37	0.07
I_K (0.50 mM)	5.41	0.77
I_K (0.75 mM)	10.93	0.79
I_K (1.00 mM)	9.94	0.19
I_K (1.25 mM)	14.34	0.25

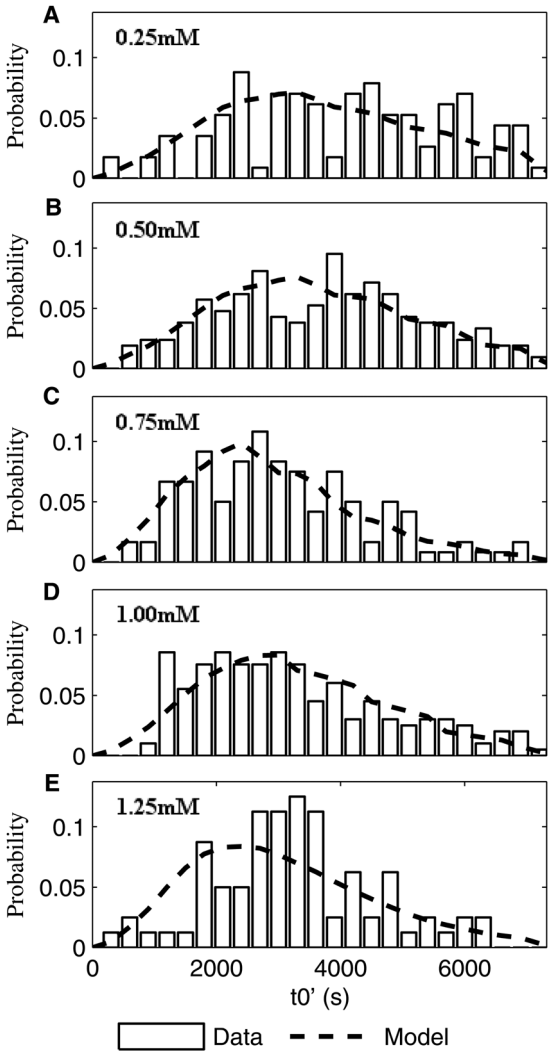


Fig. 4 Distribution of the actual samples t_0' (bars) and distribution curves inferred from the model (dashed lines). Data for (A) 0.25 mM, (B) 0.50 mM, (C) 0.75 mM IPTG, (D) 1.0 mM and (E) 1.25 mM IPTG concentrations.

Finally, we also derived alternative models of inducer-repressor interactions (reaction (12)), where more than one (namely two, three and four) inducer molecules are required to neutralize one repressor molecule.[37] In all models tested, the likelihood ratio test comparing the original model with the alternative ones yielded *p*-values smaller than 0.01, favoring the model of the first-order inducer–repressor interaction. For example, the fourth-order model, which assumes that the tetramer *lacI* requires exactly four inducer molecules to lose its binding affinity to the promoter, was rejected by the Pearson chi-squared test (*p*-value smaller than 0.01).

Intake mechanisms of IPTG at the outer membrane

We next studied the nature of the dominant intake mechanism of IPTG (*i.e.* whether it has feedback or is diffusive-like). To assess whether the intake of IPTG through the outer membrane is consistent with a process of pure diffusion (*i.e.* I_K and k_{outer} proportional to $[IPTG]_{media}$), we compared the values of I_K as a

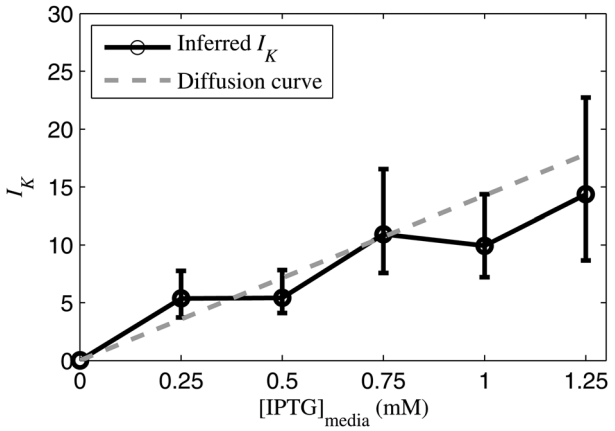


Fig. 5 Intracellular inducer level, I_K, as a function of external inducer concentration $[IPTG]_{media}$. The vertical bar indicates the margins of error, for α-value of 90%. The diffusion curve (grey dashed) is the approximation of I_K using a line through the origin (0, 0). The slope of the diffusion curve is set as the mean of $I_K/[IPTG]_{media}$ under all conditions.

function $[IPTG]_{media}$ (as inferred in the previous section from empirical data) to the values of I_K that would be expected from a purely diffusive mechanism, with a diffusion rate constant equal to $k_{outer}/[IPTG]_{media}$, averaged over all conditions (eqn (16)).

In Fig. 5, we plotted the inferred values of I_K along with the 90% margins of error from the empirical data. Note that, as the inducer concentration increases and, thus, the promoter reaches close-to-full induction faster, the margins of error for the inference of I_K also increase. Also plotted are the values of I_K expected from a pure diffusive model. We assume that, in the absence of IPTG in the media, the inducers are absent from the cells (origin of the plot). From 0 mM up to 1.25 mM, the diffusion curve (grey dashed line) is not excluded from the confidence interval of I_K and therefore the inferred model and the pure diffusion model are indistinguishable.

4. Conclusions

From the transient dynamics of transcription initiation upon the introduction of the inducer at different concentrations into the media, we characterized the mechanism of intake of IPTG, a synthetic inducer of $P_{lac\text{-}ara\text{-}1}$. We made use of *in vivo* measurements of the moments of occurrence of individual transcription events in multiple cells. Namely, we measured the intervals between consecutive RNA production events and the waiting time, t_0, for the first target RNA to appear in each cell following induction. Then, we fitted a deterministic model of inducer intake through a bilayer membrane, coupled with a stochastic multi-step model of transcription, and we studied how the kinetics of intake changes as a function of extracellular inducer concentrations.

We found that a model of a bilayer membrane fits the data well, for a transport rate of inducers through the second membrane layer as slow as $\sim 5.3 \times 10^{-4}$ s^{-1}. This suggests that the entrance of inducers into the cytoplasm, after crossing the outer membrane, is a time-consuming event that causes tangible effects on single-cell RNA numbers following the introduction of an inducer into the media, in agreement with ref. 43. A recent study on the *in vivo* intake kinetics of the MG^+ ion, whose mass is similar to that of IPTG (~300 Da), reported similar transport times (~15 to 75 min).[44] This is surprising, given their different hydrophobicity[18] and suggests that this property might not always be the main factor determining intake times.

Finally, we found that, at high concentrations, the intake at the outer membrane can be well approximated by a model of diffusion, where the intake rate is linearly dependent on the external inducer concentration. This suggests that inducers can cross the outer membrane with a dynamics similar to that of a Michaelis–Menten process, when the amount of pores contributing to the intake process is a small portion of the total amount of pores capable of intake (*i.e.* for amounts of IPTG that do not saturate the pores). In support of this hypothesis, at 2 mM or higher IPTG concentrations in the media, there are observable changes in cells, namely the formation of inclusion bodies.

Our results, aside from the empirical ones, are drawn from deterministic models of inducer intake and repressor dynamics, combined with a stochastic model of transcription. As such, the cell-to-cell diversity generated by the model (*e.g.* in the values of t_0) is only due to noise in transcription. This approximation was made based on the intake of inducers and the interactions between inducers and repressors having much shorter time scales (of the order of tens of seconds[29,45]) than the closed and open complex formations (of the order of hundreds of seconds[4,15,28]).

Intake processes can nevertheless generate tangible, transient phenotypic diversity (see, *e.g.*, ref. 5, 11 and 46), for example, due to the cell-to-cell diversity in membrane properties (such as the number of pores and permease proteins responsible for the IPTG intake) or in intracellular numbers of repressors, among others. Here, to minimize the role of such factors, we employed the strain DH5-αPRO, whose lac repressor is overexpressed. In the future, it would be of interest to investigate the contribution of noise in the intake process to the diversity in cellular responses to, *e.g.*, environmental shifts.

Even though DH5-αPRO cells cannot produce *lacY* permease[20] and, thus, cannot regulate the intake kinetics of inducers as a function of extracellular inducer concentration (shown by the linear increase in I_k as a function of IPTG concentration), note that our results are also applicable to *E. coli* strains able to produce *lacY* permease, as they apply to the regime of high IPTG concentrations, where *lacY*'s contribution to the total influx of inducer is negligible.[8,9]

In this regard, in general, the method employed here can be used to study the intake mechanisms of other inducers, by altering the target promoter and removing cellular disposal systems of the inducer (*e.g.*, the araBAD operon which catalyzes the arabinose metabolism[47] or the tetA gene responsible for aTc efflux[48]), so as to eliminate negative feedbacks controlled by the target gene.[38]

In general, the findings on the kinetics of the intake system of an inducer can be used to establish a lower bound for the response time of genetic systems to external stimuli. As such, knowledge of this process aids in understanding how cells

constantly adapt to fluctuating environments. This knowledge will also be of use in the construction of synthetic circuits. For example, when designing circuits capable of decision making or filtering based on environmental conditions (*e.g.* switches[49] or frequency filters[50]), intake times will influence the rate of decision making or the filter response. Added to that, knowledge of the intracellular inducer level as a function of the media composition aids in understanding different modes of activity of genetic circuits and, as such, we may be able to expand the ranges of applicability of the synthetic circuits. For example, using promoters of the same family with different inducer affinities (*e.g.* P_{lac} and $P_{lac\text{-}ara\text{-}1}$,[20] or P_{BAD}, P_E and P_{GFH}[47]), one should be able to construct synthetic genetic circuits exhibiting different behaviors that will be selectable by the inducer concentration in the media.

Acknowledgements

The authors thank Antti Häkkinen for valuable advice and discussions. Work supported by Academy of Finland (257603, ASR) and Portuguese Foundation for Science and Technology (PTDC/BBB-MET/1084/2012, ASR). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1 W. Gilbert and B. Muller-Hill, *Proc. Natl. Acad. Sci. U. S. A.*, 1966, **56**, 1891–1898.
2 G. C. Miyada, L. Stoltzfus and G. Wilcox, *Proc. Natl. Acad. Sci. U. S. A.*, 1984, **81**, 4120–4124.
3 M. Gossen and H. Bujard, *Proc. Natl. Acad. Sci. U. S. A.*, 1992, **89**, 5547–5551.
4 R. Lutz, T. Lozinski, T. Ellinger and H. Bujard, *Nucleic Acids Res.*, 2001, **29**, 3873–3881.
5 J. Mäkelä, M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2013, **41**, 6544–6552.
6 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.
7 F. Jacob and J. Monod, *J. Mol. Biol.*, 1961, **3**, 318–356.
8 L. H. Hansen, S. Knudsen and S. J. Sørensen, *Curr. Microbiol.*, 1998, **36**, 341–347.
9 P. R. Jensen, H. V Westerhoff and O. Michelsen, *Eur. J. Biochem.*, 1993, **211**, 181–191.
10 A. Marbach and K. Bettenbrock, *J. Biotechnol.*, 2012, **157**, 82–88.
11 E. M. Ozbudak, M. Thattai, H. N. Lim, B. I. Shraiman and A. Van Oudenaarden, *Nature*, 2004, **427**, 737–740.
12 A. Fernández-Castané, G. Caminal and J. López-Santín, *Microb. Cell Fact.*, 2012, **11**, 58.
13 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–1036.
14 M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja and A. S. Ribeiro, *BMC Syst. Biol.*, 2011, **5**, 149.
15 M. Kandhavelu, J. Lloyd-Price, A. Gupta, A. B. Muthukrishnan, O. Yli-Harja and A. S. Ribeiro, *FEBS Lett.*, 2012, **586**, 3870–3875.
16 P. J. Schlax, M. W. Capp and M. T. J. Record, *J. Mol. Biol.*, 1995, **245**, 331–350.
17 I. Golding and E. C. Cox, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11310–11315.
18 M. Vaara, W. Z. Plachy and H. Nikaido, *Biochim. Biophys. Acta, Biomembr.*, 1990, **1024**, 152–158.
19 A. S. Ribeiro, R. Zhu and S. A. Kauffman, *J. Comput. Biol.*, 2006, **13**, 1630–1639.
20 R. Lutz and H. Bujard, *Nucleic Acids Res.*, 1997, **25**, 1203–1210.
21 S. Chowdhury, M. Kandhavelu, O. Yli-Harja and A. S. Ribeiro, *J. Microsc.*, 2012, **245**, 265–275.
22 A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J. M. Fonseca and A. S. Ribeiro, *Bioinformatics*, 2013, **29**, 1708–1709.
23 M. Kandhavelu, A. Häkkinen, O. Yli-Harja and A. S. Ribeiro, *Phys. Biol.*, 2012, **9**, 026004.
24 A. B. Muthukrishnan, A. Martikainen, R. Neeli-Venkata and A. S. Ribeiro, *PLoS One*, 2014, **9**, e109005.
25 D. R. Cox, *J. R. Stat. Soc., Ser. B*, 1972, **34**, 187–220.
26 H. Koul, V. Susarla and J. Van Ryzin, *Ann. Stat.*, 1981, **9**, 1276–1288.
27 S. Chong, C. Chen, H. Ge and X. S. Xie, *Cell*, 2014, **158**, 314–326.
28 H. Buc and W. R. McClure, *Biochemistry*, 1985, **24**, 2712–2723.
29 J. Elf, G.-W. Li and X. S. Xie, *Science*, 2007, **316**, 1191–1194.
30 L. M. Hsu, *Biochim. Biophys. Acta*, 2002, **1577**, 191–207.
31 K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick and S. M. Block, *Cell*, 2006, **125**, 1083–1094.
32 Y. Cao, D. T. Gillespie and L. R. Petzold, *J. Chem. Phys.*, 2005, **122**, 14116.
33 C. W. Forsberg, J. W. Costerton and R. A. Macleod, *J. Bacteriol.*, 1970, **104**, 1338–1353.
34 D. T. Gillespie, *J. Stat. Phys.*, 1977, **16**, 311–318.
35 J. Y. Liu, P. F. Miller, J. Willard and E. R. Olson, *J. Biol. Chem.*, 1999, **274**, 22977–22984.
36 N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain and M. B. Elowitz, *Science*, 2005, **307**, 1962–1965.
37 M. Lewis, *C. R. Biol.*, 2005, **328**, 521–548.
38 J. A. Megerle, G. Fritz, U. Gerland, K. Jung and J. O. Rädler, *Biophys. J.*, 2008, **95**, 2103–2115.
39 D. T. Gillespie, *J. Comput. Phys.*, 1976, **22**, 403–434.
40 S. Cooper, *Theor. Biol. Med. Modell.*, 2006, **3**, 10.
41 N. Sriubolmas, W. Panbangred, S. Sriurairatana and V. Meevootisom, *Appl. Microbiol. Biotechnol.*, 1997, **47**, 373–378.
42 A. B. Lindner, R. Madden, A. Demarez, E. J. Stewart and F. Taddei, *Proc. Natl. Acad. Sci. U. S. A.*, 2008, **105**, 3076–3081.
43 J. A. Boezi and D. B. Cowie, *Biophys. J.*, 1961, **1**, 639–647.
44 J. Zeng, H. M. Eckenrode, S. M. Dounce and H.-L. Dai, *Biophys. J.*, 2013, **104**, 139–145.

45 P. Hammar, P. Leroy, a. Mahmutovic, E. G. Marklund, O. G. Berg and J. Elf, *Science*, 2012, **336**, 1595–1598.
46 G. Fritz, J. A. Megerle, S. A. Westermayer, D. Brick, R. Heermann, K. Jung, J. O. Rädler and U. Gerland, *PLoS One*, 2014, **9**, e89532.
47 C. M. Johnson and R. F. Schleif, *J. Bacteriol.*, 1995, **177**, 3438–3442.
48 P. McNicholas, I. Chopra and D. M. Rothstein, *J. Bacteriol.*, 1992, **174**, 7926–7933.
49 T. S. Gardner, C. R. Cantor and J. J. Collins, *Nature*, 2000, **403**, 339–342.
50 A. Häkkinen, H. Tran, O. Yli-Harja and A. S. Ribeiro, *PLoS One*, 2013, **8**, e70439.

Supplementary Material for "Kinetics of the cellular intake of a gene expression inducer at high concentrations"

Huy Tran[1], Samuel M.D. Oliveira[1], Nadia Goncalves[1] and Andre S. Ribeiro[1,*]

[1]Laboratory of Biosystem Dynamics, Department of Signal Processing, Tampere University of Technology.
TC316, Korkeakoulunkatu 10, 33720 Tampere, Finland
*andre.ribeiro@tut.fi

I. Measurements and data extraction

Media and growth condition

Cells were grown overnight at 30°C with aeration and shaking in Luria-Bertani (LB) medium, supplemented with the necessary antibiotics. Cells were then diluted in fresh M63 medium. When reaching an optical density of OD600≈0.3–0.5, cells were pre-incubated for 45 min with 100 ng/ml anhydrotetracycline (aTc) to produce enough matured MS2d-GFP proteins to detect RNAs at the start of the microscopy measurements. During the microscopy measurements, cells were kept in M63 media, so as to extend cells' division time, which increases the chances for each cell present at the start of the measurements to produce at least one target RNA before it divides. The contents of (i) LB and (ii) M63 media are:

(i) 10g/L of Tryptone (Sigma Aldrich, USA), 5g/L of yeast extract (LabM, UK) and 10g/L of NaCl (LabM, UK);

(ii) 2mM MgSO4.7H2O (Sigma-Aldrich, USA), 7.6mM (NH4)2SO4 (Sigma Life Science, USA), 30μM FeSO4.7H2O (Sigma Life Science, USA), 1mM EDTA (Sigma Life Science, USA), 60mM KH2PO4 (Sigma Life Science, USA) pH 6.8 with Glycerol 0.5% (Sigma Life Science, USA) and Casaminoacids 0.1% (Fluka Analytical, USA).

Microscopy

After pre-incubation with aTc, cells are placed on a microscope slide with 3% agarose gel to restrict movements. A peristaltic pump is used to provide cells with a constant flow of fresh, pre-warmed M63 media and of IPTG at specified concentrations throughout the measurement period. With the pump initialized at a speed of 0.3 mL/min, the collection of time lapse images by confocal microscopy is initiated as soon as the flow reaches the microscope slide (detected visually).

Microscopy time series were 3 hours long, with cells being imaged once per minute. The data from the first ~5 minutes following induction is not recorded (although time is) as the gel slide slightly shifts due to the initialization of flow of fresh media by the pump, hampering a proper cell tracking.

During the microscopy measurements, the cells' fluorescent background was found to be stable, which indicates that the ability of target RNA counting of the MS2d-GFP system does not change during the course of measurements. Also, from previous studies[1–4], the amounts of fluorescence in the cell background observed suffice to accurately report the appearance of new target RNA molecules in the cells.

Image and data analysis

Image analysis was performed as in [1]. We use a semi-automated cell segmentation strategy [5] as in [6]. Afterwards, fluorescent spots in each cell at each time moment are detected automatically (Figure S1) as in [7], by estimating the cell background intensity distribution using its median and median absolute deviation, and then performing thresholding with a given confidence level assuming that this distribution is Gaussian. Finally, we extracted the

moment when the first RNA appears in each cell and the time intervals between consecutive RNA production events are extracted from the time series of total spot intensities.

We fit a monotonically increasing piecewise-constant function to the corrected total spot intensity in a cell over time using least squares and infer on the moments of appearance of novel target RNAs as in [2,3,8]. The number of terms for the fit was selected by an F-test with a p-value of 0.01. Each discontinuity, i.e. jump, corresponds to the production of one target RNA[3]. An example of the results of applying these methods is shown in Figure S1. Validation of this method is provided in section VII of this document.

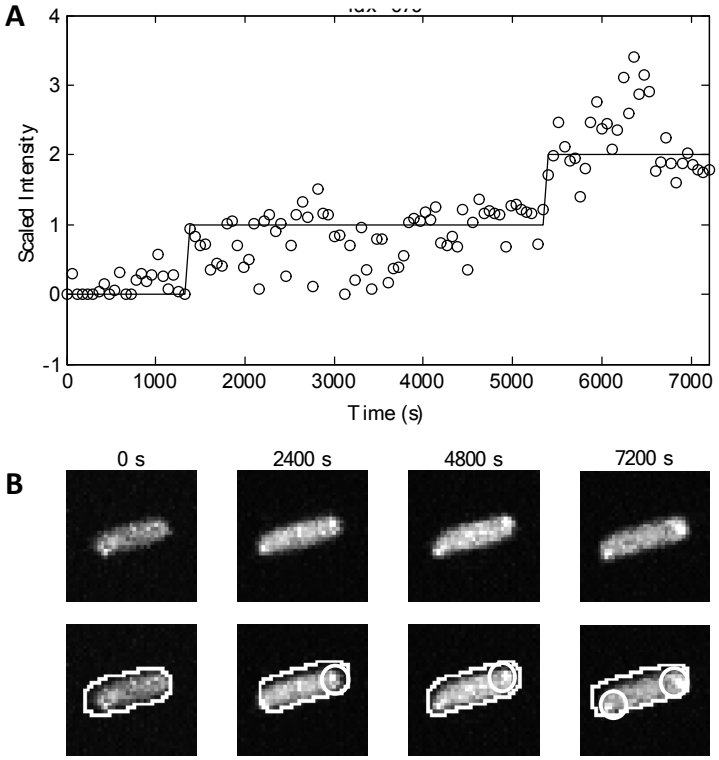


Figure S1. Tagged RNAs in *E. coli* cells. (A) Unprocessed frames and segmented cells and RNA spots. The moments when images were taken are shown for each frame. (B) Examples of time series of scaled spot intensity levels from one cell (circles) and the corresponding estimated RNA numbers (solid lines).

II. Collection and analysis of censored data

The problem of right censored data is well described in [9,10], where each individual in the population has a limited life time drawn from a random variable *Y*. We measure from each individual of the population the time for a certain event *X* to occur. We assume that the time for this event to occur, without the effect of limited life time, is a random variable *T*. Given that *X* has no effect on the 'health' of the individuals under observation, *T* and *Y* are independent of one another.

Collection of censored data

For the i^{th} individual in the population, we draw from the bivariate variable $<T, Y>$ a pair $<t_i, y_i>$, where y_i is the life time of that individual and t_i is the time for event *X* to occur. We define δ_i and z_i as follow:

$$\delta_i = [t_i < y_i] \text{ and } z_i = min(t_i, y_i) \quad (1)$$

where δ_i is the type of sample and z_i is the value of the sample of the i^{th} individual. If the event occurs before the death of the individual, we obtain an actual sample (δ_i=1), else we obtain a censored sample (δ_i=0).

Measurements of the time for the first RNA to appear in each cell, t_0, are obtained from cells present at the start of the microscopy sessions. For measurements of Δt, the intervals between consecutive RNA productions in each cell, the individuals under the observation are any cells that produce one or more RNAs during the last hour of the measurements. For both measurements of Δt and of t_0, the event X to observe is the appearance of the next novel RNA molecule in the cell. Cell 'death' is due to division or the end of the measurement time.

Likelihood function of censored data

To find the likelihood function of the parameter set θ characterizing the model of T, we calculate the possibility to obtain the outcome $<\delta_{1..n}, z_{1..n}>$ from n individuals in the population with this model: $\Lambda\left(\langle\delta_{1..n},z_{1..n}\rangle \,\middle|\, \theta\right)$. With each parameter set θ, the model of T is defined by the probability distribution function $P_{T|\theta}(t|\theta)$ and the cumulative distribution function $F_{T|\theta}(t|\theta)$.

The life time Y of individuals in the population has the probability distribution function $P_Y(y)$ and the cumulative distribution $F_Y(y)$. These distribution functions can be obtained directly by measuring the life time of the individuals in the population.

The likelihood function of the parameter set θ of T's model with the outcome $<\delta_{1..n}, z_{1..n}>$ is given by[9]:

$$\Lambda\left(\langle\delta_{1..n},z_{1..n}\rangle \,\middle|\, \theta\right) = \prod_{i=1}^{n}\left[P_{T|\theta}(z_i|\theta)\left(1-F_Y(z_i)\right)\right]^{\delta_i}\left[P_Y(z_i)\left(1-F_{T|\theta}(z_i|\theta)\right)\right]^{1-\delta_i} \tag{2}$$

Here, $P_{T|\theta}(z_i|\theta)\left(1-F_Y(z_i)\right)$ is the probability of obtaining an actual sample with the value z_i ($<\delta_i$=1, $z_i>$), and $P_Y(z_i)\left(1-F_{T|\theta}(z_i|\theta)\right)$ is the probability of obtaining a censored sample with the value z_i ($<\delta_i$=0, $z_i>$).

While probing for the value of θ that maximizes the likelihood function, the functions $P_Y(y)$ and $F_Y(y)$, which are independent of T, remain constant. Therefore, the objective function to maximize can be simplified to:

$$Obj\left(\theta \,\middle|\, \langle\delta_{1..n},z_{1..n}\rangle\right) = \prod_{i=1}^{n}\left[P_{T|\theta}(z_i|\theta)\right]^{\delta_i}\left[1-F_{T|\theta}(z_i|\theta)\right]^{1-\delta_i} \tag{3}$$

Model distribution of *T* subject to censoring

With the inferred parameter set θ, the probability distribution of T is given as $P_{T|\theta}(t|\theta)$.

The life time of an individual cell in the measurement depends on various factors, such as the division moment and the duration of the measurements. Here, the distribution of the life time Y is obtained directly from the observations of cell life times during the microscopy measurements, rather than being modeled. The inferred distribution of actual samples T' with the distribution of life time Y known is:

$$P_{T'|\theta}(t \mid \theta) = P_{T|\theta}(t \mid \theta) \times P(Y > t) = P_{T|\theta}(t \mid \theta)(1 - F_Y(t)) \tag{4}$$

By comparing $P_{T'|\theta}(t \mid \theta)$ with the empirical distribution of the actual samples (δ_i=1, z_i) using Pearson's chi-squared test, we can calculate the goodness of fit of θ's estimation.

III. Solving the deterministic model of inducer dynamics

Model of inducer dynamics

The model of inducer dynamics is described (as in equations (7) and (8) in the manuscript) as follows:

$$\delta I_m / \delta t = k_o - I_m \times k_i \tag{5}$$

$$\delta I / \delta t = I_m \times k_i - I \times d_I \tag{6}$$

We first find the solution for the inducer level in the periplasmic space (I_p):

$$\frac{\delta I_m}{k_o - I_m k_i} = \delta t \tag{7}$$

By integrating both sides of the equation, we obtain:

$$-\frac{\ln\left(k_o - I_m k_i\right)}{k_i} = t + C_1 \tag{8}$$

$$\leftrightarrow k_o - I_m k_i = C_1 . e^{-k_i t} \tag{9}$$

$$\leftrightarrow I_m = \frac{k_o - C_1 e^{-k_i t}}{k_i} \tag{10}$$

At $t=0$, $I_m(0)=0$, thus $C_1=k_o$. The solution for I_m is:

$$I_m(t) = \frac{k_o(1 - e^{-k_i t})}{k_i} \tag{11}$$

The differential equation for $I(t)$ becomes a first order linear differential equation:

$$\frac{\delta I}{\delta t} + I.d_I = k_o\left(1 - e^{-k_i t}\right) \tag{12}$$

The general solution for this equation is:

$$I(t) = \frac{\int u(t) k_o\left(1 - e^{-k_i t}\right) dt + C_2}{u(t)} \tag{13}$$

in which $u(t) = e^{\int d_I dt} = e^{d_I t}$. C_2 is a constant determining the initial condition $I(0)$. Thus:

$$I(t) = \frac{\int e^{d_I t} k_o\left(1 - e^{-k_i t}\right) dt + C_2}{e^{d_I t}}$$

$$= \frac{k_o(\int e^{d_I t} dt - \int e^{\left(d_I - k_i\right)t} dt) + C_2}{e^{d_I t}} = \frac{k_o}{d_I} - \frac{k_o e^{-k_i t}}{d_I - k_i} + \frac{C_2}{e^{d_I t}} \tag{14}$$

$$= \frac{k_o(d_I - k_i) - d_I k_o e^{-k_i t} + C_2 d_I (d_I - k_i) e^{-d_I t}}{d_I (d_I - k_i)}$$
$$= \frac{k_o\left(d_I e^{-k_i t} + C_2 d_I (k_i - d_I) e^{-d_I t} + k_i - d_I\right)}{d_I (k_i - d_I)}$$

At $t=0$, $I(0)=0$, $C_2 d_I (k_i - d_I) = -k_i$.

The final solution for the intracellular inducer quantity over time is therefore:

$$I(t) = \frac{k_o\left(d_I e^{-k_i t} - k_i e^{-d_I t} + k_i - d_I\right)}{d_I (k_i - d_I)} \quad (15)$$

IV. Model distribution of t_0

From the models of inducer intake and of transcription, we use the Chemical Master Equation (CME) [11] to calculate the first moment of open complex formation in each cell, which is followed, shortly after, by the release of a transcript [12,13]. For this, we assume that, upon this release, the promoter is unable to transcribe any subsequent RNA. Given this approximation, the master equation for the promoter in each of its three possible states is given by:

$$\delta P(\mathrm{Pr},t)/\delta t = -k_c f_R(t) \times P(\mathrm{Pr},t) \quad (16)$$

$$\delta P(\mathrm{Pr}_c,t)/\delta t = k_c f_R(t) \times P(\mathrm{Pr},t) - k_o \times P(\mathrm{Pr}_c,t) \quad (17)$$

$$\delta P(\mathrm{Pr}_o,t)/\delta t = k_o \times P(\mathrm{Pr}_c,t) \quad (18)$$

$P(Pr,t)$, $P(Pr_c,t)$ and $P(Pr_o,t)$ are the probabilities that the promoter is in its primary state, in closed complex state and in open complex state, respectively, at time t. Due to the high amount of repressors in the cells [14], we ignore the leakiness of the target gene (from our measurements, we observed that, on average, it takes more than 1 hour for ~10% of the cells to produce one spurious RNA, when not induced). Given this, we set the probability of the promoter to be in its primary state, P(Pr,0), to 1 and to be in the other two states ($P(Pr_c,0)$ and $P(Pr_o,0)$) to 0.

V. Dilution rate of regulatory molecules at various induction levels

The dilution rate of regulatory molecules (d_I) is calculated from the expansion rate of the cells' volume. As *E. coli* grows mostly by elongating through its major axis length, while leaving its minor axis length unchanged, the relative increase in cell's volume can be approximated by the increase in the cell's major axis length.

Cell growth in liquid media

To test for the effect of IPTG induction on the cells growth rate at 37 ˚C, we first measured cell growth in liquid media. Cells were grown overnight at 30 ˚C with aeration and shaking in LB media, supplemented with the appropriate antibiotics, before being diluted in fresh LB medium until an OD600 ≈ 0.1 and pre-incubated for 2 hours without inducers. In the remaining hours, cells were either left to grow normally or grown in the presence of IPTG at the concentration of 0.25mM and 1mM. The optical density (OD) curves at 0mM, 0.25mM and 1mM IPTG concentrations were sampled every 30 minutes for 5 hours (Figure S2).

From Figure S2, during the first 4 hours of the measurements there is little difference between the normalized OD curves, indicating that, in the range of concentrations tested, IPTG does not have any notable effect on cell growth.

Cell growth on agarose gel

Next, we obtained the cell growth rate during the microscopy measurements, where cells are kept on agarose gel as described in the Methods section of the main manuscript. As only a few cell cycles were observed in M63 media during 2 hour-long measurements, we estimated the cell growth rate from the elongation rate of all cells' major axis rather than the cells' doubling time.

From the time lapse confocal images, cells were segmented and the length of the major axis was extracted at each frame. For each cell, we fitted a linear function to the logarithm of the major axis length over time and obtained the slope coefficient d_I', equivalent to the cell's elongation rate. The doubling time T_d' of each cell is inferred from d_I' as follows:

$$T_d' = \frac{ln(2)}{d_I'} \tag{19}$$

The distributions of T_d' at different induction levels spans over a wide range of durations, suggesting a noisy dilution rate when cells are on the 3% agarose gel. The distributions share a mode of around ~8400 seconds. To eliminate any effects of noise in the dilution rate of regulatory molecules, for the analysis of t_0, we selected 'normal' cells with a doubling time T_d'~8400 s, using a margin for selection of 15% of the mode's value. Finally, from the value of T_d', the dilution rate d_I of the selected cells is found to be:

$$d_I = \frac{ln(2)}{T_d'} = 8.25 \times 10^{-5}\ (s^{-1}) \tag{20}$$

Since cells grew exponentially during the measurements at a rate of d_I ~8.25×10^{-5} s^{-1} (doubling time of ~140 minutes) in all conditions, it is reasonable to assume that the cells were unaffected by the inducer in the range of concentrations tested (in this regard see, e.g. [15]).

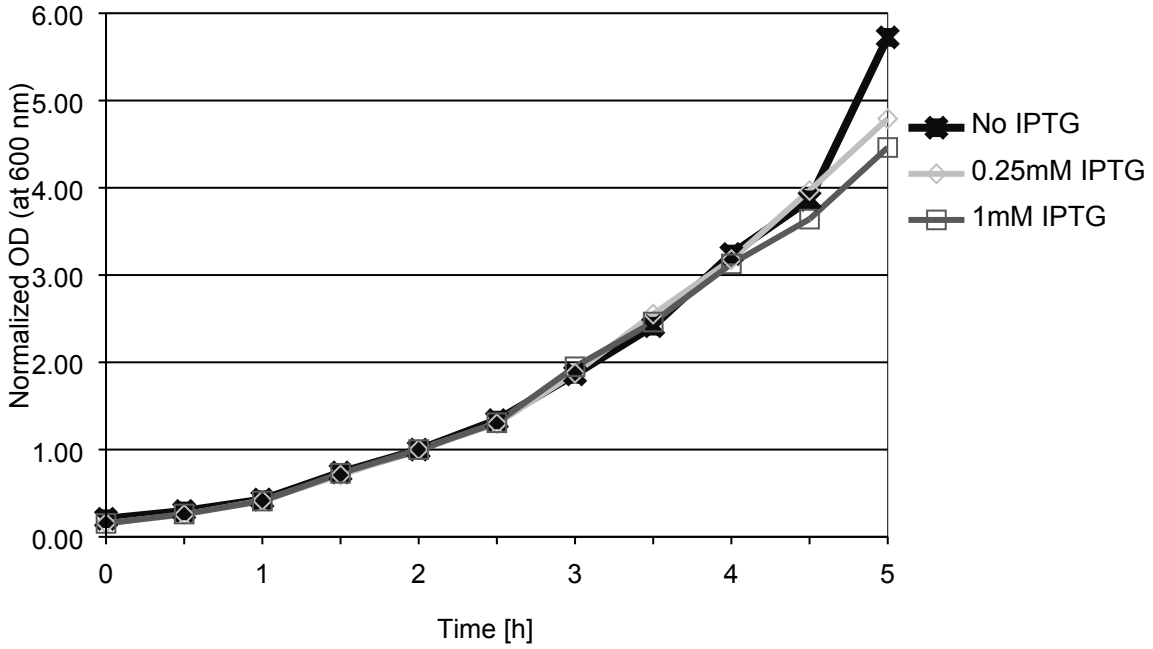
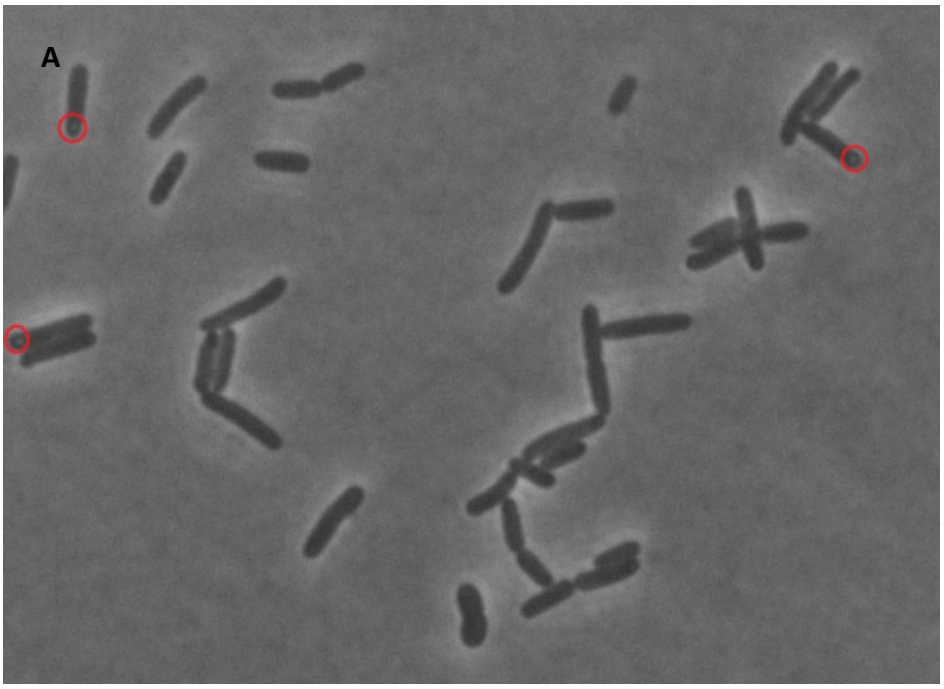
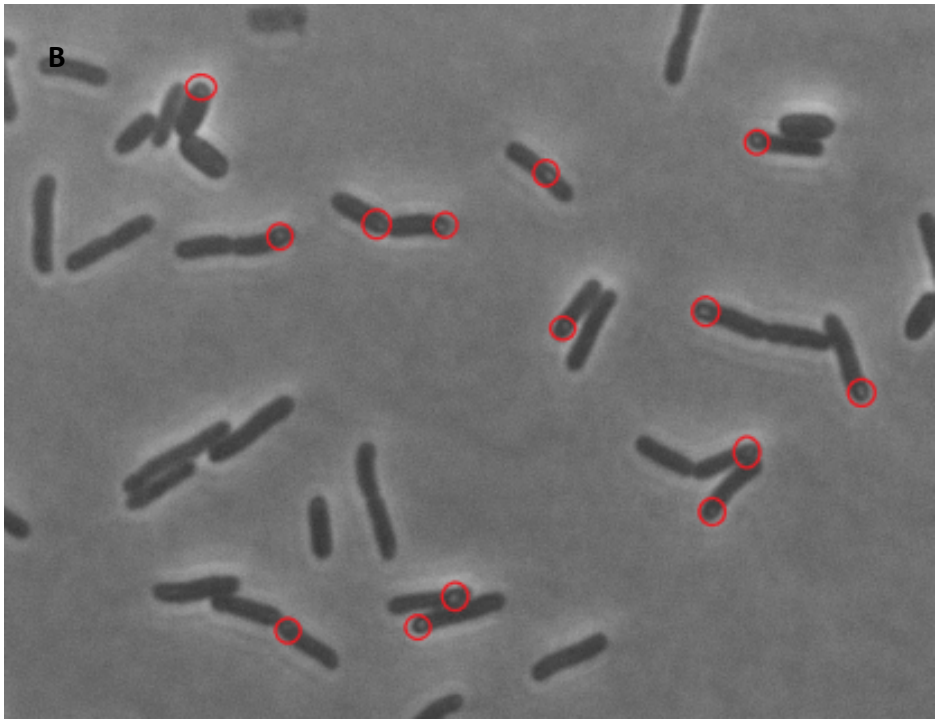


Figure S2. Normalized optical density (OD) curve at 0.25mM (diamond) and 1mM (square) IPTG and without IPTG (cross). Inducers are added at the end of the second hour, where the normalized OD's values equal 1.

VI. Formation of inclusion bodies at high inducer concentrations

We use phase contrast microscopy to examine the fraction of cells with inclusion bodies as a function of IPTG concentration in the media. Cells were grown overnight at 30 ˚C with aeration and shaking in LB media, supplemented with the appropriate antibiotics, before being diluted in fresh LB medium until an OD600 ≈ 0.1 and pre-incubated for 2 hours without inducers. In the remaining hours, cells were incubated in the presence of aTc at 100ng/L and IPTG at 1mM, 2mM and 4mM before being placed under the microscope. From the phase contrast images, we manually detected the presence of inclusion bodies (shown as a bright spot) in each cell. Example images of cells with marked inclusion bodies are shown in Figure S3.


A


B

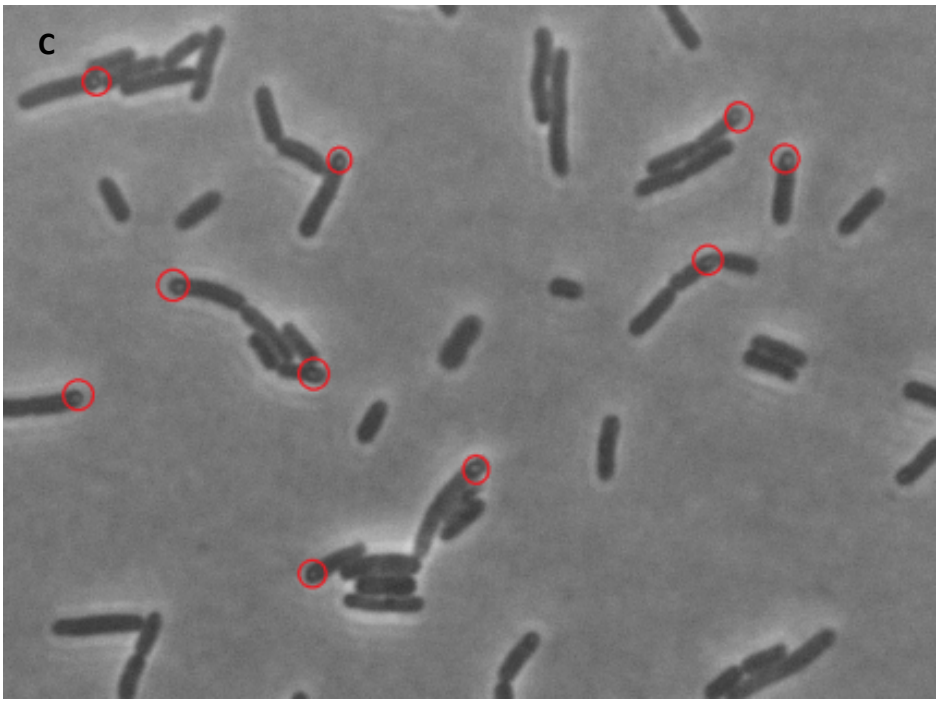


Figure S3. Phase contrast images with cells with marked inclusion bodies (appears as a bright spots), when induced with IPTG concentrations of (Top) 1mM, (Middle) 2mM and (Bottom) 4mM.

VII Temporal fluorescence intensity of MS2d-GFP tagged RNA molecules

The technique of detecting new RNA molecules in individual cells as these appear in time lapse microscopy images using the MS2d-GFP RNA-tagging system (ref. 6 in main manuscript) consists of fitting the total corrected RNA spot intensity with a step-increasing function (see section I of this document).

For this method to be valid, it is necessary that new RNA molecules appear nearly fully-tagged when first detected, so as to cause a significant "jump" in the total spots fluorescence intensity of the cell[7]. This is possible if the speed of elongation at the target gene and MS2d-GFP binding is not much longer than the interval between consecutive images, which in our measurements is 1 minute long.

Also, it is necessary that an MS2d-GFP tagged RNA, once tagged, does not degrade significantly (neither abruptly nor gradually) during the measurement period (so as to allow using a step increasing function). Note that, nevertheless, the method can tolerate infrequent "blinking" of the tagged RNAs, due to moving out of focus transiently, without loss of information[7].

To validate the two assumptions, we observed the fluorescence intensity of individual, RNA spots over time (1 min^{-1}). As newly produced RNA spots could appear and compensate for the loss of intensity (abrupt or gradual) of the existing spots (resulting in the underestimation of the spots' degradation rate), we conducted the observation on a non-induced target gene. Namely, following the protocol described in the main manuscript (except for the induction of expression of the RNA target for MS2d-GFP), we observed sufficient cells during a period of 3 hours so that at least 40 RNA spots appearances could be detected (during that period of time, less than 1 in 10 cells produced an RNA spot). Note that, by

inspection, we never observed the appearance of two new fluorescent spots in a cell at the same time moment and no cell ever contained 2 spots.

To test the first assumption, from the time-lapse images, we obtained the fluorescence intensity of 40 individual tagged RNAs for 30 minutes, since first detected. From these, we found that there is no significant RNA fluorescence increase after its detection. That is, new RNA molecules are nearly fully-tagged when first detected, as expected from the frequency of image acquisition (1 min^{-1}) and the expected speed of transcription elongation and MS2d-GFP binding (tens of seconds[16,17]). This is visible in Figure S4, where the mean spot fluorescence over time is shown. Note how, following the detection of the spots at moment 0 (synchronized for easier visualization), their mean fluorescence over time does not increase further in subsequent time moments.

To test the second assumption, we fitted the intensity of each RNA spot over time with a decaying exponential function and inferred the degradation rate of the spot intensity. We obtained a mean decaying rate of ~8.1×10^{-5} s^{-1}, corresponding to a mean half-life of ~144 mins, which is much longer than our observation window for Δt (60 mins). As such, we conclude that, during the measurement period, the fluorescence of tagged RNAs does not decrease significantly over time (gradually or abruptly), in agreement with previous reports using the same RNA detection system[2,3,16,17].

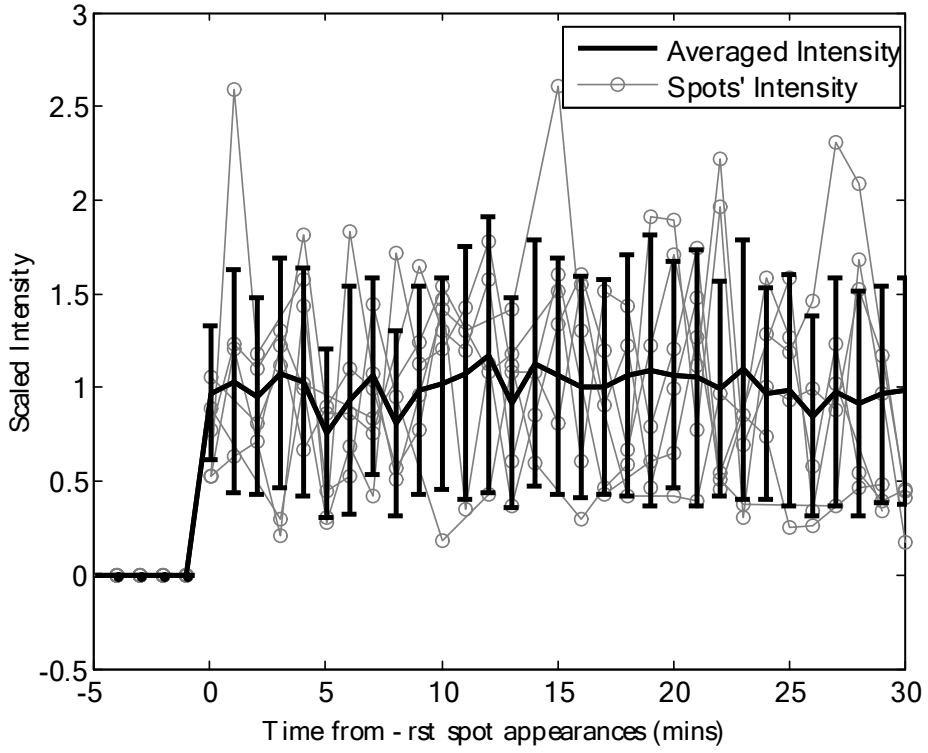


Figure S4. Fluorescence of tagged RNAs in *E. coli* cells over time. Each of the five thin lines shows the fluorescence of a single tagged RNA molecule (randomly selected from the data) since first detected, for a period of 30 minutes. The solid black line shows the mean fluorescence intensity of individual tagged RNA molecules (40 molecules tracked), along with the sample standard deviation (vertical bars).

The above results are in agreement with previous studies. Regarding the dynamics of RNA production, the present results agree with previous data on the rate of transcription elongation in *E. coli*. Namely, at 37ºC, this rate is expected to be between ~60 and ~90 base pairs (bp) per second[18–20]. Given that the target gene is ~3200 bp long[16], the RNA polymerase should produce a complete transcript in ~35 to ~50 s, which is faster than our imaging interval (60 s).

Meanwhile, regarding the lack of degradation of tagged RNAs, our results are expected given previous studies on the coat protein of bacteriophage MS2[16,21,22], which showed that most of the MS2 binding sites are constantly occupied by (at least 70) MS2d-GFP proteins, which results in the 'immortalization' of the target RNA due to isolation from RNA-degrading enzymes [16,17].

REFERENCES

1 J. Mäkelä, M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2013, **41**, 6544–52.

2 A. B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja and A. S. Ribeiro, *Nucleic Acids Res.*, 2012, **40**, 8472–83.

3 M. Kandhavelu, J. Lloyd-Price, A. Gupta, A. B. Muthukrishnan, O. Yli-Harja and A. S. Ribeiro, *FEBS Lett.*, 2012, **586**, 3870–5.

4 M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen, J. Lloyd-Price, O. Yli-Harja and A. S. Ribeiro, *BMC Syst. Biol.*, 2011, **5**, 149.

5 S. Chowdhury, M. Kandhavelu, O. Yli-Harja and A. S. Ribeiro, *J. Microsc.*, 2012, **245**, 265–75.

6 A. Häkkinen, A.-B. Muthukrishnan, A. Mora, J. M. Fonseca and A. S. Ribeiro, *Bioinformatics*, 2013, **29**, 1708–9.

7 M. Kandhavelu, A. Häkkinen, O. Yli-Harja and A. S. Ribeiro, *Phys. Biol.*, 2012, **9**, 026004.

8 H. Mannerstrom, O. Yli-Harja and A. S. Ribeiro, *Eurasip J. Bioinforma. Syst. Biol.*, 2011, **2011**, 11–15.

9 D. R. Cox, *J. R. Stat. Soc. Ser. B*, 1972, **34**, 187–220.

10 H. Koul, V. Susarla and J. Van Ryzin, *Ann. Stat.*, 1981, **9**, 1276–1288.

11 D. T. Gillespie, *J. Stat. Phys.*, 1977, **16**, 311–318.

12 L. M. Hsu, *Biochim. Biophys. Acta*, 2002, **1577**, 191–207.

13 K. M. Herbert, A. La Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick and S. M. Block, *Cell*, 2006, **125**, 1083–94.

14 R. Lutz and H. Bujard, *Nucleic Acids Res.*, 1997, **25**, 1203–10.

15 S. Cooper, *Theor. Biol. Med. Model.*, 2006, **3**, 10.

16 I. Golding and E. C. Cox, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 11310–5.

17 I. Golding, J. Paulsson, S. M. Zawilski and E. C. Cox, *Cell*, 2005, **123**, 1025–36.

18 U. Vogel and K. a J. F. Jensen, 1994.

19 P. P. Dennis, M. Ehrenberg, D. Fange and H. Bremer, *J. Bacteriol.*, 2009, **191**, 3740–3746.

20 J. Ryals, R. Little and H. Bremer, 1982, **151**, 879–887.

21 S. J. Talbot, S. Goodman, S. R. E. Bates, C. W. G. Fishwick and P. G. Stockley, *Nucleic Acids Res.*, 1990, **18**, 3521–3528.

22 D. Fusco, N. Accornero, B. Lavoie, S. M. Shenoy, J. M. Blanchard, R. H. Singer and E. Bertrand, *Curr. Biol.*, 2003, **13**, 161–167.

PUBLICATION II

N.S.M. Goncalves, L. Martins, H. Tran, S.M.D. Oliveira, R. Neeli-Venkata, J.M. Fonseca and A.S. Ribeiro, "In vivo single-molecule dynamics of transcription of the viral T7 Phi 10 promoter in *Escherichia coli", Proceedings of the 8th International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies* (BIOTECHNO 2016), June 26-30, Lisbon, Portugal. ISBN: 978-1-61208-488-6 (pp. 9-15), 2016.

In Vivo Single-Molecule Dynamics of Transcription of the Viral T7 Phi 10 Promoter in *Escherichia coli*

Nadia S.M. Goncalves[1], Huy Tran[1], Samuel M.D. Oliveira[1], Ramakanth Neeli-Venkata[1], Andre S. Ribeiro[1]
[1] Laboratory of Biosystem Dynamics, Tampere University of Technology, Finland.
e-mail: andre.ribeiro@tut.fi

Leonardo Martins[2], José M. Fonseca[2]
[2]Computational Intelligence Group, CTS/UNINOVA, Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa, Portugal.
e-mail: jmf@uninova.pt

***Abstract* – We study the dynamics of transcription initiation of the T7 Phi 10 promoter as a function of temperature, using quantitative polymerase chain reaction (qPCR) and *in vivo* single-cell, single-ribonucleic acid (RNA) time-lapse microscopy. First, from the mean and squared coefficient of variation of the empirical distribution of intervals between consecutive RNA appearances in individual cells, we find that both the mean rate and noise in RNA production increase with temperature (from 20ºC to 43ºC). Next, the process is shown to be sub-Poissonian in all conditions, suggesting the existence of more than one rate-limiting step and absence of a significant ON-OFF mechanism. Next, from the kinetics of RNA production for varying amounts of T7 RNA polymerases, we find that as temperature increases, the fraction of time that the T7 RNA polymerase spends in open complex formation increases relative to the time to commit to closed complex formation, due to changes in the kinetics of open complex, closed complex, and reversibility of the closed complex formation. We conclude that the initiation kinetics of the T7 Phi 10 promoter changes with temperature due to changes in the kinetics of its rate-limiting steps.**

Keywords – Transcription; Open and closed complex formation; T7 Phi 10 promoter

I. INTRODUCTION

The bacteriophage T7 is an obligate lytic phage that infects *Escherichia coli*, using the host system to produce up to 100 progeny phages in less than 25 min, in optimal conditions [1]. One of the major gene products of T7 bacteriophage is the T7 RNA polymerase (T7 RNAP) [2]. This is a single subunit enzyme, with a high specificity towards T7 promoters via the recognition of a highly conserved 23bp consensus sequence [3]. Early studies have shown that the T7 RNAP transcription rate is sequence dependent and depends on environmental conditions [4][5][6]. Given that the infection process of T7 bacteriophage is not only fast but it also requires a balance between the number of phages and the amount of capsid proteins produced [7], the phage needs to coordinate the dynamics of transcription of the viral genes, as this is likely critical for its success.

It is known that the dynamics of gene expression, as well as of many other cellular processes, depends on environmental factors, particularly temperature [8]. Consequently, microorganisms have evolved mechanisms that allow them to cope with both sudden as well as slow temperature changes [9][10]. *E. coli*, for example, can survive in a wide range of temperatures. Similarly, it has also been shown that the T7 bacteriophage is capable of coping with these fluctuations and wide ranges [5].

Even though robustness to sudden temperature changes and wide temperature ranges is crucial for the survival of microorganisms, so far, little is known about what are the consequences of these environmental changes on the *in vivo* transcription kinetics of the T7 promoter. In addition, most studies characterizing the transcription initiation kinetics of T7 promoters have mostly used *in vitro* measurement techniques [5][11].

To address this issue, here we use recently developed measurement strategies that use single-cell, single-RNA *in vivo* detection techniques [12] and use them to study in detail the kinetics of transcription initiation of the T7 *Phi10* (*Φ10*) promoter as a function of temperature.

The remaining of this article is organized as follows: Section II describes the methods used and measurements conducted. Section III presents the results from these experiments. In Section IV, we conclude by presenting our interpretation of the results and our assessment of their relevance, as well as additional considerations for future work.

II. METHODS

In this section, we describe the measurements conducted in this study. Each subsection presents a detailed explanation of the experiments performed.

A. Strain and plasmids

The strain *E. coli* BL21(DE3) (New England Biolabs, USA) was used to express the target and reporter genes. This strain has a copy of the T7 bacteriophage gene 1 coding for T7 RNAP controlled by the P_{lacUV5} promoter and integrated in the chromosome [13] (Figure 1A).

The single copy F-plasmid pBELOBAC11, carrying the *Φ10-mCherry-48bs* sequence (constructed for this work) was inserted in the host strain. It produces the target RNA, with an array of 48 MS2 binding sites (*48bs*) under the control of a T7 *Φ10* promoter, cloned from the plasmid pRSET/EmGFP (ThermoScientific, USA).

A second plasmid, pZA25-GFP (Green Fluorescent Protein) [14] (a gift from Orna Amster-Choder, Hebrew University of Jerusalem, Israel), was also inserted in the host strain. It contains the reporter gene *ms2-gfp*, placed under the control of P_{BAD} promoter. This reporter gene encodes for the fusion protein MS2-GFP, which binds the target RNAs and renders them visible as bright spots under the confocal microscope [15] (Figure 1B). From here onwards we refer to the T7 *Φ10* promoter as T7 promoter.

 ISBN: 978-1-61208-488-6

B. Microscopy

For live cell microscopy, BL21(DE3) cells were incubated in M63 medium supplemented with Glucose (0.4%) and the appropriate concentration of Chloramphenicol and Kanamycin (Sigma Aldrich, USA) and was grown overnight at 30°C, with shaking (250 rpm). Cells from the overnight culture were then diluted in fresh M63 medium, with an initial OD600 ~ 0.05, and incubated at 37°C, for 90 minutes with shaking (250 rpm). Then, cells were pelleted and re-suspended in ~100 μl of M63 medium. Four microliters of cells were placed between a 3% agarose gel pad, made with M63 medium, and a glass coverslip before assembling the imaging chamber (CFCS2, Bioptechs, USA). Two hours before the microscopy measurements, a flow of fresh M63 medium at 37°C containing the reporter inducer (0.8% L-arabinose) was initiated with a peristaltic pump at a rate of 1 ml/min to produce sufficient MS2-GFP molecules in the cells to detect the target RNA in all experiments. Note that we shifted the temperature in the chamber from 37°C to 20°C or to 43°C (depending on the condition studied), 20 minutes prior to inducing the target system.

To activate the target system, we induced the production of T7 RNAP, controlled by P_{lacUV5}, by introducing a new flow (1 ml/min) of M63 medium containing 0.8% L-arabinose and Isopropyl β-D-1-thiogalactopyranoside (IPTG) at various concentrations (see below). Once synthesized, T7 RNAPs will bind the T7 promoter and transcribe *48bs* RNAs, which are quickly bound by MS2-GPF molecules and appear under the confocal microscope as bright spots (Figure 1B).

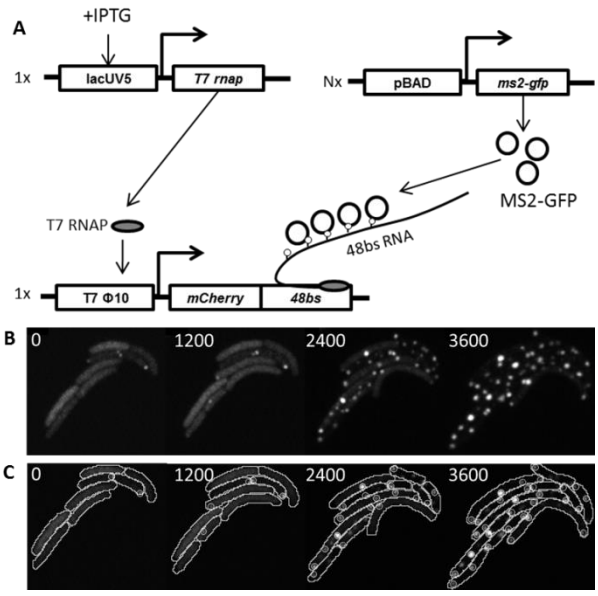


Figure 1. (A) Diagram of the measurement system, depicting the target and reporter genes along with the MS2-GFP tagging process. (B) Confocal microscope images at subsequent time points showing the cells and the MS2-GFP tagged RNA molecules inside. (C) Segmented cells and RNA spots within.

Cells imaging was started at the same time as the introduction of the flow containing IPTG. Images were captured every minute for 2 hours using an inverted microscope Nikon Eclipse (Ti-E, Nikon, Japan). Both confocal images (confocal C2+ scanner connected to LU3 laser system, Nikon) and phase contrast images (DS-Fi2 CCD-camera) were collected.

Examples of confocal images of cells are shown in Figure 1B. Note that, at the end of the time series, the fluorescent background in some cells becomes dimmed due to the produced RNAs having bound most MS2-GFP molecules in the cytoplasm.

C. Image analysis

The segmentation of cells and detection of RNA spots were performed by the software "iCellFusion" [16]. It first applies the cell segmentation on phase contrast images using a Gradient Path Labelling Algorithm [17]. Then, it performs the inter-modal image registration between phase-contrast images and the corresponding fluorescence images and exports the segmentation results on fluorescence images. The spot detection was performed as in [18]. Results from the segmentation and spot detection algorithms are shown in Figure 1C.

D. Data analysis

The cell-to-cell variability in the kinetics of intake of IPTG, which affects the activation of P_{lacUV5} [19][20], creates extrinsic variability regarding when the first RNA appears in each cell. Since we are only interested in the intrinsic noise of the transcription process, to correct for this, we fit the total spot intensity in each cell over time with an activation function:

$$x(t_{activation}, c, t) = c \times H(t - t_{activation}) \times (t - t_{actiovation}) \quad (1)$$

where t is time, $t_{activation}$ is activation time of T7 when the *48bs* RNA production reaches steady state, c is the mean increment rate of total spot intensity and H is a unit step function. With the function in (1) fitted using least mean squared, we find $t_{activation}$ for each cell. The total spot intensities are then aligned using the inferred $t_{activation}$, so as to compare the kinetics of active T7 promoters in individual cells.

We found by inspection that, at 37°C, in the first ~18 minutes, the mean curve of the aligned total spot intensities can be well fitted with a linear function, indicating that RNA production in most cells reached a steady state after their corresponding $t_{activation}$. After the 18th minute, the mean spot intensity increases with decreasing speed, visibly due to increasing shortage of free MS2-GFP. Therefore, for this condition, we select the data in the first 18 minutes for RNA quantification as in [18][21]. Note also that for different temperatures and IPTG concentrations, the window for RNA quantification differs (data not shown).

E. qPCR

Cells grown to OD600 ~0.4 were induced with the appropriate IPTG concentration (5-250 μM) for 1 hour, at the specific temperature (20°C, 37°C and 43°C). Afterwards, cells were fixed with RNAprotect bacteria reagent (Qiagen, Germay), followed by total RNA isolation, DNase I treatment (ThermoScientific, USA) and cDNA

 ISBN: 978-1-61208-488-6

synthesis (BioRad, USA). The qPCR master mix contained iQ SYBR Green supermix (Biorad, USA) with primers for the target gene, the T7 RNAP and the reference gene at a final concentration of 200 nM. The primers for the target gene were (Forward: 5' CACCTACAAGGCCAAGAAGC 3' and Reverse: 5' TGGTGTAGTCCTCGTTGTGG 3') for the mCherry region. To quantify the T7 RNAP, the primers used were (Forward: 5' TCCTGAGGCTCTCACCGC 3' and Reverse: 5' GATACGGCGAGACTTGCGA 3'). For the reference gene 16SrRNA, the primers were (Forward: 5' GCTACAATGGCGCATACAAA 3' and Reverse: 5' TT CATGGAGTCGAGTTGCAG 3'). The data from CFX Manager TM Software was used to obtain the relative gene expression and standard error [22].

F. Model of T7 promoter transcription kinetics

To study how the kinetics of the T7 promoter changes with temperature, we assume the modelling strategy of transcription proposed in [23][24][25], derived from both *in vitro* and *in vivo* studies on viral [11][26] and *E. coli* promoters [8][25][27][28][29]. The model of transcription kinetics of T7 promoter is as follows:

$$R + Pr \underset{k'_{cc}}{\overset{k_{cc}}{\rightleftarrows}} Pr_{cc} \xrightarrow{k_{oc}} Pr_{oc} \xrightarrow{\infty} Pr + R + RNA \quad (2)$$

where R is an active T7 RNAP, Pr is a free promoter, Pr_{cc} is a fully formed closed complex, and Pr_{oc} is a fully formed open complex. The closed complex formation occurs at the rate k_{cc}. Once the closed complex is formed, the promoter can either be unbound by the R at the rate k'_{cc} or undergo open complex formation at the rate k_{oc}. Due to fast promoter escape [30], the low frequency of abortive initiation [6] and the fast rate of elongation of T7 RNAP [5][11][31], we assume that the RNAP and target RNA are released soon after completion of the open complex. Note that this model does not include an ON-OFF mechanism since T7 is a constitutive promoter.

From (2), the mean of the interval distribution (Δt) between consecutive transcription events is:

$$\Delta t(R) = \frac{(k'_{cc} + k_{oc})}{R k_{cc} k_{oc}} + \frac{1}{k_{oc}} = \frac{1+K}{R k_{cc}} + \frac{1}{k_{oc}} = \tau(R) + \tau_{oc} \quad (3)$$

where R is the abundance of T7 RNAP in the cell, K is ratio between k'_{cc} and k_{oc} indicating the reversibility of the closed complex, $\tau(R)$ is the time for an RNAP to commit to the open complex formation, and τ_{oc} is time for open complex formation. From (3), the production interval $\Delta t(R)$ is a linear function of the inverse of T7 RNAP level ($1/R$), and thus:

$$\tau_{oc} = \Delta t(R = \infty) \quad (4)$$

With each set of values of $R.k_{cc}$, K, and k_{oc}, we use the Chemical Master Equation (CME) to find the distribution of intervals between consecutive RNA production events, from which the mean rate and noise in transcription are extracted.

III. RESULTS

This section comprises the results, obtained from the measurements, which are presented into three separate subsections.

A. Validation of the construct with the T7 promoter

First, to validate that the T7 promoter inserted in the F-plasmid (Methods) is active, we measured the RNA levels of the T7 RNAP and of the target gene by qPCR for varying IPTG concentrations (which control the expression of T7 RNAP). Results are shown in Figure 2.

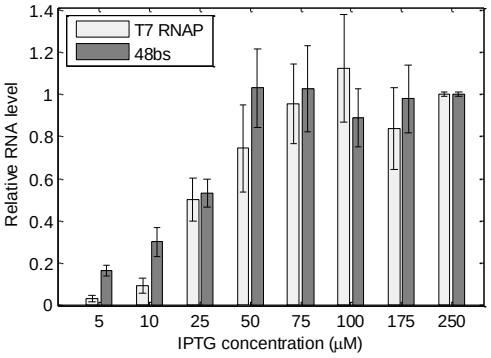


Figure 2. Relative RNA levels of T7 RNAP (light grey) and target gene (*48bs*) (dark grey) at 37°C with varying IPTG concentrations as measured by qPCR. Also shown for each condition are the standard errors from 3 technical replicates.

From Figure 2, first, both the T7 RNAP's and target gene's levels do not increase significantly with increasing IPTG concentrations beyond 100 μM, suggesting that the lacUV5 promoter is fully induced at this concentration. In Figure 2, the data is normalized by the RNA levels at 250 μM IPTG. We validated these measurements, in the case of the target RNA, by observing its production dynamic at 175 μM, 250 μM and 1000 μM IPTG at 37°C under the microscope (via MS2-GFP tagging, Methods).

While we observed changes in the mean activation time of the T7 promoter with changing IPTG concentration (data not shown), we did not observe a significant change in mean transcription rate ($\mu_{\Delta t}$ ~350 s).

Finally, we find an increase in both the T7 RNAP's and target gene's RNA expression with increasing IPTG concentration, demonstrating that both genes are active. Note the close correlation between the activities of the two genes, indicating that the T7 promoter is, as expected, under the control of the T7 RNAP.

B. T7 promoter dynamics at various temperatures

We next observed the transcription dynamics of T7 promoter at different temperatures (within sub-optimal intervals). The IPTG concentration used was 250 μM, in order to ensure that lacUV5 is fully induced in all conditions. Under the microscope, all cells appeared to grow normally, with reduced division rates at lower temperatures. In particular, cells' mean doubling times were 50 min, 60 min and 100 min at 43°C, 37°C and 20°C respectively.

 ISBN: 978-1-61208-488-6

From the RNA numbers over time in individual cells as observed by microscopy at different temperatures, we extracted the mean duration (μ) and coefficient of variation squared (CV^2) of the intervals between consecutive RNA appearances in individual cells as in [18][32]. Results are shown in Table I.

For each temperature, the number of cells observed, the number of samples collected (intervals between consecutive RNAs in individual cells), and the mean and CV^2 of the intervals between consecutive RNA appearances in individual cells are shown. The final column shows the relative RNA levels of T7 RNAP measured by qPCR (normalized by RNA levels at 37°C).

TABLE I. *IN VIVO* TRANSCRIPTION INITIATION DYNAMICS OF THE *T7* PROMOTER AT DIFFERENT TEMPERATURES MEASURED BY MS2-GFP TAGGING OF RNA.

T (°C)	No. Cells	No. Samples	μ (s)	CV^2 (σ^2/μ^2)	Relative T7 RNAp no.
43	150	508	320	0.95	0.86
37	111	311	352	0.85	1
20	68	105	518	0.62	0.46

From Table I, somewhat surprisingly but in agreement with a previous observation by *in vitro* methods [5], the mean length of the RNA production intervals, μ, increases with decreasing temperature. Overall, this indicates that the *in vivo* kinetics of transcription initiation of the T7 promoter is temperature dependent.

Notably, the mean transcription rates *in vivo* are approximately one order of magnitude smaller than those reported from *in vitro* tests [5][11]. This weaker activity in live cells is likely due to the more limited amount of T7 RNAP (bound by the limits in lacUV5's activity) and limited resources (ATP, ribonucleotides, etc.) in the host cells to support the viral transcription process.

Also in Table I, the noise in transcription (as measured by CV^2) decreases with decreasing temperature. A previous work reported a similar result for P_{tetA}, a native promoter of *E. coli* [8].

In addition, in all conditions, the RNA production appears to be a sub-Poissonian process ($CV^2<1$). This suggests that it consists of multiple rate-limiting steps rather than being dominated by an ON-OFF process [11]. Similar *in vivo* sub-Poissonian dynamics of transcription has been observed in several *E. coli* promoters, native and synthetic, when under full induction [8][28][33].

Overall, the results suggest that the process of transcription initiation of the T7 promoter by the T7 RNAP is similar to that of *E. coli* native promoters.

Meanwhile, from the relative numbers of T7 RNAP as measured by qPCR, we find that unlike when controlling with IPTG concentrations, the kinetics of RNA production of the target promoter T7 no longer follows solely the T7 RNAP numbers, as its production rate is not maximized at 37°C while T7 RNAP numbers are. Therefore, we conclude that the observed changes in the T7 promoter dynamics are due to changes in both the kinetic rates of T7 transcription and in T7 RNAP numbers.

C. Estimation of kinetic rates of the T7 promoter

We searched for changes in the underlying kinetics of transcription initiation of the T7 promoter (i.e. in the duration of the closed and open complex formation) with temperature that can explain the changes in the target RNA production with changing temperature.

To quantify how the kinetic rates of T7 promoter evolve with temperature, we followed the strategy proposed in [12] by investigating, for each temperature, how the transcription activity on T7 promoter is affected by the T7 RNAP abundance. This abundance should affect the kinetics of the closed complex formation, but not that of the steps following the closed complex [12].

Here, the T7 RNAP levels, varied by employing different IPTG concentrations (5 μM, 10 μM, 25 μM, 50 μM and 250 μM), and the T7 promoter's activity are measured relatively by qPCR. From these, we infer what would be the relative rate of RNA production given an infinite amount of T7 RNAP in cells (Methods). This rate should correspond to the fraction of time of the transcription initiation process that corresponds to the open complex formation alone [12]. Results for each temperature condition are shown in 'τ plots' in Figure 3.

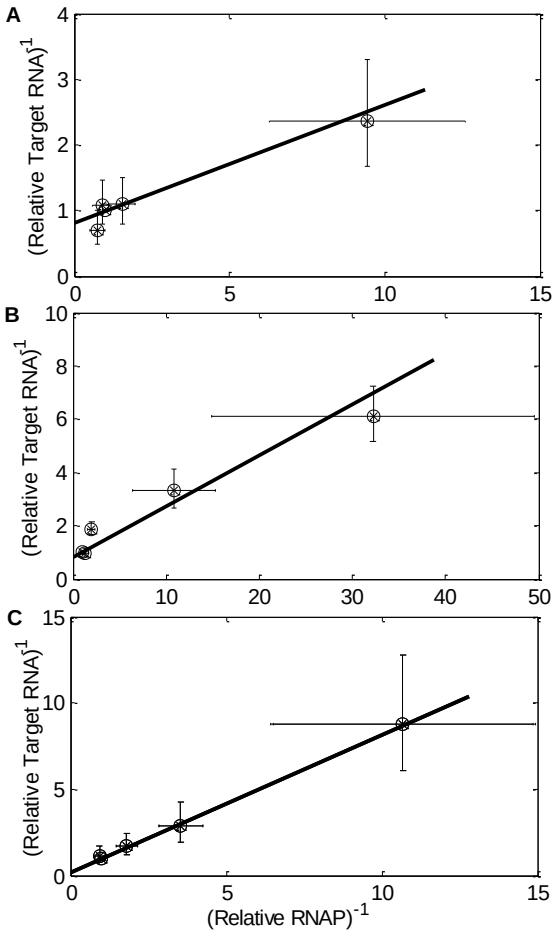


Figure 3. τ plots for T7 promoter activity at different temperatures: (A) 43°C (B) 37°C and (C) 20°C.

In Figure 3, the data is shown relative to the RNA and RNAP levels at 250 μM IPTG. Error bars represent the

standard error of the mean (SEM) of the estimate of the inverse of the relative rates of transcription for the target RNA and T7 RNAP in each condition. The lines are Weighted Total Least Squares fits [34]. Errors are calculated including the uncertainty in the 250 μM IPTG condition in the plot (thus removing the error from that point). From Figure 3, the ratio between the inverse of the T7 RNA production rate for infinite T7 RNAP numbers in the cells (R^{-1}=0) equals 0.82 at 43°C, 0.81 at 37°C, and 0.21 at 20°C. These numbers correspond also to the ratio between open complex formation (τ_{oc}) and mean transcription interval (Δt), described in Table I (Methods).

Next, from the ratio (τ_{oc} /Δt), we calculated the rate of open complex formation (k_{oc}). Given the value of k_{oc}, we can find the values of k_{cc} and K to achieve the same mean and noise (with 95% accuracy) of the transcription intervals shown in Table I (Methods). Results are shown in Table II. Shown are the rate of open complex formation (k_{oc}), the reversibility of the closed complex formation (K) and the rate of closed complex formation ($R.k_{cc}$), given the empirical values of the ratio ($\tau_{oc}/\Delta t$) extracted from Figure 3.

TABLE II. ESTIMATION OF THE KINETIC RATES OF THE T7 PROMOTER INITIATION PROCESS VERSUS TEMPERATURE.

T (°C)	$\tau_{oc}/\Delta t$	k_{oc} (s^{-1})	K	$R.k_{cc}$ (s^{-1})
43	0.822	263^{-1}	> 2.00	$> 20^{-1}$
37	0.808	284^{-1}	1.2±0.5	$(32±8)^{-1}$
20	0.206	107^{-1}	<0.11	$(351±77)^{-1}$

From Table II, the formation of the open complex, following the T7 RNAP commitment to the closed complex, is faster at 20°C and slower at 37°C and 43°C. This seemingly counterintuitive response suggests that, at higher temperatures, the open complex may be less stable and that, has a consequence, it becomes more reversible to the previous state rather than to committing to the elongation complex.

Namely, the reversibility of the closed complex (K) increases with increasing temperature. At 43°C, the closed complex appears to be highly unstable and T7 RNAP likely binds and unbinds from the T7 promoter several times before being able to form a stable open complex, thus reducing the rate of RNA production. At 37°C, the closed complex appears to be more stable, with a ~50% chance of the RNAP unbinding. At 20°C, the chance of this RNAP unbinding appears to become negligible, likely due to both more stable closed complex formation and faster rate of open complex formation.

Finally, the rate of closed complex formation ($R.k_{cc}$) becomes slower with decreasing temperature. It should be noted that this rate is highly dependent on lacUV5's strength (which determines R) and therefore is not a property of the natural system. In the future, direct measurements of the relative T7 RNAP protein levels should help revealing the temperature dependence of the closed complex (k_{cc}) of this system.

IV. CONCLUSION AND FUTURE WORK

The T7 bacteriophage has only the lytic cycle. Once infecting an *E. coli* cell, its genes transcription is activated and proceeds uninterruptedly until the replication of the viral DNA it achieved [2]. The dynamics of transcription (mean and noise), should therefore play a key role in the success rate of this process. Consequently, for this process to be successful in temperature-fluctuating environments, the transcription process ought itself to be robust to a wide range of temperature conditions.

To assess this robustness, we observed for the first time the *in vivo* transcription initiation kinetics of the T7 promoter at the single RNA level as a function of temperature. Our results suggest that, as temperature decreases, both the mean rate of RNA production and the noise in this process decrease. This somewhat surprising result appears to be made possible by the stabilization of the closed complex formation at lower temperatures.

Our results are, to some extent, similar to those reported for a natural promoter of *E. coli*, P_{tetA}. Namely, its initiation kinetics is also sub-Poissonian, with two rate-limiting steps, the closed and the open complex, whose duration is temperature dependent [8]. However, in P_{tetA}, the noise increases for decreasing temperature.

At the moment, it is unknown what specificities the configuration or composition of the T7 promoter allow this opposite behavior, but this knowledge should be of value to the future engineering of synthetic genes and circuits with robust behaviors at low temperature conditions. From the evolutionary point of view, such noise reduction with lowering temperatures could be associated with the need of the virus for balancing the numbers of phages and capsid proteins more accurately as their total numbers are reduced due to the lowering of the mean production rate [2][7].

In this regard, note from Table I that the relative increase in the interval between RNA productions as temperature decreases from 37°C to 20°C is smaller than the decrease in T7 RNAP numbers (which here are artificially controlled by the LacUV5 promoter). This suggests that, provided a constant number of T7 RNAP for changing temperature, the mean rate of transcription from the T7 promoter will not decrease heavily for decreasing temperature in this range.

In the future, we will employ the system used here and, among other, make use of different promoters controlling the expression of the T7 RNAP so as to, by comparing the various results, isolate the effects of temperature on the T7 promoter alone. Also, we observed that this system is capable of quickly depleting cells from MS2-GFP. This may allow studying the kinetics of binding and unbinding of MS2-GFP to the target RNA as a function of temperature, which might give insights, e.g., on the process by which viral RNAs are protected from the host degradation mechanisms.

ACKNOWLEDGMENT

Academy of Finland [257603, A.R.]; Vilho, Yrjö and Kalle Väisälä fund of Finnish Academy of Science and Letters [S.O.]; TUT President's Graduate Programme

[R.N.], Portuguese Foundation for Science and Technology (FCT/MCTES) [SFRH/BD/88987/2012, LM] and PTDC/BBB-MET/1084/2012, JMF and ASR] and FCT Strategic Program UID/EEA/00066/203 of UNINOVA, CTS [LM and JMF].

REFERENCES

[1] I. Molineux, "The T7 Group," in *The Bacteriophages*, Second Edi., Oxford: Oxford University Press, pp. 277–301, 2005.

[2] J. Dunn, F. Studier, and M. Gottesman, "Complete nucleotide sequence of bacteriophage T7 DNA and the locations of T7 genetic elements," *J. Mol. Biol.*, vol. 166, 1983, pp. 477-535.

[3] J. L. Oakley, R. E. Strothkamp, a H. Sarris, and J. E. Coleman, "T7 RNA polymerase: promoter structure and polymerase binding," *Biochemistry*, vol. 18, no. 3, 1979, pp. 528–37.

[4] W. T. McAllister and A. D. Carter, "Regulation of promoter selection by the bacteriophage T7 RNA polymerase in vitro," *Nucleic Acids Res.*, vol. 8, no. 20, 1980, pp. 4821–4837.

[5] R. Ikeda, A. Lin, and J. Clarke, "Initiation of transcription by T7 RNA polymerase as its natural promoters," *J. Biol. Chem.*,vol. 267, no. 4, 1992, pp. 2640–2649.

[6] R. A. Ikeda, "The efficiency of promoter clearance distinguishes T7 class II and class III promoters," *J. Biol. Chem.*, vol. 267, no. 16, 1992, pp. 11322–11328.

[7] M. De Paepe and F. Taddei, "Viruses' life history: Towards a mechanistic basis of a trade-off between survival and reproduction among phages," *PLoS Biol.*, vol. 4, no. 7, 2006, pp. 1248–1256.

[8] A.-B. Muthukrishnan, M. Kandhavelu, J. Lloyd-Price, F. Kudasov, S. Chowdhury, O. Yli-Harja, and A. S. Ribeiro, "Dynamics of transcription driven by the tetA promoter, one event at a time, in live Escherichia coli cells," *Nucleic Acids Res.*, vol. 40, no. 17, 2012, pp. 8472–8483.

[9] F. Arsène, T. Tomoyasu, and B. Bukau, "The heat shock response of Escherichia coli," *Int. J. Food Microbiol.*, vol. 55, 2000, pp. 3–9.

[10] K. Yamanaka, "Cold shock response in Escherichia coli," *J. Mol. Microbiol.*, vol. 1, 1999, pp. 193–202.

[11] G. M. Skinner, C. G. Baumann, D. M. Quinn, J. E. Molloy, and J. G. Hoggett, "Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. A single-molecule view of the transcription cycle," *J. Biol. Chem.*, vol. 279, no. 5, 2004, pp. 3239–3244.

[12] J. Lloyd-Price, S. Startceva, J. G. Chandraseelan, V. Kandavalli, N. Goncalves, A. Häkkinen, and A. S. Ribeiro, "Dissecting the stochastic transcription initiation process in live Escherichia coli," *DNA Res.*, 2016, in press.

[13] F. W. Studier and B. A. Moffatt, "Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes," *J. Mol. Biol.*, vol. 189, no. 1, 1986, pp. 113–130.

[14] K. Nevo-Dinur, A. Nussbaum-Shochat, S. Ben-Yehuda, and O. Amster-Choder, "Translation-independent localization of mRNA in E. coli," *Science*, vol. 331, no. 6020, 2011, pp. 1081–1084.

[15] I. Golding and E. Cox, "RNA dynamics in live Escherichia coli cells," *Proc. Natl. Acad. Sci,.* vol. 101, no. 31, 2004, pp. 11310-11315.

[16] J. Santinha, L. Martins, A. Häkkinen, J. Lloyd-Price, S. M. D. Oliveira, A. Gupta, T. Annila, A. Mora, A. S. Ribeiro, and J. R. Fonseca, "iCellFusion: Tool for fusion and analysis of live-cell images from time-lapse multimodal microscopy," in *Biomedical image analysis and mining techniques for improved health outcomes, IGI Global*, 2015, pp. 71-99.

[17] A. D. Mora, P. M. Vieira, A. Manivannan, and J. M. Fonseca, "Automated drusen detection in retinal images using analytical modelling algorithms," *Biomed. Eng. Online*, vol. 10, no. 1, 2011, p. 59.

[18] A. Häkkinen, M. Kandhavelu, S. Garasto, and A. S. Ribeiro, "Estimation of fluorescence-tagged RNA numbers from spot intensities.," *Bioinformatics*, 2014, pp. 1–8.

[19] J. Mäkelä, M. Kandhavelu, S. M. D. Oliveira, J. G. Chandraseelan, J. Lloyd-Price, J. Peltonen, O. Yli-Harja, and A. S. Ribeiro, "In vivo single-molecule kinetics of activation and subsequent activity of the arabinose promoter," *Nucleic Acids Res.*, vol. 41, no. 13, 2013, pp. 6544–6552.

[20] H. Tran, S. M. D. Oliveira, N. Goncalves, and A. S. Ribeiro, "Kinetics of the cellular intake of a gene expression inducer at high concentrations," *Mol. Biosyst.*, vol. 11, no. 9, 2015, pp. 2579–2587.

[21] A. Häkkinen and A. S. Ribeiro, "Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data," *Bioinformatics*, vol. 31, no. 1, 2015, pp. 69–75.

[22] K. J. Livak and T. D. Schmittgen, "Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method," *Methods*, vol. 25, no. 4, 2001, pp. 402–408.

[23] W. McClure, "Rate-limiting steps in RNA chain initiation," *Proc. Natl. Acad. Sci.*, vol. 77, no. 10, 1980, pp. 5634–5638.

[24] H. Buc and W. R. McClure, "Kinetics of open complex formation between Escherichia coli RNA polymerase and the lac UV5 promoter. Evidence for a sequential mechanism involving three steps," *Biochemistry*, vol. 24, no. 11, 1985, pp. 2712–2723.

[25] W. R. Mcclure, "Mechanism and control of transcription initiation in prokaryotes," *Annu. Rev. Biochem.*, vol. 54, 1985, pp. 171–204.

[26] D. K. Hawley and W. R. Mcclure, "Nucleic compilation and analysis of Escherichia coli promoter DNA sequences," *Nucleic Acids Res.*, vol. 11, no. 8, 1983, pp. 2237–2255.

[27] R. Lutz and H. Bujard, "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.*, vol. 25, no. 6, 1997, pp. 1203–1210.

[28] M. Kandhavelu, J. Lloyd-Price, A. Gupta, A.-B. Muthukrishnan, O. Yli-Harja, and A. S. Ribeiro, "Regulation of mean and noise of the in vivo kinetics of transcription under the control of the lac/ara-1 promoter," *FEBS Lett.*, vol. 586, no. 21, 2012, pp. 3870–3875.

[29] E. Bertrand-Burggraf, J. F. Lefèvre, and M. Daune, "A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322," *Nucleic Acids Res.*, vol. 12, no. 3, 1984, pp. 1697–1706.

[30] L. M. Hsu, "Promoter clearance and escape in prokaryotes," *Biochim. Biophys. Acta - Gene Struct. Expr.*, vol. 1577, no. 2, 2002, pp. 191–207.

[31] S. L. Heilman-Miller and S. A. Woodson, "Effect of transcription on folding of the Tetrahymena ribozyme," *RNA*, vol. 9, no. 6, 2003, pp. 722–733.

[32] C. Zimmer, A. Häkkinen, and A. S. Ribeiro, "Estimation of kinetic parameters of transcription from temporal single-RNA measurements," *Math. Biosci.*, vol. 271, 2015, pp. 146–153.

[33] M. Kandhavelu, H. Mannerström, A. Gupta, A. Häkkinen,

J. Lloyd-Price, O. Yli-Harja, and A. S. Ribeiro, “In vivo kinetics of transcription initiation of the lar promoter in Escherichia coli. Evidence for a sequential mechanism with two rate-limiting steps,” *BMC Syst. Biol.*, vol. 5, no. 1, 2011, p. 149.

[34] M. Krystek and M. Anton, “A weighted total least-squares algorithm for fitting a straight line,” *Meas. Sci. Technol.*, vol. 19, no. 7, 2008, p. 079801.

 ISBN: 978-1-61208-488-6

PUBLICATION III

N.S.M. Goncalves, S.M.D. Oliveira, V.K. Kandavalli, J.M. Fonseca and A.S. Ribeiro "Temperature Dependence of Leakiness of Transcription Repression Mechanisms of *Escherichia coli*", *Proceedings of the Computational Methods in Systems Biology* (CMSB 2016), September 21-23, Cambridge, United Kingdom, 2016.

Temperature Dependence of Leakiness of Transcription Repression Mechanisms of *Escherichia coli*

Nadia Goncalves[1], Samuel M.D. Oliveira[1], Vinodh K. Kandavalli[1], Jose M. Fonseca[2] 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.
andre.ribeiro@tut.fi

Abstract. In organisms such as *Escherichia coli*, transcription repression plays a key role, as the untimely production of RNAs can significantly perturb cellular functioning. While repression mechanisms have, thus, been under much selective pressure, their failure rates are non-negligible, causing repressed genes to have leaky transcription events. Here, we measured with single RNA-molecule sensitivity the rate of leakiness of the promoter *lacO3O1* under optimal and sub-optimal temperature conditions in live, individual cells. After establishing that the rate of leaky RNA production is temperature dependent, we assume a standard model of transcription and perform additional measurements, in order to dissect which model parameters are most temperature-dependent. First, we show from empirical data that RNA polymerase numbers and k_t, the rate at which an RNA polymerase (RNAp) finds a free *lacO3O1* promoter and executes transcription, are weakly temperature dependent. Consequently, the parameter of repression efficiency (dependent on the number of repressors and binding and unbinding rates of repressors to the promoter) is found to be heavily temperature dependent. We conclude that the degree of leakiness of the *lacO3O1* promoter increases at low temperatures, due to the strong temperature-dependence of its repression mechanism.

Keywords: • Transcription • Repression • Leakiness • lacO3O1 • MS2-GFP RNA detection • Time-lapse Confocal Microscopy.

1 Introduction

Repression mechanisms of gene expression play (at least) two major roles that are essential for cellular functioning. First, they allow establishing when a gene is expressed, which is of importance to the cellular global functioning (see e.g. [1]). Success in this task can be vital, if the resulting protein plays a role that goes against what needs to be executed at a particular time [2]. Second, they contribute to the difference in transcription rates of a gene under full repression and under full induction, which will define the degree to which the numbers of the resulting protein will vary from

low to high [3]. This width can be of importance since, due to noise in gene expression [4][5], in general, e.g. the wider it is, the more precise will be the functioning of a mechanism relying on the crossing of a threshold between the two 'states'.

Due to their importance for the functioning of gene regulatory networks, repression mechanisms have been under much selective pressure and, currently, multiple forms of repression mechanisms exist in any living cell, including bacteria (for a review see [6]). However, these highly evolved mechanisms remain subject to failures, as it is possible to observe the occurrence of transcription events, e.g. even in the absence of inducers. These transcription events occurring in the absence of favorable conditions for gene expression (e.g. in the absence of activator molecules), are usually referred to as 'leaky' transcription events, and the rate at which they occur defines the leakiness of the promoter regulating gene expression.

While there are several studies on transcription leakiness (see e.g. [7]), little is known about the extent to which sub-optimal temperature conditions may hamper repression mechanisms and, thus, the leakiness of a gene. This is of importance for better understanding not only how gene regulatory mechanisms function (which usually requires analyzing how they respond to changes in internal variables and external conditions), but also to learn about their limitations and, ultimately, in order to explain why and how bacteria alter the behavior so widely as a function of temperature.

Here, we investigate this question for one of the best known repression systems, namely, the one regulating the lacI-lac operon, which is present in the chromosome of *Escherichia coli*. For this, we performed *in vivo* single-cell, single-RNA measurements of the transcription dynamics of the *lacO3O1* promoter, in the absence of inducers, and analyzed the results with the support of a model and parallel empirical data that allowed dissecting the contribution of the various players in transcription to the changes in leakiness with temperature.

2 Materials and Methods

To conduct this study, we used the *lacO3O1* promoter, known to be functional at optimal temperatures, and whose transcription dynamics requires an external inducer (e.g. Isopropyl β-D-1-thiogalactopyranoside, IPTG) [8]. We placed this promoter on a single-copy F-plasmid, controlling the expression of a synthetic RNA with multiple binding sites for MS2-GFP. In addition, we also inserted in the cells a multi-copy plasmid coding for MS2-GFP. The *E. coli* strain BW25113 was used because it expresses the repressor of *lacO3O1*, LacI, in similar quantities to the natural system [9].

2.1 Cells, Plasmids and Growth

The strain *Escherichia coli* BW25113 (*lacI*$^{+}$ *rrnB*$_{T14}$ Δ*lacZ*$_{WJ16}$ *hsdR514* Δ*araBAD*$_{AH33}$ Δ*rhaBAD*$_{LD78}$) [10] from Keio Collection (Japan), was used to transfer and express the reporter and target genes. The target plasmid, a single copy F-plasmid, pBELO-P$_{lacO3O1}$-mCherry-48BS (constructed for this work), was inserted into the host strain. It contains the *lacO3O1* promoter and the coding sequence of a red fluorescent

protein (mCherry), followed by the sequence of the target RNA, an array of 48 binding sites (48bs) for the viral coat protein (MS2). This promoter was obtained from the native *lac* promoter, but the O2 repressor binding site downstream from the start site has been removed. Instead, the start site is followed by DNA coding for mCherry.

The reporter plasmid, pZA25-P_{BAD}-MS2-GFP [11] (a kind gift from Orna Amster-Choder, Hebrew University of Jerusalem, Israel) was also inserted into the host strain. The reporter plasmid carries the reporter gene, *ms2-gfp*, under the control of the *BAD* promoter. This reporter gene encodes for the fusion protein MS2-GFP, which binds to target RNA sequences. The presence of 48 copies of the MS2 binding sequence per target RNA, and the flooding of MS2-GFP proteins in the cytoplasm, makes a single RNA to appear as a bright spot in fluorescence microscope images, as soon as it is transcribed [12] (Fig. 1).

E. coli BW25113 cells were streaked from -80°C glycerol stocks on LB medium agar plates, containing 34 µg/ml of Chloramphenicol and 35 µg/ml of Kanamycin (Sigma-Aldrich, USA) and incubated overnight at 37°C. From the plate, a colony was picked and cultured overnight at 30°C, with aeration and shaking (250 rpm) in LB medium, supplemented with the appropriate concentrations of Chloramphenicol and Kanamycin. From the overnight culture, cells were diluted in fresh M9 medium supplemented with Chloramphenicol, Kanamycin and Glycerol (0.4 % final concentration) (Sigma-Aldrich, USA), with an initial optical density (OD_{600}) of 0.05, and allowed to grow at 37°C until reaching an OD_{600} of 0.3. Cells from the 24°C group were then incubated for 1 hour at the respective temperature. To induce the production of MS2-GFP proteins, 0.4 % of L-arabinose (Sigma-Aldrich, USA) was added to the culture, and incubated at 24°C or 37°C for at least 30 minutes. To induce the production of the target RNA, cells were then incubated at the appropriate temperature, with 1 mM of IPTG (Sigma-Aldrich, USA) for 2 hours, while no IPTG was added for the non-induced group. Time-lapse measurements were performed to measure RNA production when the target is fully-induced (1 mM IPTG). For that, target induction was continued during image acquisition (see 'Microscopy' section).

To measure RNA polymerase (RNAp) intracellular concentrations, we used *E. coli* strain RL1314 (a kind gift from Robert Landick, University of Wisconsin-Madison, USA), carrying RNAp fluorescently tagged with GFP [13]. Starting from one colony, RL1314 cells were cultured overnight at 30°C with aeration (250 rpm), in LB medium supplemented with 35 µg/ml of Kanamycin. A pre-culture was prepared, by diluting the overnight culture in fresh M9 medium, supplemented with 0.4 % of Glycerol and 35 µg/ml of Kanamycin, to an initial OD_{600} of 0.05. Cells were then incubated at 37°C, 250 rpm, until reaching an OD_{600} of 0.3. At this point, cells from the group 37°C were prepared for imaging, while cells at 24°C were still incubated at the respective temperature for 1 hour prior to image acquisition.

2.2 Microscopy

Cells were visualized by a Nikon Eclipse (Ti-E, Nikon) inverted microscope equipped with a 100x Apo TIRF (1.49 NA, oil) objective. Confocal images were taken by a C2+ (Nikon) confocal laser-scanning system. To visualize MS2-GFP-RNA

'spots', we used a 488 nm laser (Melles-Griot) and an emission filter (HQ514/30, Nikon). Phase contrast images were attained by an external phase contrast system and CCD camera (DS-Fi2, Nikon). Images were obtained by Nikon Nis-Elements software.

For measurements of cell populations, fluorescence and phase contrast images were taken once and simultaneously. For time-lapse measurements of live cells, fluorescence images were taken once per minute while phase contrast images were taken, simultaneously, once every 5 minutes, for at least 4 hours. In both measurements, 5 µl of cell culture were placed on, respectively, 2% and 2.5% agarose gel pads of M9 medium, supplemented with 0.4% of Glycerol, 0.4% of L-Arabinose and 1 mM IPTG, and kept in between a microscope slide and a cover slip. In time series imaging, a peristaltic pump provided continuous flow of fresh M9 media (supplemented with inducers for target and reporter genes in the appropriate concentrations) to the cells, at 0.3 ml/min, through the thermal chamber (CFCS2, Bioptechs, USA). Meanwhile, the temperature was kept as desired during microscopy sessions by a cooling/heating microfluidic system, which provided continuous flow of deionized water at stable temperature (not in contact with the cells) into the thermal imaging chamber.

2.3 Quantitative PCR

Cells were harvested as described above, followed by the addition of RNA protect bacteria reagent (Qiagen, Germany) and immediate mixing by vortexing. Samples were incubated for 5 minutes at room temperature, and then centrifuged at 5000 × g for 10 minutes. The supernatant was discarded and any residual supernatant was removed by inverting the tube once onto a paper towel. The entire RNA content was isolated using the RNeasy kit (Qiagen) according to the instructions of the manufacturer. Samples were quantified using a Nanovue plus spectrophotometer (GE Healthcare Life Sciences) and the quality of the isolated RNA was assessed by measuring the ratio of absorbance at 260 and 280 nm (A260/A280 ratio) of the sample (2.0–2.1). DNaseI treatment was then performed to avoid DNA contamination. The cDNA was synthesized from 1 µg of RNA using the iScript Reverse Transcription Supermix (Biorad, USA). The cDNA templates with a final concentration of 10 ng/µl were added to the qPCR master mix containing iQ SYBR Green supermix (Biorad, USA) with primers for the target and reference genes at a final concentration of 200 nM. The 16S RNA gene was used for internal reference. The primers set for the target RNAs and the reference gene (16S RNA) are as follow: mCherry (Forward: 5' CACCTACAAGGCCAAGAAGC 3', Reverse: 5' TGGTGTA GTCCTCGTT GTGG 3'), 16S RNA (Forward: 5' CGTCAGCTCGTGTTGTGAA 3', Reverse: 5' G GACCGCTGGCAACAAAG 3').

The qPCR experiments were performed by 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 for each cDNA replicate. All reactions were performed in 3 replicates for each condition. No-RT and no-template controls were used to crosscheck non-specific signals and contamination. PCR efficiencies of these reac-

tions were greater than 95%. Data from CFX Manager™ Software was used to calculate the relative gene expression and its standard error using the $2^{-\Delta\Delta CT}$ method [14].

2.4 Image Analysis

Cells were detected from phase contrast images as in [15]. Phase contrast and fluorescence images were aligned using cross-correlation maximization and then cells were automatically segmented from phase contrast images using MAMLE [16], followed by manual correction. In time-series imaging, cell lineages were determined by CellAging [17], from overlapping areas of segments between consecutive images.

Next, cell segmentation results were aligned with fluorescence images by manually selecting 5-7 landmarks in both images, and using thin-plate spline interpolation for the registration transform [18]. After, spots and their intensities were detected from fluorescence images by a Gaussian surface-fitting algorithm [19]. Namely, the intensity of each cell is fit with a surface representing the cellular background intensity, which is subtracted to obtain the foreground intensity. The foreground intensity is fit with a set of Gaussian surfaces, whose volume represents the total spot intensity.

Example images of a cell over time, and results of cell segmentation and detection of spots (bright spots in the green fluorescent channel) are shown in Figure 1.

Finally, the number of RNA molecules produced in individual cells and their corresponding production rates were obtained. Since MS2-GFP-tagged RNA's lifetime is much longer than cell division times [20][21], the cellular foreground intensity is expected to always increase (by 'jumps'), with a jump in intensity corresponding to the appearance of a new tagged RNA. The number of RNAs in each cell was estimated from the individual spot intensities by a manual RNA rounding method [21]. In this, the expert selects the first and second peaks of spot intensity distribution that likely correspond to the appearance of one and two RNAs in the cell, respectively, as the intensities are expected to cluster around multiples of the average single RNA intensity. Using these two peaks as reference, the method converts spot intensities into RNA numbers for all cells by subtracting the offset and rounding.

Finally, RNAp abundance in each cell was quantified from the background corrected mean pixel intensity as obtained from the fluorescence images.

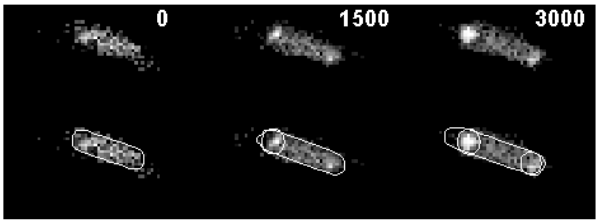


Fig. 1. MS2-GFP tagged RNAs in an *E. coli* cell over time. Unprocessed frames (top) along with the segmented cell and RNA spots (bottom). The moment images were captured is shown at the top of each frame.

2.5 Effects of cell division on mean RNA numbers

To account for cell division effects on the mean RNA number per cell at a given moment, we calculated the rate of RNA production normalized by the rate of RNA dilution caused by cell division. We do not account for RNA degradation, as the binding by multiple MS2-GFP molecules makes the tagged RNAs virtually immortal, and their fluorescence levels constant, for the duration of the measurements. This was shown by measurements of the dissociation rate of MS2 coat proteins from their RNA binding sites (on the order of several hours [22]), and by measurements of the lifetimes of fluorescence of MS2-GFP tagged RNAs kept under observation for more than 2 hours [12][20][21][24].

In accordance with the direct method [23], the cell mean division times (*Div*) between two time moments (t_1 and t_2) can be estimated from OD measurements of cell cultures in the exponential growth phase by:

$$Div = \frac{(t_2 - t_1) \times \ln(2)}{(\ln OD_2 - \ln OD_1)} \tag{1}$$

One can then estimate the expected rate of RNA dilution (k_d) due to cell division:

$$k_d = \frac{\ln(2)}{Div} = \frac{(\ln OD_2 - \ln OD_1)}{(t_2 - t_1)} \tag{2}$$

Given a certain mean number of RNAs per cell at a given time (RNA(t)), it follows that the mean rate of RNA production, λ_{RNA}, necessary for observing those numbers is:

$$\lambda_{RNA} = \frac{k_d \times RNA(t)}{1 - e^{-k_d \times t}} \tag{3}$$

3 Results and Conclusions

3.1 Control Experiments

To study the leakiness of *lacO3O1* and assess the effects of changing temperature on this phenomenon, we first conducted a control experiment, to determine if MS2-GFP proteins can form significant clusters in the absence of target RNA that could be mistaken by MS2-GFP-RNA complexes or if the image analysis wrongly detects false RNA 'spots' where none are visible to a human observer. For this, we analyzed images with cells containing the reporter system (producing MS2-GFP fluorescent proteins), but lacking the target system (coding for the RNA target for MS2-GFP).

We found that the automatic spot detection method rarely detected artefacts in the image (0.02 false positives per cell, Table 1). Upon visual inspection, these were found to be due to errors (false positives) of the image analysis algorithm rather than real fluorescent spots due to, e.g. MS2-GFP clustering in the absence of target RNA.

Overall, we conclude that the RNA tagging method combined with the spot detection methods used here, produces a negligible number of 'false positive' RNA spots.

3.2 Induction and leakiness as a function of temperature

First, we measured mean RNA numbers in cells at 37°C as a function of the inducer (IPTG) concentration by qPCR. Results are shown in Fig. 2.

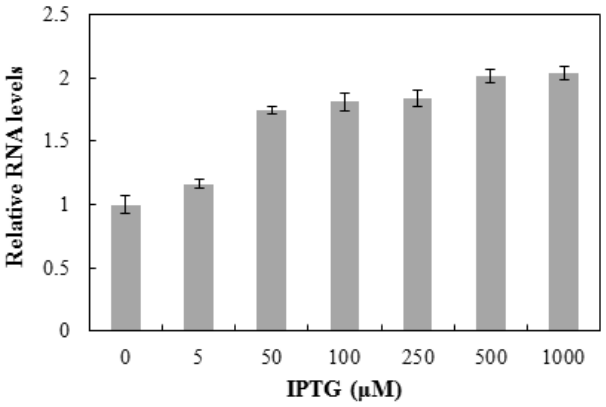


Fig. 2. Relative RNA levels under the control of *lacO3O1* promoter as a function of IPTG concentration, as measured by qPCR. Values are relative to the condition of 0 μM IPTG. The bars show the standard errors in RNA numbers, obtained from 3 biological replicates, per condition.

From Fig. 2, *lacO3O1* promoter is responsive, while not strongly, to changes in the concentration of IPTG. Given these results, from here onwards, we assume that for 1 mM IPTG induction level, the promoter is fully induced (i.e. we assume that all repressor molecules, LacI, are rendered inactive), similar to the native Lac gene, and in agreement with previous studies [2][8].

Next, we compared the kinetics of transcription of *lacO3O1* promoter when not induced, in live individual cells at 24°C and 37°C. Measurements were performed from microscope images, at the single cell level (Methods).

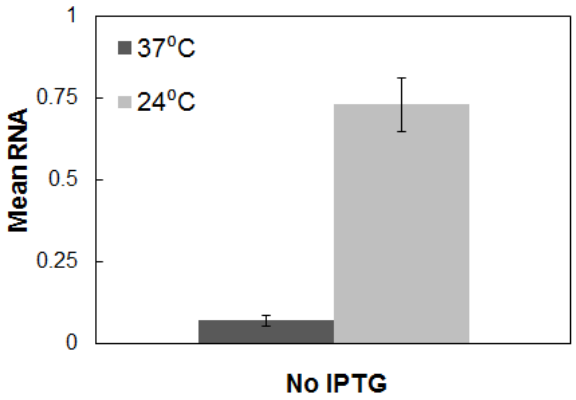


Fig. 3. RNA numbers resultant from the transcription activity of *lacO3O1* promoter as a function of IPTG concentration and temperature, obtained from microscope images. *E. coli* BW25113 cells containing *lacO3O1* promoter controlling the production of RNA target for MS2-GFP, were grown in liquid M9 medium at 37°C, until reaching mid-log phase. Cells were then incubated at 24°C or 37°C, before being imaged, as soon as they were placed in the microscope. Single-cell RNA numbers were extracted from the images. The bars show the standard uncertainties in RNA numbers, obtained from 3 technical replicates per condition.

Next, to obtain the rates of RNA production at the single cell level, we also measured the cell growth rates at each temperature (Methods). Results are shown in Fig. 4.

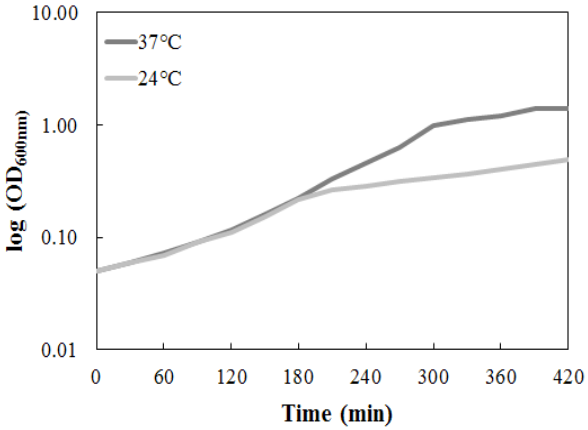


Fig. 4. Growth curves (OD_{600}, measured with an Ultrospec 10 Cell Density Meter, Amersham Bioscience) of cells at 37°C and 24°C. Cells were incubated at 37°C at 250 rpm until reaching the mid-log phase (first ~3 h), and then placed at the appropriate temperature for 4 hours. During these 4 hours, the OD_{600} was measured every 30

minutes. The slopes of the fits were obtained from this time period, and correspond to the cell doubling times (Table 1).

From the data in Fig. 3 along with the cell growth rates in Fig. 4, we estimated the rate of RNA dilution due to cell division, along with the number of RNAs in each cell at the time of the imaging. From these, and equation (3), we estimated the 'actual' rate of *leaky* (non-induced) RNA production, $\lambda_{RNA}^{\mathrm{Rep}}$. In addition, we obtained the leakiness ratio at 24°C, relative to 37°C. These quantities are shown in Table 1.

Table 1. Mean number of tagged RNAs produced under the control of *lacO3O1* promoter in non-induced cells at different temperatures while in liquid culture, as observed by confocal microscopy. Shown are the number of cells observed, the mean RNA numbers, the estimated rate of RNA dilution due to cell division, the 'actual' mean rate of leaky RNA production accounting for dilution, and the ratio of leakiness rate at 24°C relative to the control condition (37°C). In the case of the control experiment, cells lack the coding ability of the RNA target for MS2-GFP and thus, in this, any 'detected' RNA molecules are false positives.

Condition	No. *cells*	*Mean RNA no. per cell*	*Rate of RNA Dilution,* k_d (h^{-1})	$\lambda_{RNA}^{\mathrm{Rep}}$ (h^{-1})	Ratio of Leakiness *(relative to 37°C)*
Control, 37°C	159	0.02	-	-	-
37°C	526	0.07	0.31	0.047	-
24°C	365	0.73	0.09	0.40	8.52

From Table 1, first, as expected, cell division rates are maximal at 37°C, which results in maximal rate of RNA dilution due to cell division. Accounting for this (Methods), we find that the 'actual rate' of leakiness is approximately 8.5 times higher at 24°C than at 37°C. We conclude that the leakiness of *lacO3O1* promoter is heavily temperature dependent, being much higher at lower temperatures, and that this is not an artifact caused by differences in cell division rates with temperature.

This allows concluding that the strength, or 'efficiency', of repression by LacI is temperature dependent, being much weaker at lower temperatures.

3.3 Possible sources of the temperature dependence of leakiness

Next, we searched for underlying causes for the strong temperature-dependence of the leakiness phenomenon. For this, we assume the following model of transcription and its repression mechanism:

$$\mathrm{Pro}_{OFF}.\mathrm{Rep} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} \mathrm{Pro}_{ON} + \mathrm{Rep} \tag{4}$$

$$\mathrm{Pro}_{ON} + \mathrm{RNAp} \xrightarrow{k_t} \mathrm{Pro}_{ON} + \mathrm{RNAp} + \mathrm{RNA} \tag{5}$$

Reaction (4) models the repression mechanism, where 'Pro' is the promoter (which can be 'ON', i.e. available for transcription, or 'OFF' due to being blocked by a repressor) and 'Rep' is a repressor (LacI tetramer) [25], while k_{on} and k_{off} are the rates of binding and unbinding of the repressor to the promoter. Meanwhile, reaction (5) models transcription as a single-step process, where k_t is the rate at which an RNA polymerase (RNAp) finds the promoter and, once doing so, produces an RNA. While much evidence shows that this process in general has, *in vivo*, two tangible rate-limiting steps [18][24][26], this simplification is made here as it does not affect our conclusions.

Given this model, the average rate of RNA production, λ_{RNA}, should equal:

$$\lambda_{RNA} = \mathrm{P}_{ON} \times \left(k_t \times RNAp \right) \tag{6}$$

where P_{ON} is the probability that the promoter is "ON". From (4) and (5), this quantity should equal:

$$P_{ON} = \frac{k_{on}}{\mathrm{Rep} \times k_{off} + k_{on}} \tag{7}$$

Consequently, from (6) and (7):

$$\lambda_{RNA} = \left(\frac{k_{on}}{Rep \times k_{off} + k_{on}} \right) \times RNAp \times k_t \tag{8}$$

Equation (8) informs on the factors determining the mean rate of RNA production based on the model. These are the RNAp concentration in the cells, the transcription rate of a free promoter, k_t, and, finally, the fraction of time that the promoter is free from repressors, which depends on the number of repressors (Rep) and their binding (k_{off}) and unbinding (k_{on}) rates to the promoter. In this regard, since our empirical data does not allow dissecting the parameters associated with the repression mechanism (see below), we define β as the repression strength, which equals the inverse of the time that the promoter is free from repressors:

$$\beta = \frac{Rep \times k_{off} + k_{on}}{k_{on}} \tag{9}$$

From (9), assuming that virtually all repressor molecules can be made inert by activator molecules, IPTG, when under full induction, then, in those conditions β will equal 1. In that case, from (8) and (9), we can define the average rate of RNA production under full induction (λ_{RNA}^{Act}) and full repression ($\lambda_{RNA}^{\mathrm{Rep}}$), to equal, respectively:

$$\lambda_{RNA}^{Act} \cong RNAp \times k_t \tag{10}$$

$$\lambda_{RNA}^{\mathrm{Rep}} \cong \beta^{-1} \times RNAp \times k_t \tag{11}$$

Equation (10) informs on the factors determining the mean RNA production rate of an active (free from repressors) promoter. This is informative to us, as it implies that, if we obtain empirical values for λ_{RNA}^{Act} (Methods and Table 3) of a fully active promoter, and RNAp concentrations (or at least, a relative concentration, see Methods), than we can calculate the value of k_t. Namely, from (10) one can write:

$$\frac{k_t}{k_{t(control)}} = \frac{RNAp_{control}}{RNAp} \times \frac{\lambda_{RNA}^{Act}}{\lambda_{RNA}^{Act}(control)} \tag{12}$$

Having obtained k_t from empirical data on fully active (induced) promoters, next, one can determine which factor(s) (RNAp, k_t, or β) most contribute(s) to the measured differences with temperature in the leakiness of non-induced promoters (Table 1). For this, we need to determine which factors in (11) most change with temperature. For that, from (11), one can write:

$$\frac{\beta}{\beta_{control}} = \frac{RNAp}{RNAp_{control}} \times \frac{k_t}{k_{t(control)}} \times \left(\frac{\lambda_{RNA}^{\mathrm{Re}\,p}}{\lambda_{RNA}^{\mathrm{Re}\,p}(control)} \right)^{-1} \tag{13}$$

Having (12) solved from empirical data on RNAp concentrations and λ_{RNA}^{Act}, (13) can be solved by that solution, along with empirical values of the other two terms of the multiplication. With this aim, next, we performed measurements of RNAp concentrations and of RNA production rates in cells where *lacO3O1* is fully induced, at different temperatures, so as to estimate the values of λ_{RNA}^{Act} in each condition.

3.4 Empirical values of RNAp concentrations

Using confocal microscopy, we measured relative RNAp concentrations of cells of the strain RL1314 expressing fluorescently tagged RpoC [13] under the two temperature conditions. Results are shown in Table 2.

Table 2. Measurements of the mean intracellular RNAp concentrations at 24°C and 37°C from microscope images. Shown are the number of cells observed, and the mean RNAp:GFP fluorescence levels in arbitrary units (a.u.), at each temperature. Also shown are the RNAp levels at 24°C relative to 37°C. After growing cells (RL1314, RpoC::GFP strain) in liquid M9 medium at 37°C, cells were incubated at the appropriate temperature for 1 hour. Next, cells were placed under the microscope and images were collected to extract RNAp fluorescence levels.

Condition	No. cells	Mean RNAp level (a.u.)	RNAp level Ratio (to 37°C)
37°C	139	0.98	-
24°C	262	1.17	1.19

From Table 2, RNAp concentrations are mildly higher at the lower temperature. While this is expected to contribute to the higher leakiness, this difference is too small to fully explain the observed differences in leakiness shown in Table 1.

3.5 Empirical values of RNA production rates of fully induced *lacO3O1*

The rate constant k_t, which is the rate at which one RNAp finds a free promoter and executes transcription, cannot be measured directly *in vivo*. However, from time-lapse fluorescence microscopy of cells expressing RNA target for MS2-GFP under the control of a single *lacO3O1* promoter and MS2-GFP reporters, one can measure the mean RNA production rate under full induction (λ_{RNA}^{Act}), which allow estimating k_t using equation (11).

For this, we performed microscopy measurements of RNA production rates in cells with *lacO3O1* promoter fully induced (1000 µM IPTG). Unlike when measuring RNA production rates due to leakiness, here, as RNA production rates are much higher, approximations based on division times are no longer as reliable. Instead, first, we performed time-lapse microscopy measurements (4 hours long, with images taken every minute, using a temperature control chamber, see Methods). Then, we performed cell segmentation throughout the time series, to identify when cells are first born and then divide (Methods). RNA production rates of active promoters were obtained from the initial and final number of RNA molecules in each tracked cell, and averaged over all cells. As such, the extracted quantities are independent of cell division rates.

Note that the MS2-GFP tagging system of RNA molecules is well-known to lose reliability in the counting of RNAs from a single cell for large numbers of target RNAs (larger than ~7 [21]). However, here this does not pose a problem since, even under full induction, virtually no cell produced close to such amounts of target RNAs during the measurement period. Results are shown in Table 3.

Table 3. RNA production rates under the control of *lacO3O1* promoter fully induced with 1000 µM IPTG at different temperatures, as observed by confocal microscopy. Shown are the number of cells observed, the rate of RNA production in individual cells, the ratio of λ_{RNA}^{Act} relative to the control (37°C), and the ratio of k_t relative to the control (37°C).

Condition	No. cells	λ_{RNA}^{Act} (h^{-1})	Ratio $\lambda_{RNA}^{Act} / \lambda_{RNA}^{Act}$ (control)	Ratio k_t / k_t (control)
37°C (control)	116	0.42	-	-
24°C	106	0.26	0.62	0.52

From Table 3, under full induction (λ_{RNA}^{Act}), RNA production under the control of *lacO3O1* promoter is higher at 37°C.

3.6 Dissection of causes for increased leakiness in transcription at sub-optimal temperatures

Having collected empirical values of the relative changes in RNAp numbers, in k_t rates, and in λ_{RNA}^{Act} rates from cells at 24°C when compared to the control condition (37°C), we next examine the contribution of each factor in equation (13) to the measured differences with temperature in the rate of leakiness of *lacO3O1* promoter. For this, we use the data in Table 1 (λ_{RNA}^{Rep}), Table 2 (RNAp numbers), and Table 3 (k_t):

$$\frac{\beta_{24°C}}{\beta_{control}} = \frac{RNAp_{24°C}}{RNAp_{control}} \times \frac{k_{t(24°C)}}{k_{t(control)}} \times \left(\frac{\lambda_{RNA}^{Rep}(24°C)}{\lambda_{RNA}^{Rep}(control)} \right)^{-1} = 0.07 \qquad (14)$$

By comparing the degree of change of each of the three parameters (β, RNAp numbers, and k_t) with temperature, we conclude that the major contributor for the change with decreasing temperature in the mean rate of leakiness, λ_{RNA}^{Rep}, of repressed *lacO3O1* promoters, is the strong loss in efficiency of the repression mechanism (β), which changes by an order of magnitude, rather than the weak changes in RNAp numbers (Table 2) and in the rate of active transcription, i.e. k_t (which changes in an opposite fashion, Table 3).

4 Discussion

Little is known about the temperature-dependence of repression mechanisms of transcription initiation. Here, we performed an exploratory study of this temperature-dependence for the *lacO3O1* promoter, based on time-lapse confocal microscopy of single-RNA production dynamics at the single cell level, making use of a specially tailored thermal chamber to allow for a wide range of temperatures while under microscope observation. In addition, we performed measurements of RNA polymerase numbers at the single cell level, qPCR measurements of transcription rates, and standard measurements of cell growth rates under each temperature condition.

We found that, at 37°C, *lacO3O1* promoter exhibits both higher production rates when induced and lower leakiness when non-induced. As temperature is decreased, while the production rate when active decreases, the rate of leakiness increases widely. This increase appears to be caused by the strong-dependence of the repression mechanism, rather than due to changes in the kinetics of transcription or due to alterations in RNAp numbers. Given this, we expect the temperature-dependence of leakiness to vary widely between promoters, as their repressions mechanisms also differ widely.

In the future, three avenues of research should be of interest. One consists of further dissecting the causes for increased leakiness with decreasing temperatures at the level of the repression mechanism. Namely, at the moment, from our study, we cannot establish, e.g., if the increased leakiness at low temperatures is due to alterations in the number of functional repressor molecules, or in the binding and unbinding rate

constants of these molecules to the promoter region. The second avenue of interest is to study what occurs in promoters with different repression mechanisms. We expect that, in certain promoters whose functioning is critical for the well-being of the cell at low temperatures, we will encounter much greater robustness in their degree of leakiness with decreasing temperature. Finally, it should be of interest to explore to what extent this temperature-dependence in leakiness may affect the robustness of small genetic circuits to temperature fluctuations.

References

1. Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S.: Quantifying E. coli Proteome and Transcriptome with Single-Molecule Sensitivity in Single Cells. Sci., 329, 533-538 (2010).
2. Choi, P.J., Cai, L., Frieda, K., Xie, X.S.: A Stochastic Single-Molecule Event Triggers Phenotype Switching of a Bacterial Cell. Sci., 322, 442-446 (2008).
3. Kaern, M., Elston, T., Blake, W.J., Collins, J.J.: Stochasticity in Gene Expression: From Theories to Phenotypes. Nat. Rev. Genet., 6, 451-464 (2005).
4. McAdams, H., Arkin, A.: Stochastic Mechanisms in Gene Expression. Proc. Natl. Acad. Sci., 94, 814-819 (1997).
5. Elowitz, M.B., Levine, A.J., Eric, D.S., Swain, P.S.: Stochastic Gene Expression in a Single Cell. Sci., 297, 1183-1186 (2002).
6. Rojo, F.: Repression of Transcription Initiation in Bacteria. J. of Bact., 181, 2987 (1999).
7. Huang, L., Yuan, Z., Liu, P., Zhou, P.: Effects of Promoter Leakage on Dynamics of Gene Expression. BMC Sys. Biol., 9, 16 (2015)
8. Oehler, S., Eismann, E.R., Kramer, H., Muller-Hill, B.: The Three Operators of the Lac Operon Cooperate in Repression. EMBO Jour., 9, 973-979 (1990).
9. Penumetcha, P., Lau, K., Zhu, X., Davis, K., Eckdahl, T.T., Campbell, A.M.: Improving the Lac System for Synthetic Biology. BIOS, 81, 7-15 (2010).
10. Baba, T., Ara T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., Datsenko, K.A., Tomita, M., Wanner, B.L., Mori, H.: Construction of Escherichia coli K-12 in-frame, Single-Gene Knockout Mutants: The Keio collection. Mol. Syst. Biol., 2, 11 (2006).
11. Nevo-Dinur, K., Nussbaum-Shochat, A., Ben-Yehuda, S., and Amster-Choder, O.: Translation-Independent Localization of mRNA in E. coli. Sci., 331, 1081-1084 (2011).
12. Golding, I. Cox, E.C.: RNA Dynamics in Live Escherichia coli Cells. Proc. Natl. Acad. Sci., 101, 11310-11315 (2004).
13. Bratton, B.P., Mooney, R.A., Weisshaar, J.C.: Spatial Distribution and Diffusive Motion of RNA Polymerase in Live Escherichia coli. J. Bacteriol., 193, 5138-5146 (2011).
14. Livak, K.J., Schmittgen, T.D.: Analysis of Relative Gene Expression Data using Real-Time Quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 25, 402–408 (2001).
15. Gupta, A., Lloyd-Price, J., Oliveira, S.M.D., Yli-Harja, O., Muthukrishnan, A.-B., Ribeiro, A.S.: Robustness of the Division Symmetry in Escherichia coli and Functional Consequences of Symmetry Breaking. Phys. Biol., 11, 066005 (2014).
16. Chowdhury, S., Kandhavelu, M., Yli-Harja, O., Ribeiro, A.S.: Cell Segmentation by Multi-Resolution Analysis and Maximum Likelihood Estimation (MAMLE). BMC Bioinf., 14, S8 (2013).

17. Hakkinen, A., Muthukrishnan, A.B., Mora, A., Fonseca, J.M., Ribeiro, A.S.: CellAging: A Tool to Study Segregation and Partitioning in Division in Cell Lineages of Escherichia coli. Bioinf., 29, 1708–1709 (2013).
18. Lloyd-Price, J., Startceva, S., Kandavalli, V., Chandraseelan, J.G., Goncalves, N.S.M., Oliveira, S.M.D., Hakkinen, A., Ribeiro, A.S.: Dissecting the Stochastic Transcription Initiation Process in Live Escherichia coli. DNA Res., 23, 203-214 (2016).
19. Häkkinen, A., Kandhavelu, M., Garasto, S., Ribeiro, A.S.: Estimation of Fluorescence-Tagged RNA Numbers from Spot Intensities. Bioinf., 30, 1146–1153 (2014).
20. Tran, H., Oliveira, S.M.D., Goncalves, N.S.M, Ribeiro, A.S.: Kinetics of the Cellular Intake of a Gene Expression Inducer at High Concentrations. Mol. Biosyst., 11, 2579–2587 (2015).
21. Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C.: Real-Time Kinetics of Gene Activity in Individual Bacteria. Cell, 123, 1025–1036 (2005).
22. Johansson, H.E., Dertinger, D., LeCuyer, K.A., Behlen, L.S., Greef, C.H., Uhlenbeck, O.C.: A Thermodynamic Analysis of the Sequence-Specific Binding of RNA by Bacteriophage MS2 Coat Protein, Proc. Natl. Acad. Sci., 95, 9244–9244 (1998).
23. Widdel, F.: Theory and Measurement of Bacterial Growth. Basic Pratical Micobiology, 4th Semester, (ed. Widdel F), pp. 11 (2007).
24. Muthukrishnan, A.-B., Kandhavelu, M., Lloyd-Price, J., Kudasov, F., Chowdhury, S., Yli-Harja, O., Ribeiro, A.S.: Dynamics of Transcription Driven by the TetA Promoter, one Event at a Time, in Live Escherichia coli Cells. Nuc. Aci. Res., 40, 12 (2012).
25. Rutkauskas, D., Zhan, H., Matthews, K.S, Pavone, F.S., Vanzi, F.: Tetramer Opening in LacI-Mediated DNA Looping. Proc. Natl. Acad. Sci., 106, 16627-16632 (2009).
26. Kandhavelu, M., Hakkinen, A., Yli-Harja, O., Ribero, A.S.: Single-Molecule Dynamics of Transcription of the Lac Promoter. Phys Biol, 9, 026004 (2012).

PUBLICATION IV

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

Temperature-dependence of the single-cell variability in the kinetics of transcription activation in *Escherichia coli*

Nadia S.M. Goncalves[1], Sofia Startceva[1], Cristina S.D. Palma[1,2], Mohamed N.M. Bahrudeen[1], Samuel M.D. Oliveira[1] and Andre S. Ribeiro[1,2,3,4]

Running Head: Temperature-dependence of transcription activation in *E. coli*

Keywords: Single-cell intake kinetics; gene expression activation times; *Escherichia coli*; critically low temperatures;

Accepted for Publication: 28 November 2017

doi: 10.1088/1478-3975/aa9ddf/meta

Link for the Version of Record of this article:
http://iopscience.iop.org/article/10.1088/1478-3975/aa9ddf/meta

[1] Laboratory of Biosystem Dynamics, BioMediTech Institute and Faculty of Biomedical Sciences and Engineering, Tampere University of Technology, 33101, Tampere, Finland.
[2] CA3 CTS/UNINOVA. Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal.
[3] Multi-scaled Biodata Analysis and Modelling Research Community, Tampere University of Technology, 33101, Tampere, Finland.
[4] Corresponding author. E-mail: andre.ribeiro@tut.fi, Tel: +358408490736.

Abstract

From *in vivo* single-cell, single-RNA measurements of the activation times and subsequent steady-state active transcription kinetics of a single-copy Lac-ara-1 promoter in *Escherichia coli*, we characterize the intake kinetics of the inducer (IPTG) from the media, following temperature shifts. For this, for temperature shifts of various degrees, we obtain the distributions of transcription activation times as well as the distributions of intervals between consecutive RNA productions following activation in individual cells. We then propose a novel methodology that makes use of deconvolution techniques to extract the mean and the variability of the distribution of intake times. We find that cells, following shifts to low temperatures, have higher intake times, although, counter-intuitively, the cell-to-cell variability of these times is lower. We validate the results using a new methodology for direct estimation of mean intake times from measurements of activation times at various inducer concentrations. The results confirm that *E. coli*'s inducer intake times from the environment are significantly higher, following a shift to a sub-optimal temperature. Finally, we provide evidence that this is likely due to the emergence of additional rate-limiting steps in the intake process at low temperatures, explaining the reduced cell-to-cell variability in intake times.

Introduction

RNA and protein numbers differ between cells of monoclonal populations, due to the stochastic nature of the chemical reactions composing gene expression ('intrinsic' noise) [1,2] and the cell-to-cell variability in the numbers of the molecules involved ('extrinsic' noise) [3].

Besides these 'constant' sources of cell-to-cell variability, recent studies have shown that, following the appearance of an inducer of gene expression in the media, there is an additional transient cell-to-cell diversity in RNA and protein numbers of the target gene [4–6], which cannot be explained by the intrinsic and extrinsic noise of active gene expression. This additional source can be strong enough and the transient long enough to affect the phenotypic diversity of cell lineages for generations [4–12].

The origin of this transient phenotypic diversity has been shown to be the noise in the intake time of the inducers, which causes the time for transcription to be activated (following the introduction of the inducers in the media) to differ widely between cells [5]. At the RNA numbers level, this transient diversity can be higher than the diversity caused by the intrinsic and extrinsic noise in active transcription for long periods of time [5].

Similarly to noise in gene expression, noise in intake times has two sources. One is the stochasticity of the intake process, caused by the random nature of the chemical reactions and the membrane crossing processes [2,6]. The other is likely a non-negligible degree of cell-to-cell heterogeneity in the efficiency of the mechanisms involved in the intake of inducers [5]. This heterogeneity can be caused by, among other, cell-to-cell diversity in the number of transmembrane proteins involved in the active uptake of

inducer/repressor molecules [5]. One example is the lactose permease (LacY), which, while being produced by an all-or-nothing system that minimizes cellular heterogeneity, it nevertheless exhibits significant cell-to-cell diversity in numbers, following the appearance of the inducer (e.g. TMG) in the media [13].

As natural environmental conditions fluctuate and many genes in *E. coli* are only activated in specific conditions, cellular heterogeneity in gene expression activation times is expected to affect significantly the phenotypic diversity of cell populations.

One environmental parameter that we expect to have a tangible impact on both the mean and variability of intake times of external inducers and repressors of gene expression is temperature. This assumption originates from the fact that temperature affects not only proteins functionality and numbers in cells [14], but also the physical properties of cell walls, periplasm and cytoplasm (e.g. the cytoplasm's viscosity is temperature dependent [15]), and these variables are expected to affect the kinetics of intake of inducers from the environment.

However, there is yet no direct experimental validation and, as many variables are involved, model-based predictions of the quantitative degree of changes with temperature in inducers intake times and subsequent transcription initiation times are unreliable.

Here, we characterize quantitatively the changes in cell-to-cell variability in gene expression activation times of the Lac-ara-1 promoter and, more importantly, of the intake times of its inducer, Isopropyl β-D-1-thiogalactopyranoside (IPTG), caused by rapid physical changes following temperature shifts.

For this, we use time-lapse microscopy measurements of RNA production at the single-cell, single-RNA level at various temperatures, along with several recently developed techniques [6,14,16], including a new strategy here proposed to dissect the kinetics of the intake process. Our results provide novel information for the understanding of the effects of temperature shifts of bacterial populations at the single-cell level.

Methods

Bacterial strains and plasmids

We use *E. coli* strain DH5α-PRO, generously provided by I. Golding, University of Illinois, U.S.A. The genotype is deoR, endA1, gyrA96, hsdR17(rK^- mK^+), recA1, relA1, supE44, thi-1, Δ(lacZYA-argF)U169, Φ80δlacZΔM15, F^-, λ^-, P_{N25}/tetR, P_{lacIq} /lacI, SpR. The strain contains two genes, *lacI* and *tetR*, constitutively expressed under the control of P_{laci}^{q} and P_{N25} promoters, respectively [17]. Relevantly, the native lac operon (*lacZYA*) is mutated, to prevent production of permease (*lacY*) and activation of the lactose metabolic system [18]. I.e., these cells lack the native positive feedback mechanism involving lactose [6,19].

In addition to this strain, we also use *E. coli* JW0334 strain. The genotype is F^- (Δ (araD-araB)567 ΔlacY784 ΔlacZ4787(::rrnB-3) λ^-rph-1 Δ(rhaD-rhaB)568 hsdR514) [18].

This strain also lacks the ability to produce lacY [18]). Here, we only make use of this strain to show that the changes in the target gene activation time with temperature are, qualitatively, only weakly strain dependent. Unless stated otherwise, measurements are made using DH5α-PRO cells.

Both strains lack the ability to express lacY permease [18], which is responsible for a feedback response to the intake of IPTG, which would result in more complex, time-dependent single-cell intake times, as they would not be solely determined by the induction level and temperature.

Two constructs were added to DH5α-PRO cells: pROTET-K133 with $P_{LtetO-1}$-MS2d-GFP and pIG-BAC, a single-copy plasmid with $P_{Lac-ara-1}$-mRFP1-MS2d-96bs [20] (Figure 1). In the case of JW0334 cells, another reporter is used (P_{RHAM}-MS2d-GFP), as these cells lack the ability to express TetR.

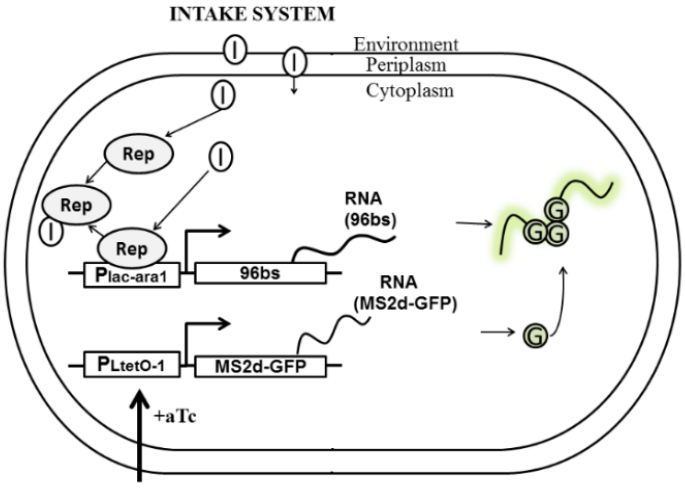


Figure 1. Diagram of the target gene and its RNA tagging system, along with the intake system of inducers of the target gene: IPTG molecules (I) are added to the media and enter the cytoplasm by passing through two membrane layers, with a periplasmic space in between. When in the cytoplasm, they neutralize lacI repressors (R) by forming inducer-repressor complexes (RI). This allows $P_{Lac-ara-1}$ to express RNAs that include an array of 96 MS2d-binding sites. Meanwhile, MS2d-GFP expression is controlled by the $P_{LtetO-1}$ promoter and anhydrotetracycline (aTc). Once produced, each target RNA is rapidly bound by multiple tagging MS2d-GFP proteins (G), and appears as a bright spot, significantly above background fluorescence, under the confocal microscope [6,20]. The tagging provides the RNA a long lifetime, with constant fluorescence, beyond our observation times [6].

Finally, it is noted that previous measurements [6] have shown that, provided full induction of the reporter gene (1 hour) prior to induction of the target gene, any newly produced target RNA molecule becomes ‘fully fluorescent’ (i.e. its RNA MS2-GFP binding

sites become fully occupied) in less than 1 minute. These measurements were conducted in the same strain and media employed here. Given this, and since our microscopy time-lapse images are separated by 1 minute intervals, it is reasonable to assume that, once a new RNA appears, the full occupation of its MS2-GFP binding sites will take less time than the time between two consecutive images. This is agreement with measurements in [21].

Growth Conditions, Microscopy, Data Extraction on Transcription Activation Times

Cells were grown overnight at 30 °C with aeration and shaking in lysogeny broth (LB) medium, supplemented with the appropriate antibiotics (35 μg/ml Kanamycin and 34 μg/ml Chloramphenicol). From the overnight cultures, cells were diluted into fresh LB medium, supplemented with antibiotics, to an optical density of $OD_{600} \approx 0.05$, and allowed to grow at 37 °C, 250 rpm, until reaching an $OD_{600} \approx 0.3$. Next, 100 $ng.ml^{-1}$ anhydrotetracycline (aTc) was added to induce $P_{LtetO-1}$ and produce MS2d-GFP, and 0.1% L-Arabinose to pre-activate the target gene, controlled by $P_{Lac-ara-1}$ [17,20]. Afterwards, cells were centrifuged (8000 rpm, for 1 minute), and re-suspended in the remaining LB medium. From this, a few microliters of cells were taken and placed between a 3% agarose gel pad and a glass coverslip, before assembling the FCS2 imaging chamber (Bioptechs, see Figure S1). Finally, the chamber was heated to the desired temperature (24 °C, 30 °C, 37 °C and 41 °C) and placed under the microscope.

We observed that, in the absence of IPTG, the cells produce the same (spurious) amount of RNA, with or without Arabinose (data not shown), in agreement with previous studies [20]. However, pre-induction by Arabinose much prior to induction by IPTG, enhances slightly the RNA production rate [16,18]. As such, we pre-induced cells with Arabinose [17,20] 45 minutes prior to introducing IPTG in the media. As such, we pre-induced cells with Arabinose [17]. This implies that, by the time IPTG is added, the cells already contain a constant amount of Arabinose. This is ensured by the presence of Arabinose in the original media and by the constant replenishment of this media during microscopy measurements (Methods and Figure S1). Thus, we do not expect any potential feedback mechanism associated to the Arabinose intake process to influence the transcription activation times measured here, following the introduction of IPTG in the media.

Cells were visualized by a 488 nm argon ion laser (Melles-Griot), and an emission filter (HQ514/30, Nikon) using a Nikon Eclipse (Ti-E, Nikon) inverted microscope with a 100x Apo TIRF (1.49 NA, oil) objective. Fluorescence images were acquired by C2+ (Nikon), a point scanning confocal microscope system, and Highly Inclined and Laminated Optical sheet (HILO) microscopy, using an EMCCD camera (iXon3 897, Andor Technology). The laser shutter was open only during exposure time to minimize photobleaching. All images were acquired with NIS-Elements software (Nikon). While imaging, cells were supplied with a constant flow of fresh LB medium (pre-warmed to the

same temperature as in the chamber), containing 1 mM of IPTG, 0.1% of L-Arabinose, and 100 ng.ml^{-1} of aTc, using a peristaltic pump (Bioptechs), at a rate of 0.1 mL min^{-1}. Images were taken once per minute for 2.5 hours. At each moment, we imaged 6 specific locations, to attain information on multiple lineages.

After performing a semi-automated cell segmentation and lineage construction [22], the moment of production of the first RNA by each cell lineage was obtained by selecting cells absent of RNA spots at the start of the imaging period (i.e., without leaky expression), and then detecting by visual inspection (from fluorescence images) when the first production occurs in each branch of each lineage (Figure 2B), after introducing the inducers.

Aside from visual inspection, fluorescent RNA spots and their intensities were also detected from the confocal images using the Gaussian surface-fitting algorithm proposed in [23] specifically for the purpose of detecting and quantifying MS2-GFP tagged RNAs. We found no significant difference between using this automatic algorithm and the visual inspection of the moment when the first RNA appears.

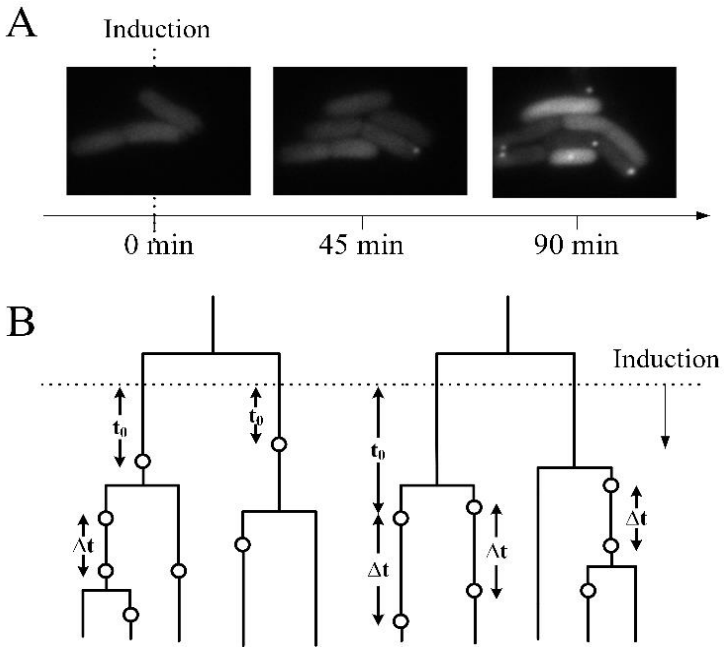


Figure 2. Data collection: (A) Cells are placed under the microscope at t=0 min and continuously supplemented with fresh medium. The reporter system (MS2d-GFP) is induced in liquid culture at t = -45 min. At t = 0 min, with the cells already having sufficient MS2d-GFP proteins for accurate RNA detection, transcription of the target RNA for MS2d-GFP is induced. (B) Illustration of RNA production events (circles) in cell lineages. Shown are the time for the first RNAs to appear (t_0) and the subsequent time intervals between consecutive RNA production events (Δt) in single cells. A dotted line indicates when the inducer of the target promoter is introduced.

As a side note, we found the rate of leaky expression to be very weak (less than 1 spot per ~20 cells prior to induction).

Finally, we note that the data on time intervals between consecutive RNA productions in individual cells used here was entirely obtained from [15]. There, time lapse microscopy was conducted on cells of the same strain, with the same constructs, and under the same induction and growth conditions as the ones used here.

Quantitative PCR for mean RNA quantification

Quantitative PCR (qPCR) was used to attain the induction curve of $P_{Lac\text{-}ara\text{-}1}$ as a function of IPTG concentration at 37 °C (for details, see Supplementary Material). This induction curve is shown in Figure S2. Visibly, for 0.5 mM IPTG and above, $P_{Lac\text{-}ara\text{-}1}$ is fully induced.

Estimation of intake times by deconvolution from empirical data on activation times and active transcription interval duration

The empirical method of MS2-GFP tagging of RNA allows for new RNAs containing multiple MS2-GFP binding sites to be detected shortly after they are produced [20]. From this data, one can directly extract the time intervals between consecutive RNA productions in individual cells following induction, as well as the time for the first RNA to be produced once inducers are added in the media. However, one cannot directly measure the time that inducers take to enter the cells and activate the target promoter. To obtain this information, we next propose a methodology based on deconvolution techniques for extracting this information from the data.

Given the model above, the mean time for the first RNA to appear in a cell following the addition of inducers in the media (here named t_0) depends on the time for inducers to enter the cell (reactions (1-2) in Supplementary Material) [4,5], here denoted as t_{int}. Also, it depends on the time for RNA production by an active promoter (which depends on the rate-limiting steps in transcription) [24,25], determined by reactions (3-6) in Supplementary Material, and here represented by Δt since, under full induction, this time should equal the time between consecutive RNA productions in active promoters [5]. In particular, we have:

$$t_0 = t_{int} + \Delta t \tag{1}$$

As the inducer intake and the production of the first RNA are independent, consecutive processes, one can use deconvolution to obtain a distribution of values of t_{int} (and, thus, mean and variance) from the data. Namely, for each temperature, one can deconvolve the probability density function (PDF) of the Δt distribution from the PDF of the t_0 distribution, provided that these two distributions are known [26].

For this, we estimate the PDFs of Δt and t_0 distributions as their best-fitted gamma distributions to the respective empirical distributions. We choose the gamma distribution as a model, since such distributions allow the mean and the variance to change independently, thus facilitating the fitting to the empirical distribution [14].

First, we use the gamma fits to the empirical Δt distributions reported in a previous work [14]. This fit used censored intervals between productions of consecutive transcripts extracted from live-cell measurements. The censoring accounts for the effects of finite sampling rate (60 s sampling interval), and thus improves the accuracy of the parameter estimation [27]. It also accounts for right-censored intervals, to compensate for the truncation of the right tail of the Δt distribution due to the finite cell division times. This fitting follows the maximum likelihood criteria [14].

Afterwards, to the measured t_0 distributions, we apply the same censored fitting procedure, but without right-censoring (as t_0 durations are not restricted by cell lifetime). Finally, we obtain the PDF of the t_{int} distribution using the Fast Fourier Transform (FFT) deconvolution method, as proposed by Sheu and Ratcliff [26], except that we do not apply frequency filtering, since our estimated t_0 and Δt PDFs do not contain high-frequency noise. As outlined by Sheu and Ratcliff [26], the result of the deconvolution may contain negative values, even though the PDF, by definition, cannot have values below zero. Those negative values should be interpreted as resulting from the uncertainty in the best-fit gamma distributions to t_0 and Δt empirical data, which, in turn, originates from uncertainty in the t_0 and Δt measurements. However, even if the selected models do not precisely depict the PDFs of the corresponding processes, the results of the deconvolution are still interpretable, even though the uncertainty in the deconvolution product is undefined [26]. Here, to allow such interpretation, we set the negative values of the t_{int} PDF to zero.

To estimate the uncertainty of our findings, we constructed bootstrap 95% confidence intervals (CIs) for mean and noise of the t_{int} distribution using non-parametric resampling of t_0 and Δt empirical data [28,29]. For this, for each temperature condition, we perform 2000 random resamples with replacement of the t_0 and Δt empirical distributions (using an original amount of samples), and obtain the t_{int} PDF for each resampled pair of t_0 and Δt distributions, which then allows obtaining the bootstrap distributions of the mean and CV^2 (squared coefficient of variation) of the t_{int} PDFs. We take 0.05 and 0.95 percentiles of those distributions as the 95% CIs of the estimated mean and CV^2 of the t_{int} distribution.

Estimation of intake times using Lineweaver-Burk equation

Aside from the method above, we make use of the Lineweaver-Burk equation [30] to estimate mean intake times. For this, from (1) and the model of gene expression (reactions 1-6 in Supplementary Material), note that as the amount of inducers in the media is increased, in a first stage, the inducers inside the cell will increase in number. As such, during this stage, both t_{int} and Δt will decrease with increasing inducer concentration. However, beyond a certain concentration of inducers in the media, further increases in this concentration will no longer lead to increases in the rate of RNA production (i.e. when the regime of full induction is reached), due to the rate-limiting steps in transcription and the

finite number of RNA polymerases inside the cell (reaction 6 in Supplementary Material). This well-known fact is also demonstrated here by Figure S2, which shows that, beyond a certain inducer concentration (both in the microscopy measurements and in the qPCR measurements) the rate of RNA production no longer increases with further increases in the IPTG concentration in the media.

Meanwhile, the time taken by the cell to intake inducers should continue to decrease with increases in inducer concentration in the media, even in the regime of full induction of transcription. Namely, in theory, for an infinite amount of inducers in the media, t_{int} should equal zero. In this regime, following the introduction of infinite number of inducers in the media, the total mean time taken to produce the first RNA will be equal to the duration of subsequent intervals between consecutive RNA productions, i.e.:

$$t_0\left(\left[\mathrm{IPTG}\right]=\infty\right)=\Delta t \tag{2}$$

Thus, provided that the decrease in t_{int} with the decrease of the inverse of the inducer concentration is linear (as assumed in our model reactions (1) and (2) in Supplementary Material), we can derive t_{int} in the 'control condition' using the Lineweaver-Burk equation [30] as follows:

$$t_{int}=\frac{[\mathrm{IPTG}]_2(t_{0_2}-t_{0_1})}{([\mathrm{IPTG}]_1-[\mathrm{IPTG}]_2)} \tag{3}$$

In equation (3), t_{0_1} and $[\mathrm{IPTG}]_1$ are, respectively, the mean t_0 and the inducer concentration in the control condition. Meanwhile, t_{0_2} and $[\mathrm{IPTG}]_2$ are the corresponding values in a condition where the inducer concentration differs from the control, and is above the minimum concentration to achieve maximum RNA production rate.

Also, one can calculate 95% CIs for the obtained mean t_{int} value based on the method of propagation of errors [31].

As a side note, this methodology is similar to the usage of τ plots, from which, by fitting a line to the results of measurements of the transcription rate for increasing RNA polymerase concentrations one can extract the duration of the events following the initiation of the open complex formation [16,24,32].

Inference of the number and duration of the sequential steps in the intake process by fitting with a sum of exponential steps

Our model of intake (reactions (1) and (2) in Supplementary Material) assumes 2 steps, each with a duration following an exponential distribution, in agreement with measurements at optimal temperatures [5,6]. However, as noted, our modelling strategy allows considering the possibility that, at different temperatures, additional or less steps may be rate-limiting.

To determine the number of steps, one can perform fittings of *d*-steps models (each step following an exponential distribution) for increasing number of steps, until adding a step no longer improves the fitting. In such a model, as more steps are added and if the overall mean duration of the d-steps process is kept constant, the variance of the durations between events will decrease. The closer the d-exponential steps distribution is to a gamma distribution with a shape parameter set to d, the smaller will be its variance.

The d-exponential step model was chosen due to how we model transcription, namely, as a set of consecutive of chemical reactions, each of which having a distribution of intervals between consecutive occurrences that is expected to follow an exponential distribution. Also, there is significant accumulated evidence that, in *E. coli*, this model fits very well, in a statistical sense, the empirical distributions of many promoters [5,6,16,33,34].

Here we perform this fitting to a *d*-steps model for each temperature condition. For this, by deconvolution of the empirical data, we obtain a distribution of the duration of the intake process. From it, we determine the maximum likelihood fit of a model with *d* statistically independent steps, whose time lengths each follow an exponential distribution, with possibly different rates.

The likelihoods are compared using the likelihood ratio test, and the model with smallest *d* that cannot be rejected at the significance level 0.01 is selected in favor of a higher order model.

Note that this method does not allow determining the order of the steps, only their number and durations. Note also that, while changing temperature may not alter the number of rate-limiting steps, it may instead (or also) cause them to no longer be well modeled by elementary reactions as our model assumes. In that case, we expect the fitting to d exponential steps to require a higher number of steps than if the steps were elementary.

Results and Conclusions

$P_{Lac\text{-}ara\text{-}1}$ transcription activation kinetics is temperature dependent

We first studied, at the single cell level, the temperature dependence of the kinetics of transcription activation of $P_{Lac\text{-}ara\text{-}1}$ by IPTG. All empirical data were obtained from observing individual cells over time, using MS2d-GFP tagging of the target RNA, fluorescence microscopy, and image analysis techniques (Methods).

For this, we placed *E. coli* cells (DH5α-PRO) with a single-copy plasmid coding for the RNA target for MS2d-GFP under the control of $P_{Lac\text{-}ara\text{-}1}$, and fully activated its expression by adding IPTG (1 mM) to the media (Figure S2) while already under microscope observation (Figure 2). The MS2d-GFP reporters, expressed by a multi-copy plasmid, were induced prior to this, so that cells were flooded with MS2d-GFP by the time $P_{Lac\text{-}ara\text{-}1}$ was induced (Methods).

From the time series obtained (~2.5 hour long, with images taken every minute), for each temperature, we extracted t_0, the time taken by individual cells to produce the first RNA, following the addition of inducers in the media (Methods). Note that only one such event per lineage is considered and that cells already with one or more RNAs at the start of the observation period were discarded.

From these data, we calculated the mean, standard error, and CV^2 of t_0 values. Finally, we performed Kolmogorov-Smirnov (KS) tests to compare each distribution of t_0 values with the distribution at 37 °C (named 'control' condition). Results are shown in Table 1.

Table 1. Measurements of t_0 vs temperature. Shown are the number of measurements (N_{t_0}), mean (μ_{t_0}) standard error (SE) and CV^2 of the distribution of t_0 values ($CV^2_{t_0}$). The table also shows the *p*-value from the KS tests comparing the t_0 distributions at each temperature, with the distribution at 37 ˚C (control). For *p*-values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution can be rejected.

T (˚C)	N_{t_0}	$\mu_{t_0} \pm$ SE (s)	$CV^2_{t_0}$	KS-test for t_0 values vs 37 ˚C (*p*-value)
24	93	2743 ± 102	0.13	< 0.01
30	162	3020 ± 119	0.25	< 0.01
37	60	2109 ± 215	0.63	-
41	93	2379 ± 144	0.34	0.19

From the data in Table 1, we find that for temperatures lower than 37 °C, the activation time t_0 differs significantly from the control (in a statistically sense), with its mean (μ_{t_0}) being higher and its $CV^2_{t_0}$ (surprisingly) being lower for lower temperatures.

Qualitatively similar results were obtained (Table S1) using the *E. coli* JW0334 strain (see section 'Bacterial strains and plasmids').

Cell-to-cell variability of t_{int} decreases with decreasing temperature

Next, we investigate how the time for inducers to enter the cell, t_{int}, changes with temperature. For this, besides the data above, we make use of the data from [14], which consists of empirical distributions of intervals between consecutive RNA productions by active promoters in individual cells (Δt), under the same temperature conditions as above. These data therefore informs on the kinetics of active transcription (i.e. is not affected by intake times).

As mentioned in Methods, in accordance to our model (reactions 1-6 in Supplementary Material) and equation 1, the time for the production of the first RNA in each cell, following the introduction of inducers in the media (t_0), should consist of the time for the intake of the inducer by the cell (t_{int}) and the time taken by the active promoter to produce the first RNA (Δt). As these processes are consecutive and independent, it should

be possible to obtain the time-length for intake of the inducers (t_{int}) by deconvolving Δt from t_0.

For this, we performed model fitting with censoring to the data from live-cell measurements of t_0 (Table 1) and used the model fitting of empirical Δt values from [14]. In Figure 3, we show the empirical distribution and the best gamma fits of t_0.

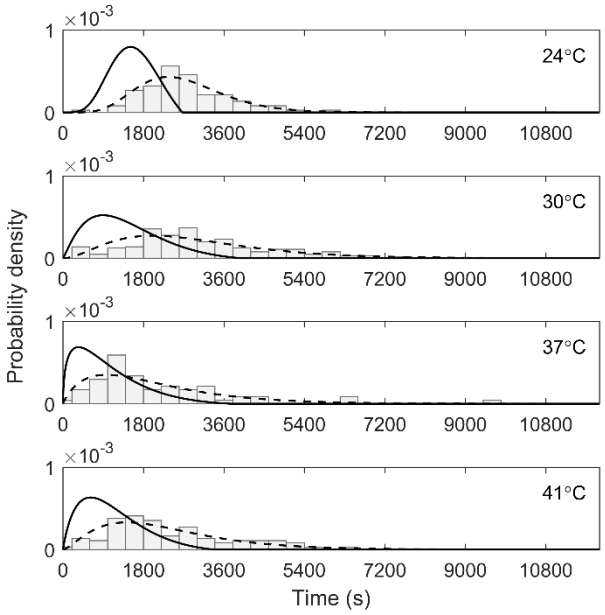


Figure 3. Empirical distribution of t_0 (histogram), along with the best gamma fit to t_0 (dashed line) and the deconvolved t_{int} (solid line), as function of temperature.

Next, we obtained the t_{int} distribution for each temperature condition from the deconvolution of Δt from t_0 (Methods). Results for the mean and CV^2 values of the distributions of t_{int} obtained from this deconvolution are shown in Table 2, along with the 95% CI. It is noted that the values at 37 °C are in agreement with previously reported measurements [5,6].

Meanwhile, the deconvolved distributions are shown in Figure 3. From these, we find a clear change in the shape of the t_{int} distribution as temperature is lowered.

Table 2. Mean and CV^2 of the deconvolved t_{int}, along with the 95% CI for each temperature condition.

T (˚C)	$\mu_{\widehat{t_{int}}}$ (s)	95% CI of $\mu_{\widehat{t_{int}}}$ (s)	$CV^2_{\widehat{t_{int}}}$	95% CI of $CV^2_{\widehat{t_{int}}}$
24	1548	[1316, 1799]	0.10	[0.06, 0.18]
30	1369	[1113, 1671]	0.32	[0.20, 0.48]
37	986	[726, 1329]	0.52	[0.28, 0.95]
41	1083	[807, 1402]	0.37	[0.23, 0.63]

From Table 2, we find that the mean duration of the intake process, $\mu_{\widehat{t_{int}}}$, is the lowest while the variability, $CV^2_{\widehat{t_{int}}}$, is the highest at 37 °C. Meanwhile, at the lowest temperature tested (24 °C) the opposite occurs ($\mu_{\widehat{t_{int}}}$ is the highest and $CV^2_{\widehat{t_{int}}}$ is the lowest).

Also, from the values of t_0 (Table 1) and t_{int} (Table 2), we find that the dynamics of intake plays a major role in the dynamics of transcription activation in all temperature conditions, both regarding the mean duration of activation and its cell-to-cell variability. Thus, it is not a surprise that t_{int} behaves similarly to t_0 with changes in temperature.

Finally, note that the fact that noise is reduced with decreasing temperature suggests that the process becomes more sub-Poissonian, which could occur, e.g., if the number of the rate-limiting steps in the intake process increases with decreasing temperature.

As a side note, we also conducted similar experiments in the absence of IPTG, so as to estimate the level of toxicity due to induction by 1 mM IPTG. We found no difference in cell growth rate between the two conditions, and thus conclude that the levels of toxicity are not significant.

Validation of the inferred mean t_{int} using the Lineweaver-Burk equation

It is possible to empirically validate the mean value of the deconvolved t_{int} using the Lineweaver-Burk equation (Methods). For this, from individual cells at 24 °C, 37 °C and 41 °C, we measured the time between the moment of induction and the moment when the first RNA is produced for IPTG concentrations of 1 mM and of 0.5 mM. Note that both of these concentrations suffice to reach maximum induction in cells under the microscope (as shown in Figure S2). Because of this, Δt does not differ between the two conditions, and only affects t_{int}. From the measurements of t_0 in these two induction levels at a given temperature, using the Lineweaver-Burk equation, one can extrapolate the value of t_0 for infinite inducer concentration, which allows estimating the mean intake time at that temperature (Table 3).

Table 3. Mean t_{int} ($\mu_{t_{int}}$) obtained from the Lineweaver-Burk equation and 95% CI of $\mu_{t_{int}}$ for various temperatures.

T (˚C)	$\mu_{t_{int}}$ (s)	95% CI of $\mu_{t_{int}}$ (s)
24	2434	[1949, 2918]
37	1322	[842, 1801]
41	1459	[1113, 1804]

From Table 3, we find that, in accordance with the results of deconvolution (Table 2), the mean t_{int} is highest at 24 °C, and is similar at 37 °C and 41 °C, being slightly smaller at 37 °C.

Quantitatively, we find that these values are ~ 35% larger (for 37 °C and 41 °C) and ~50% for 24 °C than those in Table 2. This is expected, as the deconvolution method is known to underestimate the peak value of the PDF [26].

Finally, we note that the value at 37 °C is also in clear agreement with a previous estimation of intake times at this temperature [6].

Number and duration of the rate-limiting steps of the intake process differs with temperature

To investigate the hypothesis that temperature affects the number and duration of the rate-limiting steps of the intake process, next, from the deconvolved t_{int} distributions of each temperature condition, we estimated the number and duration of these steps in maximum likelihood sense (Methods).

For this, we generalize the model of intake depicted by reactions (1) and (2) in Supplementary Material to a *d*-steps model, each exponentially distributed in duration, so that the number and duration of the rate-limiting steps are allowed to differ between the temperature conditions.

Results of this estimation are shown in Table 4, where we present the number and duration of the steps of the best fit model, along with the log-likelihood values. Meanwhile, in Table S2, we show the results for each condition when assuming specifically 1, 2, 3, and 4 steps, along with the *p*-values of the tests comparing pairs of models that are used to select the best model. Finally, in Figure 4 we show the best fit to the deconvolved t_{int} for each condition.

Table 4. Rate-limiting steps in the intake process determined by maximum likelihood estimation. Shown are the number of steps, the log-likelihood, the durations of the steps of the inferred models for each condition, and the CV^2 of the best fit. We fit the models to 10^5 random samples from the deconvolved t_{int} distribution. Note that there is no implied temporal order of the steps.

T (˚C)	No. Steps	Steps Durations	Log-likelihood	$CV^2_{\widehat{t_{int}}}$
24	≥4	(387, 387, 387, 386)	-774461	0.25
30	3	(457, 457, 457)	-801201	0.33
37	2	(667, 319)	-783576	0.56
41	3	(532, 532, 20)	-783350	0.48

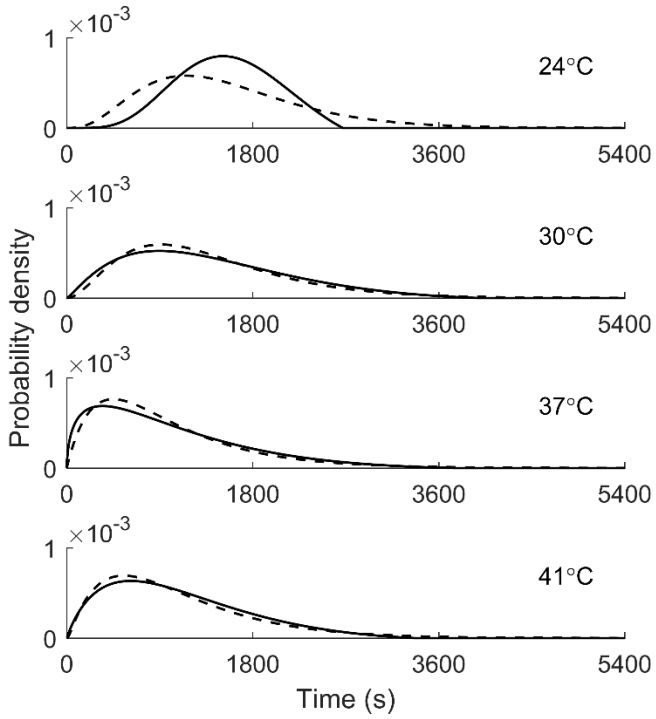


Figure 4. Deconvolved t_{int} distributions (solid line) and their best-fit d-steps model (dashed line).

Importantly, this result is in agreement with previous studies using data from cells at 37 °C [5].

From Tables 4 and S1, for all conditions, the test rejects the 1-step model in favor of a higher order model. This is expected, given the existence of the two membranes in the cell walls of *E. coli* cells and the time that inducers are expected to take to cross the periplasm in between [6].

Also, interestingly, the 2-steps model is the preferred one for cells at 37 °C and 41 °C (the step with a 20 s duration for the 41 °C condition can be disregarded, as the microscopy images are separated by 60 s intervals).

Meanwhile, at lower temperatures, higher order models (3 or more steps) are preferred, indicating that other steps become rate-limiting (in agreement with the deconvolution results), and/or that the steps duration may no longer follow an exponential distribution.

In this regard, we interpret the fact that a 4-steps model did not suffice to model the 24 °C condition (see Figure 4) as evidence for a significant change in the kinetics of intake with temperature, which renders the multi-step, exponentially distributed model incapable of fully capturing the dynamics. We hypothesize that this may be the consequence of increased viscosity of the cytoplasm and periplasm [14], along with changes in the physical properties and functionality of the intake 'machinery' in the cell walls.

Note that the CV^2 values of the best fits for 30 °C, 37 °C and 41 °C match the estimated values of the corresponding t_{int} distributions deconvolved from the fits to the empirical data. While the best fit in 24 °C condition has higher CV^2 than the deconvolved t_{int} (which is expected from the fact that the 4-steps model did not suffice to model the 24 °C condition), the trend in CV^2 of the deconvolved distributions and of their best fits is the same.

Finally, note that, in several cases, the time scales of the steps are identical. This may be due to an unknown artefact of the inference method or be representative of the real kinetics of intake of this inducer.

Discussion

In this work, we studied the single-cell dynamics of intake of IPTG, an inducer of the promoter $P_{Lac\text{-}ara\text{-}1}$, as a function of temperature. Rather than focusing on biological cellular adaptations, we focused solely on rapid physical changes due to temperature shifts in the process of inducer intake and consequent transcription kinetics.

For this, we first measured *in vivo* the time taken by individual cells to produce the first RNA, following the start of induction. From this, and previously collected data on the dynamics of RNA production by $P_{Lac\text{-}ara\text{-}1}$ [14], we applied two novel, independent methods to obtain the single-cell intake kinetics of the inducers, for each temperature condition. These methods' results were consistent with one another.

From this, first, we established that the response of the distribution of intake times of individual cells to temperature changes remains similar to that of the distribution of transcription activation times as temperature is changed, much due to the fact that most of the activation time is spent in the intake process in all conditions. Interestingly, the mean value of these distributions increases while their variability decreases for decreasing temperatures.

Since the intake process is bound to consist of multiple consecutive steps (in the case of IPTG, it was previously shown to be well modeled by a 2-step process for cells at 37 °C [5,6], we hypothesize that the decrease in variability could be the result of the emergence of additional rate-limiting steps in this process with decreasing temperature. The results of the maximum likelihood estimation tests support this view.

Further, they suggest that, at the lowest temperature condition tested here, the process is, from a dynamical point of view, 'too complex' to be well fitted by a sum of a small number (less than 5) of exponential steps. We hypothesize that this is clear evidence that the duration of one of the steps of the intake process becomes non-exponential-like at low temperatures. There are several potential causes for this (and perhaps multiple causes), and they are likely not accounted by our model (else, the increase in number of exponential steps would have allowed to fit the data well). We expect these potentials causes to range from malfunctioning of the porins in the membrane responsible for the diffusive intake of

the inducers, increased viscosity of the cytoplasm and periplasm, alteration of the physical properties of the outer and inner membranes, etc.

It is worth noting that the application of the Lineweaver-Burk equation to extract the mean value of the intake times is a methodology that has not been previously used, but we expect it to be of use in future works as well. It requires measuring transcription activation times for various inducer concentrations (at least 2) above the minimum concentration required for maximum induction. It is limited by the fact that the speed of intake is assumed to change linearly with the inverse of the inducer concentration, which may not always be the case. However, we expect this to be the case within certain ranges of inducer concentrations for simpler intake (mostly diffusion-based), mechanisms. Thus, it should be applicable to the study of a wide range of cellular intake mechanisms.

Overall, we conclude that different environmental conditions cause significant changes in the single-cell distributions of intake times of transcription inducers, which is expected to have a significant effect on the degree of heterogeneity in cell populations and cell lineages, due to the longevity of the transients during which this phenomenon has a strong effect in RNA numbers.

In the future, one important aspect that requires further research is the cause for the reduced cell-to-cell diversity in response times with decreasing temperatures, which we believe to be due to the emergence of rate-limiting steps in the intake process. Which steps and how they emerge are open questions, whose answers will help better understanding the robustness of the intake systems of *E. coli*.

Acknowledgements

Work supported by Academy of Finland (295027 and 305342 to ASR), Jane & Aatos Erkko Foundation (610536 to ASR), Tampere University of Technology President's Graduate Program (SS), Finnish Academy of Science and Letters (SO), Doctoral Programme of Computing and Electrical Engineering of TUT (NG) and Erasmus+ program 2919(713)2915/2016/SMS (CP). The funders had no role in study design, data collection and analysis, decision to publish, or manuscript preparation.

References

[1] Arkin A, Ross J and McAdams H 1998 Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells *Genetics* **149** 1633–48

[2] van Kampen NG, Reinhardt WP 1983 Stochastic Processes in Physics and Chemistry Physics Today **36(2)** 78–80

[3] Elowitz B, Siggia D, Levine AJ, Swain PS, Siggia E, and Swain P 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6

[4] Megerle J, Fritz G, Gerland U, Jung K, and Rädler J 2008 Timing and dynamics of single cell gene expression in the arabinose utilization system *Biophys J* **95** 2103–15

[5] Mäkelä J, Kandavalli V, and Ribeiro AS 2017 Rate-limiting steps in transcription dictate sensitivity to variability in cellular components *Sci Rep* **7** 10588

[6] Tran H, Oliveira SMD, Goncalves N, and Ribeiro AS 2015 Kinetics of the cellular intake of a gene expression inducer at high concentrations *Mol Biosyst* **11** 2579–87

[7] Hensel Z, Feng H, Han B, Hatem C, Wang J, and Xiao J 2012 Stochastic expression dynamics of a transcription factor revealed by single-molecule noise analysis *Nat Struct Mol Biol* **19** 797–802

[8] Rosenfeld N, Young J, Alon U, Swain P, and Elowitz M 2005 Gene regulation at the single-cell level *Science* **307** 1962–5

[9] Robert L, Paul G, Chen Y, Taddei F, Baigl D, and Lindner AB 2010 Pre-dispositions and epigenetic inheritance in the *Escherichia coli* lactose operon bistable switch *Mol Syst Biol* **6** 357

[10] Kiviet D, Nghe P, Walker N, Boulineau S, Sunderlikova V, and Tans SJ 2014 Stochasticity of metabolism and growth at the single-cell level *Nature* **514** 376–9

[11] Yun HS, Hong J, and Lim HC 1996 Regulation of Ribosome Synthesis in *Escherichia coli*: Effects of Temperature and Dilution Rate Changes *Biotechnol Bioeng* **52** 615–24

[12] Gupta A, Lloyd-Price J, Neeli-Venkata R, Oliveira SMD, and Ribeiro AS 2014 *In vivo* kinetics of segregation and polar retention of MS2-GFP-RNA complexes in *Escherichia coli Biophysical Journal* **106(9)** 1928–37

[13] Choi PJ, Cai L, Frieda K, and Xie XS 2008 A stochastic single-molecule event triggers phenotype switching of a bacterial cell *Science* **322** 442–6

[14] Oliveira SMD, Häkkinen A, Lloyd-Price J, Tran H, Kandavalli V, and Ribeiro AS 2016 Temperature-Dependent Model of Multi-step Transcription Initiation in *Escherichia coli* Based on Live Single-Cell Measurements *PLoS Comput Biol* **12** e1005174

[15] Oliveira SMD, Neeli-Venkata R, Goncalves N, Santinha JA, Martins L, Tran H, Mäkelä J, Gupta A, Barandas M, Häkkinen A, Lloyd-Price J, Fonseca JM and Ribeiro AS 2016 Increased cytoplasm viscosity hampers aggregate polar segregation in *Escherichia coli Mol Microbiol* **99** 686–99

[16] Lloyd-Price J, Startceva S, Chandraseelan JG, Kandavalli V, Goncalves N, Häkkinen A, and Ribeiro AS 2016 Dissecting the stochastic transcription initiation process in live *Escherichia coli DNA Res* **23** 203–14

[17] Lutz R, and Bujard H 1997 Independent and tight regulation of trasncriptional units in *Escherichia coli* via the LacR/O, the TetR/O and Arac/I1-I2 regulatory elements *Nucleic Acids Res* **25** 1203–10

[18] Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, Datsenko KA, Tomita M, Wanner BL, and Mori H 2006 Construction of *Escherichia coli* K-12 in-frame, single-

gene knockout mutants: the Keio collection *Mol Syst Biol* **2** 1234–44

[19] Marbach A, and Bettenbrock K 2012 Lac operon induction in *Escherichia coli*: Systematic comparison of IPTG and TMG induction and influence of the transacetylase LacA *J Biotechnol* **157** 82–8

[20] Golding I, Paulsson J, Zawilski S, and Cox EC 2005 Real-time kinetics of gene activity in individual bacteria *Cell* **123** 1025–36

[21] Golding I, and Cox EC 2004 RNA dynamics in live *Escherichia coli* cells *Proc Natl Acad Sci U S A* **101** 11310–5

[22] Häkkinen A, Muthukrishnan A, Mora A, Fonseca JM, and Ribeiro AS 2013 CellAging: a tool to study segregation and partitioning in division in cell lineages of *Escherichia coli Bioinformatics* **29** 1708–9

[23] Häkkinen A, and Ribeiro AS 2015 Estimation of GFP-tagged RNA numbers from temporal fluorescence intensity data *Bioinformatics* **31** 69–75

[24] McClure WR 1985 Mechanism and control of transcription initiation in prokaryotes *Annu Rev Biochem* **54** 171–204

[25] Lutz R, Lozinski T, Ellinger T, and Bujard H 2001 Dissecting the functional program of *Escherichia coli* promoters: the combined mode of action of Lac repressor and AraC activator *Nucleic Acids Res* **29** 3873–81

[26] Sheu CF, and Ratcliff R 1995 The application of fourier deconvoution to reaction time data: a cautionary note *Psychol Bull* **118** 285–99

[27] Häkkinen A and Ribeiro AS 2016 Characterizing rate limiting steps in transcription from RNA production times in live cells *Bioinformatics* **32** 1346–52

[28] DiCiccio T and Efron B 1996 Bootstrap Confidence Intervals *Stat Sci* **11** 189–228

[29] Carpenter J and Bithell J 2000 Bootstrap confidence intervals: When, which, what? A practical guide for medical statisticians *Stat Med* **19** 1141–64

[30] Lineweaver H and Burk D 1934 The Determination of Enzyme Dissociation Constants *J Am Chem Soc* **56** 658–66

[31] Press W, Teukolsky S, Vetterling W, and Flannery B 1992 *Numerical Recipes in C* (Cambrigde: Cambrigde University Press) 661–5

[32] Bertrand-Burggraf E, Lefèvre J, 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** 1697–706

[33] Kandavalli VK, Tran H, and Ribeiro AS 2016 Effects of σ factor competition are promoter initiation kinetics dependent *BBA - Gene Regul Mech* **1859** 1281–8

[34] Muthukrishnan A, Martikainen A, Neeli-Venkata R, and Ribeiro AS 2014 In vivo transcription kinetics of a synthetic gene uninvolved in stress-response pathways in stressed *Escherichia coli* cells *PLoS One* **9** e109005

Supplementary Material for "Temperature-dependence of the single-cell variability in the kinetics of transcription activation in *Escherichia coli*"

Nadia S.M. Goncalves[†], Sofia Startceva[†], Cristina S.D. Palma[†, ‡], Mohamed N.M. Bahrudeen[†], Samuel M.D. Oliveira[†] and Andre S. Ribeiro[†, §, *]

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

[‡] CA3 CTS/UNINOVA. Faculdade de Ciencias e Tecnologia, Universidade Nova de Lisboa, Quinta da Torre, 2829-516, Caparica, Portugal.

[§] Multi-scaled Biodata Analysis and Modelling Research Community, Tampere University of Technology, 33101, Tampere, Finland.

[*] Corresponding author. E-mail: andre.ribeiro@tut.fi, Tel: +358408490736.

1. Supplementary Methods

Quantification of target gene activity by qPCR and microscopy

Cells with the plasmid carrying the target gene (pIG-BAC-$P_{lac\text{-}ara\text{-}1}$-mRFP1-96xMS2) were grown overnight at 30 °C with aeration and shaking in lysogeny broth (LB) medium, supplemented with the appropriate antibiotics (35 μg/ml Kanamycin and 34 μg/ml Chloramphenicol). From the overnight cultures, cells were diluted into fresh LB medium, supplemented with antibiotics, to an optical density of $OD_{600} \approx 0.05$, and allowed to grow at 37 °C, 250 rpm, until reaching an $OD_{600} \approx 0.3$.

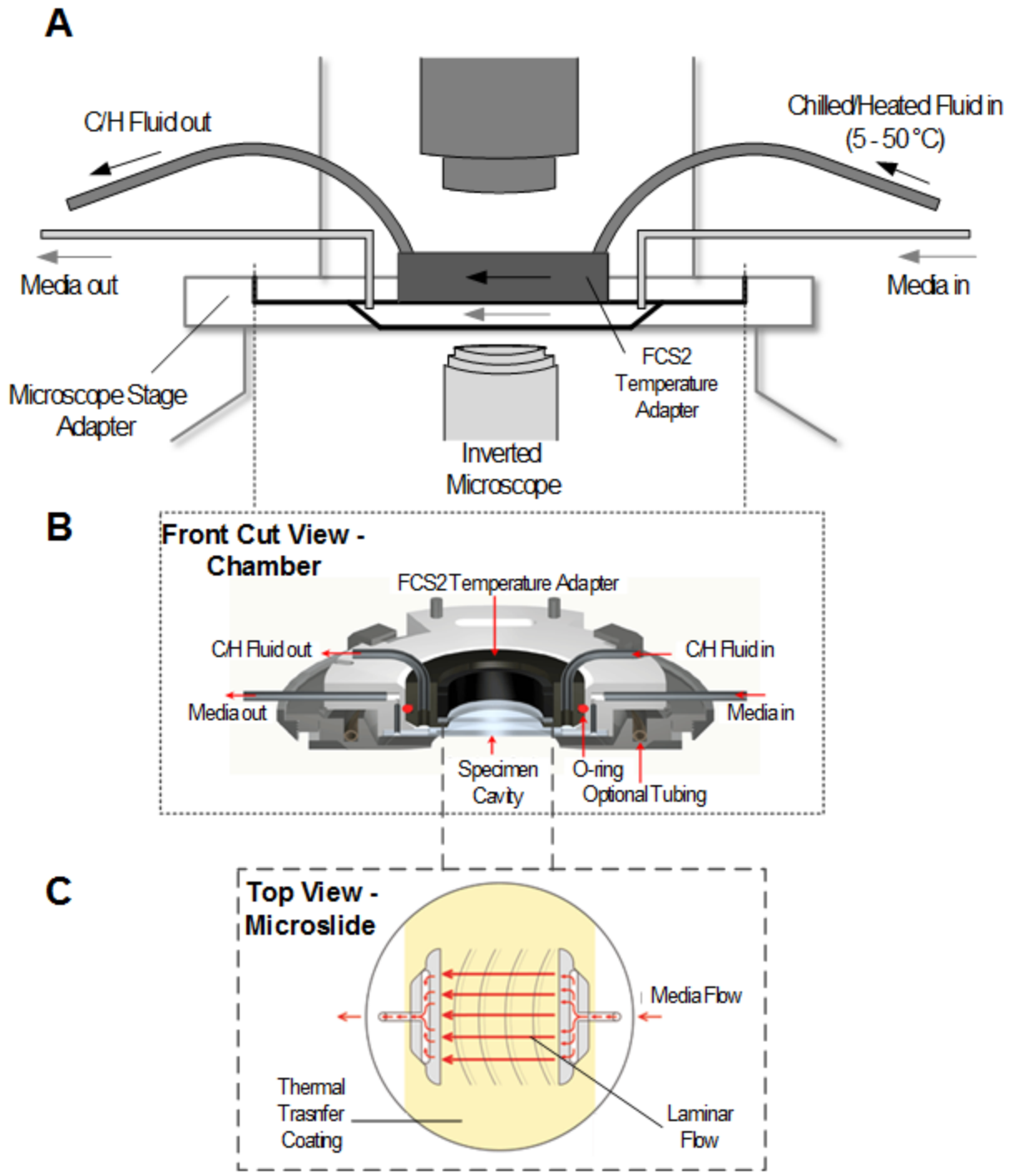
Next, qPCR was performed to analyze the fold change in mRNA production with induction of the target gene. From the culture described above, cells were then grown in LB media, at 37 ºC. To obtain the induction curve of this promoter, we first pre-induced with L-Arabinose (0.1%). Next, we induced with different concentrations of IPTG (0, 0.05, 0.1, 0.25, 0.5 and 1 mM). After collecting the cells by centrifugation at 8000 rpm for 5 minutes, twice the cell culture volume of RNA protect reagent (Qiagen) was added to the reaction tube, following the addition of Tris EDTA Lysozyme (15mg/ml) buffer (pH 8.0) for enzymatic lysis. Total RNA was isolated using RNeasy kit (Qiagen), according to the manufacturer instructions. Samples with total isolated RNA were treated with DNase for residual DNA removal. The A260/A280nm ratio of the isolated RNA samples was assessed using a Nanovue plus spectrophotometer (GE Healthcare) with the value obtained (2.0-2.1) indicating a highly purified RNA. Additionally, the resulting RNA yield was used to normalize the RNA concentration in the samples with varying IPTG concentrations. Following that, iSCRIPT reverse transcription super mix (Biorad) 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

qPCR reaction was carried out in technical triplicates with a total reaction volume of 20 µL. To amplify the target gene mRPF1 and reference gene 16SrRNA, we used the following primers, respectively: i) forward 5' TACGACGCC GAGGTCAAG 3' and reverse 5' TTGTGGGAGGTGATGTCCA 3', and ii) forward: 5' CGTCAG CTCGTGTTGTGAA 3' and reverse: 5' GGACCGCTGGCAA CAAAG 3'. The qPCR experiments were performed using a MiniOpticon Real time PCR system (Biorad). The following thermal cycling protocol was used: 40 cycles at 95 ºC for 10 s, 52 ºC for 30 s, and 72 ºC for 30 s. No-RT and No-Template controls were used to crosscheck non-specific signals and contamination, and the efficiency of the PCR reactions were found to be greater than 95%. The data from the CFX Manager™ software was then used to calculate the fold change in mRNA production and its standard error [1], which are presented in Figure S2.

Meanwhile, microscopy measurements were conducted as described in Methods (section "Growth Conditions, Microscopy, Data Extraction on Transcription Activation Times").

We note two main differences between the qPCR and the microscopy measurements. First, in the qPCR measurements, the report system is not activated, since it is not required to obtain the measurements and since this does not cause significant differences in target RNA production rates (data not shown). Second, in the qPCR measurements, the activation of the target gene is performed in liquid. In general, this results in higher absolute expression levels than if the induction is performed under the microscope. However, this does not constitute a problem as it does not alter the inducer concentration at which maximum induction is reached (Figure S2).

FCS2 imaging system



Supplementary Figure S1. (A) Schematic illustration of the CFCS2 microfluidics and the temperature control system for cell cultures while under microscope observation. The CFCS2 chamber is mounted on the stage of an inverted microscope. This device is comprised of two independent fluidic systems. One is a thermo-chiller device (not shown), which is connected to two inlets and two outlets of the CFCS2 chamber. This device controls the temperature of the system (i.e. of the metal chamber and the optical cavity, where cells are placed) through the flow of heat/chilled fluidics, whose temperature can range from 5 ºC to 50 ºC ± 0.2 ºC. The second

device, a micro-perfusion device (not shown), is connected to one inlet and one outlet of the CFCS2. It constantly provides cells with fresh media and chemicals required for cell growth. (B) Illustrative front cut view of the optical cavity of the cooled FCS2 adapter (CFCS2). The CFCS2 is a modified version of the FCS2 system, in that it has an additional, independent tubing system to facilitate the circulation of a heat/chilled fluid, that increases/reduces temperature of the metal base and of the optical cavity of the chamber. (C) Schematic top view of the micro-aqueduct slide, which is placed inside the optical cavity. The slide allows laminar flow of fluids, when a uniform and rapid exchange of media is required across the cell population. Images shown in (B) and (C) are adapted from Bioptechs Inc. (http://www.bioptechs.com).

Model of Inducer Intake and Active Transcription

We assume the following models of transcription activation and active transcription [2]. First, regarding activation, when an inducer is added to the media, the gene is only activated after a multi-step process that includes events such as the entry and diffusion of inducers in the periplasm and then cytoplasm, binding of an inducer to a transcription factor repressing gene activity, etc. These events and their kinetics differ with the induction and repression systems of the gene [3, 4].

Relevantly, as mentioned, the strain used here lacks the ability to produce LacY [5]. As such, we expect the intake process of inducers (IPTG) to be purely diffusive-like. Also, as *E. coli* is gram-negative, the cell walls have an outer and inner membranes, with a periplasmic space in between. Thus, the activation process is expected to have at least two, consecutive rate-limiting steps: entering of inducers into the periplasm, followed by entering into the cytoplasm. In agreement, previous studies of the intake of IPTG at optimal temperatures (37 ºC) [2] have shown that the activation process of our target promoter, $P_{Lac\text{-}ara\text{-}1}$, in cells lacking LacY as those used here [5], is well modelled by a 2-step stochastic process of the form [29]:

$$I_{env} \xrightarrow{Int_1} I_{peri} \quad (1)$$

$$I_{peri} \xrightarrow{Int_2} I \quad (2)$$

Reaction (1) represents the entrance of an inducer molecule (I) into the periplasm, while reaction (2) models the passage of that inducer from the periplasm into the cytoplasm, where it can activate the target gene, e.g. by interacting with repressor molecules. In general, additional rate-limiting steps could, in theory, be modelled by a sequence of d-steps with exponential duration: $I_1 \rightarrow \ldots \rightarrow I_d$. This is particularly important when selecting a strategy to decompose the rate-limiting steps from the empirical data.

Meanwhile, transcription activation is modeled as follows:

$$Rep + I \underset{k_{inact}}{\overset{k_{act}}{\rightleftarrows}} Rep.I \quad (3)$$

$$P_{OFF} + I \xrightarrow{k_{act}} P_{ON} + Rep.I \quad (4)$$

$$P_{OFF} \underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}} P_{ON} + \text{Rep} \quad (5)$$

In (3), an inducer (I) binds to the repressor (Rep), creating a complex (Rep.I) that cannot repress the promoter (see reactions (5)). In our case, the repressor are LacI tetramers, and IPTG, the inducer, acts by binding to these tetramers, greatly reducing their binding affinity to the promoter [6]. We assume that this binding reaction is very weakly reversible. Also, the reactions necessary to form LacI tetramers are not explicitly considered since most LacI molecules in the cell are present in the form of tetramers.

In reactions (4), again an inducer binds to a repressor, but the repressor is bound to the promoter, which frees the promoter. We note that, for such to occur, the LacI tetramer must unbind from both DNA binding sites [6].

Reaction (5) allows for the repression of the promoter by free repressors and for the possibility of a repressor unbinding the promoter, without direct intervention of inducers.

Finally, active transcription by a free promoter is modeled as a multi-step process [7-11]:

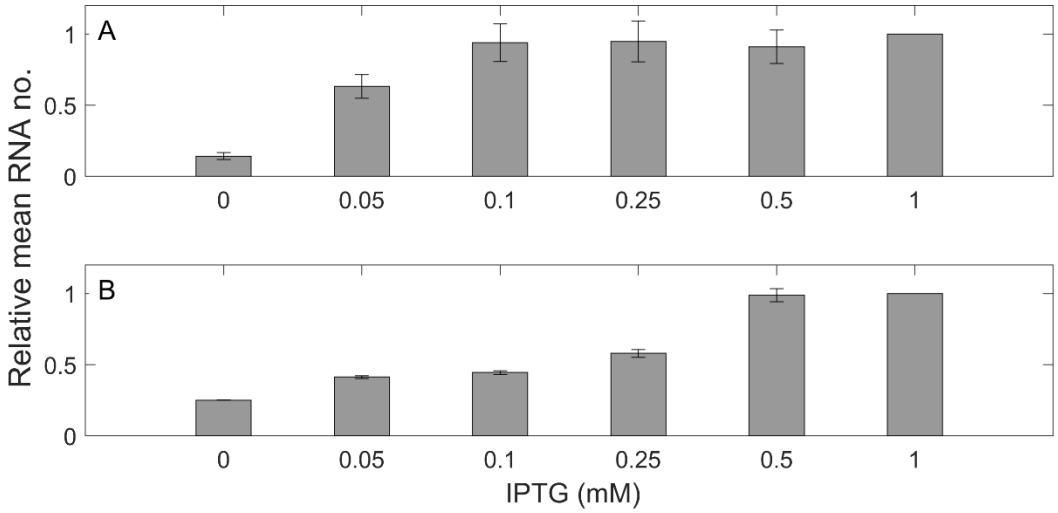
$$R+P_{ON} \xrightarrow{k_{cc}} RP_{cc} \xrightarrow{k_{oc}} RP_{oc} \xrightarrow{\infty} P_{ON} + R+RNA \quad (6)$$

In (6), R is the RNA polymerase. Once bound to the promoter, it forms a closed complex (RP_{cc}), which is followed by the open complex (RP_{oc}) formation, elongation (not rate-limiting, and thus not represented), and, finally, RNA production and RNA polymerase release (also not rate-limiting, and thus having an 'infinite' rate,∞).

It is of importance to note that, while not represented, the steps in (6) are all considered to be reversible (except the open complex formation, which, once initiated, is nearly irreversible [9]. I.e., the reactions in (6) are not to be interpreted as elementary transitions. Instead, they represent effective rates of the rate-limiting steps in transcription, thus defining the promoter strength [11].

2. Supplementary Results

2.1 Induction curve at 37 ºC



Supplementary Figure S2. Induction curve of $P_{Lac/ara-1}$ in cells at 37 ºC as obtained by microscopy (A) and qPCR (B). In the case of microscopy, from the single cell measurements, we calculated the mean and standard uncertainty from the distribution of RNA numbers produced in each cell during 1 hour following induction. In the case of qPCR, the mean RNA fold change and its standard uncertainty in each condition were extracted from 3 technical replicates. In both (A) and (B), values relative to the 1 mM induction condition were calculated in each condition using the Delta Method [12]. Also in both, measurements were conducted after induction of the target gene (IPTG added 1 hour prior to the measurements and 0.1% of L-Arabinose added 1 hour and 45 minutes prior to the measurements, see Methods).

2.2 Measurements of t_0 vs temperature for *E. coli* JW0334 strain

T (˚C)	N_{t_0}	$\mu_{t_0} \pm$ SE (s)	$CV^2_{t_0}$	KS-test for t_0 values vs 37 ˚C (p-value)
24	76	1939 ± 99	0.20	< 0.01
30	64	1725 ± 108	0.25	< 0.01
37	97	1089 ± 67	0.37	-
41	106	1729 ± 86	0.26	< 0.01

Supplementary Table S1. Measurements of t_0 vs temperature for *E. coli* JW0334 strain. Shown are the number of measurements (N_{t_0}), mean (μ_{t_0}) standard error (SE) and CV^2 of the distribution of t_0 values ($CV^2_{t_0}$). The table also shows the p-value from the KS tests comparing the t_0 distributions at each temperature, with the distribution at 37 ˚C (control). For p-values smaller than 0.01, the null hypothesis that the two sets of data are from the same distribution can be rejected.

2.3 Maximum log-likelihood fit to the deconvolved distributions of intake times

24 ˚C			
d	Log-likelihood	Durations	*p*-values ($d_1 = d_0 + 1$)
1	-834550	(1549)	0.00
2	-801008	(775, 775)	0.00
3	-784454	(517 , 517, 516)	0.00
4	-774461	(388 , 388, 388, 387)	-
30 ˚C			
d	Log-likelihood	Durations	*p*-values ($d_1 = d_0 + 1$)
1	-822278	(1370)	0.00
2	-803246	(685, 685)	0.00
3	-801201	(457, 457, 457)	1.00
4	-801204	(458, 456, 456, 0)	-
37 ˚C			
d	Log-likelihood	Durations	*p*-values ($d_1 = d_0 + 1$)
1	-789385	(986)	0.00
2	-783576	(667, 319)	1.00
3	-783576	(667, 319, 0)	1.00
4	-783576	(667, 319, 0, 0)	-
41 ˚C			
d	Log-likelihood	Durations	*p*-values ($d_1 = d_0 + 1$)
1	-798760	(1083)	0.00
2	-783444	(542, 542)	0.00
3	-783350	(532, 532, 20)	1.00
4	-783350	(532, 532, 20, 0)	-

Supplementary Table S2. Log-likelihood and durations of the steps of the inferred models with d-steps, for each temperature condition. The table shows, first, the number of steps (d) assumed, followed by the log-likehood, and the duration of the steps (the order of these steps cannot be determined by this method). The last column shows the *p*-values of the likelihood-ratio tests between pairs of models for each condition. The null model is the $d_0 = [1:3]$ step model, while the alternative model is the $d_1 = d_0 + 1$ step model.

References

[1] Livak K J and Schmittgen T D 2001 Analysis of relative gene expression data using real-time quantitative PCR and the 2-DDCT method. *Methods* **25** 402–8

[2] Tran H, Oliveira S M D, Goncalves N and Ribeiro A S 2015 Kinetics of the cellular intake of a gene expression inducer at high concentrations *Mol. Biosyst.* **11** 2579–87

[3] Megerle J A, Fritz G, Gerland U, Jung K and Rädler J O 2008 Timing and dynamics of single cell gene expression in the arabinose utilization system *Biophys. J.* **95** 2103–15

[4] Schleif R 2000 Regulation of the L-arabinose operon of Escherichia coli *Trends Genet.* **16** 559–65

[5] Lutz R and Bujard H 1997 Independent and tight regulation of trasncriptional units in Escherichia coli via the LacR/O, the TetR/O and Arac/I1-I2 regulatory elements. *Nucleic Acids Res.* **25** 1203–10

[6] Lewis M 2005 The lac repressor *Comptes Rendus - Biol.* **328** 521–48

[7] Ribeiro A S, Zhu R and Kauffman S A 2006 A General modeling strategy for gene regulatory networks with stochastic dynamics *J. Comput. Biol.* **13** 1630–9

[8] Zhu R, Ribeiro A S, Salahub D and Kauffman S A 2007 Studying genetic regulatory networks at the molecular level: Delayed reaction stochastic models *J. Theor. Biol.* **246** 725–45

[9] McClure W R 1985 Mechanism and control of transcription initiation in prokaryotes *Annu. Rev. Biochem.* **54** 171–204

[10] DeHaseth P L, Zupancic M L and Record M T 1998 RNA polymerase-promoter interactions: The comings and goings of RNA polymerase *J. Bacteriol.* **180** 3019–25

[11] Mulligan M E, Hawley D K, Entriken R and McClure W R 1984 Escherichia coli promoter sequences predict in vitro RNA polymerase selectivity *Nucleic Acids Res.* **12** 789–800

[12] Casella G and Berger R L 2001 The Delta Method. Statistical Inference, 2nd ed. Duxbury Press, Pacific Grove, CA, 240–5

Tampereen teknillinen yliopisto
PL 527
33101 Tampere

Tampere University of Technology
P.O.B. 527
FI-33101 Tampere, Finland

ISBN 978-952-15-4186-5
ISSN 1459-2045